



HAL
open science

**L'article de recherche en Sciences du langage et en
Lettres. Figure de l'auteur et identité disciplinaire du
genre.**

Fanny Rinck

► **To cite this version:**

Fanny Rinck. L'article de recherche en Sciences du langage et en Lettres. Figure de l'auteur et identité disciplinaire du genre.. Linguistique. Université Grenoble 3, 2006. Français. NNT : 2006FRE39039 . tel-01881445

HAL Id: tel-01881445

<https://hal.science/tel-01881445>

Submitted on 15 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**UNIVERSITE GRENOBLE III - STENDHAL
U.F.R. DE SCIENCES DU LANGAGE**

THESE
Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITE GRENOBLE III
Discipline : Sciences du Langage

Soutenance publique le 17 novembre 2006

Fanny RINCK

L'article de recherche en Sciences du langage et en Lettres.
Figure de l'auteur et identité disciplinaire du genre

Directeurs de thèse :

Madame Françoise BOCH
Monsieur Francis GROSSMANN

Membres du jury :

Madame Kjersti FLØTTUM (Université de Bergen), rapporteur.
Monsieur Alain RABATEL (Université Lyon 2), examinateur.
Monsieur François RASTIER (CNRS, Paris), rapporteur.

Remerciements

Ce travail est le fruit d'échanges et de rencontres et je voudrais adresser ici ma reconnaissance à tous ceux qui ont contribué pendant les 4 années de la thèse et sans doute en amont déjà à sa réalisation. Grâce à eux, elle représente avant tout pour moi une période heureuse et stimulante.

Mes remerciements vont d'abord à Françoise Boch et Francis Grossmann. La gratitude que je souhaite leur exprimer dépasse les formules consacrées qui me viennent maladroitement en tête. Leurs conseils sont précieux, leur soutien porteur, les discussions très vivifiantes, et je leur dois beaucoup de m'avoir laissée explorer des chemins en linguistique, tout en les balisant de rigueur et d'éthique.

Je tiens à remercier aussi Kjersti Fløttum, Alain Rabatel et François Rastier pour leur bienveillance et pour leur dynamisme dans le champ des recherches sur le discours scientifique et sur l'énonciation. Mon travail a beaucoup gagné à rencontrer leurs textes et participer aux séminaires qu'ils ont organisés, et je leur suis très reconnaissante d'avoir accepté de l'examiner.

Merci également à Agnès Tutin pour nos collaborations et, avec elle, à toute l'équipe du Département d'Informatique Pédagogique pour la découverte du monde du traitement automatique des langues, à Christine Barré-De Miniac pour la richesse des séminaires des doctorants en didactique, et enfin à Jean-Pierre Chevrot pour ses enseignements en statistique.

Un grand merci aussi à Céline Poudat pour nos études communes sur l'article, occasions de séances de travail tout aussi sympathiques qu'exaltantes.

Mes remerciements iront enfin à tous ceux qui font la vie de laboratoire et sa convivialité, et notamment aux doctorant(e)s du Lidilem pour leurs multiples attentions, leur sagacité en linguistique et ailleurs, et, il faut le dire, leur humour. Une pensée particulière à Céline, Carmen, Aurélie, Laurence et Pantxo pour leur aide et leurs encouragements.

Sommaire

Sommaire du tome I

Remerciements	I
Sommaire du tome II. Annexes.....	IX
Table des figures.....	XI
Introduction	1

Chapitre 1

Aborder les pratiques d'écriture sous l'angle de la problématique des genres..... 5

1.1. Mise au point terminologique : genre, texte et discours 5

1.1.1. Texte et discours.....	6
1.1.2. Options terminologiques et conceptuelles.....	8

1.2. L'interdépendance entre usages linguistiques et fonctionnements sociaux : les genres comme entités socio-historiques et formelles 9

1.2.1. Pratiques discursives et formations sociales	9
1.2.2. Le genre comme notion « bi-face ».....	10
- Le genre dans sa dimension textuelle	10
- Le genre dans sa dimension contextuelle.....	11
1.2.3. Aborder les relations entre propriétés sociales et formelles : approches taxinomistes <i>versus</i> empiriques des genres	12

1.3. Le genre comme principe de régulation des pratiques discursives 13

1.3.1. L'incidence des genres sur la langue et son étude	13
1.3.2. Le rôle des genres du point de vue du sujet et de ses pratiques discursives : les genres comme modèles textuels	15
- Des modèles pour l'interprétation.....	15
- Des modèles pour la production	16
- Stabilité des genres et prescriptivité du modèle.....	16
- Le déjà-là des genres : un obstacle à la création verbale ?	17
1.3.3. Aux fondements des genres : l'hétérogénéité constitutive du discours et du sujet	18
- De la communication à l'interaction.....	18
- La production verbale est dialogique	19
- L'hétérogénéité constitutive du sujet.....	20
1.3.4. Retour à la question de la production verbale : les genres comme modèles de textes, contraintes, et compétence.	22
- Le principe explicatif de l'adoption par le locuteur de modèles du genre : intertexte et évaluation sociale	22
- Les genres comme contraintes de la production verbale	24

- Les contraintes externes au texte ou la question du contexte	25
- Vers une problématisation de l'articulation texte/ contexte : le genre.....	27
1.3.5. Conscientisation des modèles du genre et compétence générique.....	30
1.4. La secondarité de certains genres : des textes à l'action	32
1.5. En guise de bilan : la caractérisation empirique des genres, et du genre de l'article de recherche.....	36

Chapitre 2

La figure de l'auteur comme figure du genre.

Pour une analyse énonciative et polyphonique des textes.	45
2.1. En filigrane dans les analyses des genres et de l'écrit de recherche, l'intérêt de la dimension énonciative des textes.....	46
2.2. Enoncé, énonciation et énonciation représentée dans l'énoncé	50
2.2.1. L'énonciation comme mise en fonctionnement de la langue.....	51
2.2.2. Les relations entre l'énoncé et l'énonciation : la réflexivité énonciative.....	53
2.2.3. Les marques étudiées par les approches énonciatives.....	56
2.3. De la polyphonie à la définition polyphonique d'une figure de l'auteur	58
2.3.1. Pour introduire la question de la polyphonie	59
2.3.2. Problèmes et enjeux de l'analyse polyphonique	62
- La définition de l'instance à la source des points de vue.....	62
- La définition du point de vue	64
- Le problème du niveau d'analyse	65
2.3.3. L'analyse du point de vue et de la polyphonie dans la théorie scandinave de la polyphonie linguistique	67
- Le cadre général d'analyse de la ScaPoLine.....	67
- Les points de vue et l'instanciation de leur source	69
- Les êtres discursifs et la saturation des points de vue dans le texte.....	71
2.3.4. L'analyse du point de vue et des postures énonciatives chez A. Rabatel	75
- L'analyse du point de vue et de sa prise en charge chez A. Rabatel	75
- Relations entre points de vue et construction de L1/E1 dans la topique énonciative	79
2.3.5. Quel cadre pour l'analyse de la polyphonie ?	83
- Les fondements de l'analyse	83
- Le point de vue et sa prise en charge ou la question de la modalisation	84
- La polyphonie comme dédoublement des points de vue	87
2.4. Bilan. La polyphonie au niveau textuel : la notion de figure de l'auteur.....	89

Chapitre 3

Problématique et méthodologie de la recherche.	93
3.1. Le corpus.....	95
3.1.1. Un corpus de textes du même genre : critères de sélection des textes.....	96
- Des articles de recherche français.....	97

- Le genre en synchronie	97
- La dimension éditoriale du genre.....	97
<i>Des textes publiés</i>	98
<i>Des textes publiés en tant qu'articles de revue</i>	98
<i>Des articles publiés dans des revues « à comité »</i>	99
- Une identité énonciative	99
3.1.2. Le contrôle de la variabilité du genre : diversification des textes et définition de variables externes	100
- La diversification des revues, des numéros et des auteurs.....	100
- La variable « discipline » : sciences du langage <i>versus</i> lettres	101
- La variable « statut de l'auteur » : doctorant <i>versus</i> non-doctorant.....	103
3.1.3. La composition du corpus	105
3.1.4. Acquisition des données et pré-traitement du corpus.....	107
3.2. Une analyse quanti-qualitative partiellement automatisée	109
3.2.1. L'analyse automatique et quantitative.....	110
3.2.2. Le recours aux statistiques	114
- Principes de l'analyse statistique	115
<i>Point de départ</i>	115
<i>Statistiques descriptives</i>	116
<i>Statistiques inférentielles</i>	117
- L'interprétation des résultats de l'analyse statistique	120
3.2.3. Visualisations graphiques.....	122
- Des histogrammes pour la distribution des variables nominales.....	122
- La comparaison d'une variable continue entre deux groupes (groupes « discipline » et « statut de l'auteur ») : histogramme pour les moyennes et distributions en fréquence	123
3.2.4. L'analyse du genre dans un va-et-vient entre échantillon de données et textes.....	127
3.3. Les traits textuels retenus.....	130
3.3.1. Les 97 variables Cordial.....	130
3.3.2. Le lexique : restriction au lexique transdisciplinaire	133
3.3.3. Les traits analysés manuellement.....	133
Chapitre 4	
Modèles de textes et rapports contrastés au lecteur et à l'objet	137
4.1. Le volume du corps du texte et des notes : premier indice d'une diversité insoupçonnée du genre.....	138
4.1.1. Volume total : proximité inter-disciplinaire et disparité intra-disciplinaire.....	138
4.1.2. Contribution du corps du texte et des notes au volume total : différence disciplinaire dans l'usage des notes	140
4.1.3. La variabilité entre revues	142
4.1.4. La variabilité liée au statut de l'auteur	145
4.2. Des espaces péri-textuels investis différemment dans les deux disciplines	146
4.2.1. Texte et péri-texte : introduction	146
4.2.2. Résumés et abstracts.....	147
4.2.3. Titre et signature.....	148
4.2.4. Annexes.....	149

4.2.5. Notes.....	150
4.2.6. Bibliographies	152
4.2.7. Une alternative disciplinaire entre notes et bibliographies dans la gestion des références	154
4.2.8. Bilan sur le péritexte de l'article : un contraste disciplinaire dans les degrés d'éclatement du texte	157
4.3. L'irréductible diversité des plans de texte : vers un continuum entre deux tendances opposées et disciplinairement marquées	158
4.3.1. Quelle grille pour l'analyse des plans de texte ?.....	158
4.3.2. Pratique des titres et de la numérotation : la division du texte en composantes structurelles	159
4.3.3. Un continuum entre deux modèles de texte opposés	162
4.3.4. Structuration de plus bas niveau, interne au paragraphe	164
4.4. Pluri-sémioticit� de l'article et identit�s disciplinaires des mod�les de texte.....	165
4.4.1. Les formalismes observables dans le corps des articles.....	165
4.4.2. Formalismes et mod�les de texte : le contraste disciplinaire	168
4.4.3. Bilan sur les mod�les de texte et les identit�s disciplinaires qu'ils manifestent	170
4.5. Les exemples et les citations : pratiques textuelles et �pist�mologies	172
4.5.1. Le poids des exemples et citations dans les textes	173
- Poids des exemples	173
- Poids des citations.....	174
- Un poids relatif des exemples par rapport � celui des citations ?	176
4.5.2. Sources documentaires et empiriques : pistes pour l'analyse	177
- Premieres remarques formelles.....	177
- Dans les articles de LET : l'intertexte comme mat�riau et comme objet	178
- Dans les articles de SCL : partage strict entre exemple et citation et diversit� des exemples.....	180
- Pistes pour l'analyse de la notion d'exemple.....	182
4.6. Macro-analyse du genre : bilan	184

Chapitre 5

�l�ments pour une caract�risation linguistique du genre	187
5.1. Les aspects morpho-syntaxiques.....	188
5.1.1. Le d�coupage du texte en phrases.	189
5.1.2. La syntaxe de la phrase	192
5.1.3. Les temps verbaux.....	196
- Les temps du pass�.....	198
- Le futur et l'imp�ratif.....	204
5.1.4. Les personnes linguistiques.....	208
5.2. Les ponctuations.....	213
5.2.1. Nombre de ponctuations dans le texte.....	214
5.2.2. Aper�u g�n�ral du type de ponctuations en usage	215

5.2.3. Des tendances contrastées dans l'usage des ponctuations, liées à la structure des textes et à des préférences idiomatiques.....	217
- Marqueurs de pause : virgule et points-virgules.....	217
- Les deux-points.....	218
- Les parenthèses.....	219
5.2.4. Les points de suspension et le point d'exclamation : des emplois rares... et peut-être insoupçonnés !.....	221
- Les points de suspension.....	221
- Le point d'exclamation.....	223
5.2.5. Le point d'interrogation : question rhétorique ou rhétorique de la découverte ?.....	226
5.2.6. Les guillemets : l'importance d'un « simple » signe dans le genre.....	232
- En guise d'introduction : la complexité des guillemets.....	232
- Le poids des guillemets.....	233
- Les emplois autonymiques.....	234
- De l'exemple à la citation.....	234
- Nommer et catégoriser.....	236
- Marquer sa distance.....	237
- Bilan sur les guillemets.....	239
5.3. Le lexique transdisciplinaire.....	240
5.3.1. Un lexique au service de la construction du savoir.....	241
5.3.2. Un lexique entre langue générale et langue de spécialité.....	242
5.3.3. Constitution d'une liste de termes transdisciplinaires.....	244
5.3.4. Un plus grand nombre d'occurrences tous termes confondus en SCL.....	246
5.3.5. Des termes inégalement employés en LET et en SCL.....	247
5.3.6. Concordances des termes transdisciplinaires : de quelques traits communs et des spécificités disciplinaires.....	249
- La diversité des emplois : les usages spécifique, général et transdisciplinaire.....	249
<i>Un usage conceptuel</i>	249
<i>L'usage transdisciplinaire comme spécification de l'usage général</i>	250
<i>Un usage transdisciplinaire propre aux articles de LET, dédié à l'œuvre littéraire, sa production et sa réception</i>	251
- Tendance marquée à la systématisation de l'usage des termes transdisciplinaires en SCL <i>versus</i> en LET.....	253
5.3.7. Bilan sur les usages disciplinaires des termes transdisciplinaires : des épistémologies plurielles.....	255
5.3.8. Des différences liées au statut de l'auteur?.....	258
5.4. Des caractéristiques attendues et figées, et des usages contrastés : bilan sur la morpho-syntaxe, la ponctuation et le lexique du genre.....	262

Chapitre 6

Le genre et ses usages singuliers, les styles :	
une entrée pour l'analyse de la diversité interne du genre.....	263
6.1. Le style en sciences : une question occultée.....	263
6.2. Une définition du style relative aux genres.....	266

6.3. Méthodologie pour l'étude des styles.....	268
6.3.1. Le corpus : un corpus « genre » et un corpus « auteurs »	268
6.3.2. Les traits textuels retenus	270
6.3.3. Analyses et méthodes statistiques	271
6.4. De quoi sont faits les styles ? Les traits stylistiques discriminants.....	273
6.4.1. Un biais dans l'analyse des styles	274
6.4.2. Style et lexique : une entrée pour l'étude des concepts et des épistémologies	275
6.4.3. Style et morphosyntaxe	277
- Des usages du genre marqués par le type d'études menées par l'auteur	277
<i>Des textes comportant une dimension historique.....</i>	<i>277</i>
<i>Des textes diversement exemplifiés</i>	<i>278</i>
- Des styles au sens de spécificités idiolectales.....	280
<i>Les ponctuations, la complexité syntaxique et les relations logiques.....</i>	<i>280</i>
<i>Les aspects énonciatifs : personne et non-personne, images du scripteur et rapport au</i>	
<i>lecteur.....</i>	<i>282</i>
<i>La dimension argumentative des textes.....</i>	<i>283</i>
6.4.4. Styles et lexique transdisciplinaire	285
6.4.5. Bilan sur les traits stylistiques.....	286
6.5. À quel degré un style se distingue-t-il ? L'homogénéité des textes d'un même auteur	
et leur singularité au sein du corpus.....	287
6.5.1. De l'homogénéité et de la singularité des styles d'auteur	288
6.5.2. Validation de la distinctivité des styles par la méthode Bootstrap.....	292
6.6. Retour au genre et à la question de sa stabilité	294
6.7. Conclusion sur la variation stylistique du genre de l'article.....	299

Chapitre 7

La subjectivité et la polyphonie de l'article.

Pour une approche énonciative de la rhétorique scientifique.....	301
7.1. Introduction : l'article comme type énonciatif de texte.....	302
7.2. Les emplois de « je » et « nous » ou l'irréductible polyphonie de la première	
personne linguistique.	304
7.2.1. L'usage autonymique des marques de personne	305
7.2.2. L'usage conceptuel du « je » et du « moi ».....	305
7.2.3. Le « je » comme marque personnelle en usage générique	306
7.2.4. « Je » et « nous » comme marques personnelles qui renvoient à l'auteur	308
7.2.5. La question de l'inclusion du lecteur dans le « nous ».....	310
7.2.6. Le « nous » comme collectif : le sujet épistémique et culturel	313
7.2.7. Bilan : la multiplicité interprétative de « je » et de « nous ».....	315
7.3. Entre effacement énonciatif et prise en charge personnelle.....	316
7.3.1. Le « il » impersonnel.....	317
- Les occurrences de « il » et le taux de « il » impersonnel	317
- Les constructions les plus fréquentes avec le « il » impersonnel	318

7.3.2. Des profils d'articles en termes d'effacement et de marquage de la personne linguistique	322
- « Je » ou « nous » : quels modes d'auto-désignation de l'auteur dans son texte ?	322
- Tendance à l'abondance <i>versus</i> tendance à la restriction de marques personnelles. ...	325
<i>Usage du « on »</i>	327
<i>« Il » impersonnel et passif</i>	328
7.4. La représentation de la « circulation des discours » dans l'article.....	331
7.4.1. Une quête de traçabilité dans le renvoi aux sources.....	331
7.4.2. La multiplication des références : la tentation de l'exhaustivité ?	336
7.4.3. Identifier les voix en présence : l'attribution énonciative des discours représentés	340
7.4.4. Modes de désignation des points de vue attestés dans le champ, ou comment l'univers de référence est-il cartographié ?	343
- Typologie des modes de désignation des points de vue attestés dans le champ de recherches.....	344
- Référence au document et référence à un auteur : communauté scientifique <i>versus</i> communauté savante ?	347
- Une vue surplombante sur le champ fournie par les noms de courant	349
- Des modes de désignation des points de vue au service d'un enjeu de démarcation ? 350	
7.5. Des objectifs et une axiologie au service de la légitimation de la recherche	352
7.5.1. La définition de l'objet et de l'objectif de l'étude dans l'introduction	353
- Se référer à l'existant	354
- Un enjeu de démarcation	357
- Une centration sur l'objet et son intérêt intrinsèque ?	360
7.5.2. Les paradigmes axiologiques récurrents dans l'évaluation des points de vue et des approches.....	362
- Des valeurs propres à l'activité de recherche, et communes aux deux disciplines considérées	363
- La valeur de scientificité dans les articles de SCL	366
- Matérialité et portée littéraire du texte étudié en LET	368
7.6. Bilan : surénonciation et argumentation.....	370
Conclusion.....	373
Références	378
Index	393

Sommaire du tome II. Annexes

Avant-propos	1
Annexes 1. Composition du corpus.....	3
Annexes 1.1. Articles de Lettres	3
Annexes 1.2. Articles de Sciences du langage	11

Annexes 2. Macro-analyse du genre : volume et composantes structurelles (cf. chapitre 4)	19
Annexes 3. Phrases, ponctuations, lexique (cf. chapitre 5)	25
Annexes 3.1. Nombre de phrases et longueur	25
Annexes 3.2. Composantes de la phrase	58
Annexes 3.3. Temps verbaux	72
Annexes 3.4. Marques de personne et de non-personne	88
Annexes 3.5. Ponctuations	113
Annexes 3.6. Lexique transdisciplinaire	127
Annexes 4. Les styles d'auteur (cf. chapitre 6)	219
Annexes 4.1. Composition du corpus « auteurs »	219
Annexes 4.2. Liste des étiquettes morpho-syntaxiques de Cordial et TnT	226
Annexes 4.3. Caractéristiques des 15 styles d'auteur	233
Annexes 5. Caractérisation énonciative du genre (cf. chapitre 7)	287
Annexes 5.1. Marques personnelles et tours impersonnels	287
Annexes 5.2. Modes de désignation de l'existant et enjeu d'opposition dans les articles de SCL	294

Table des figures

Les différentes représentations (tableaux, graphiques, schémas) sont désignées comme figures, et numérotées de la manière suivante : le premier chiffre correspond au chapitre où elles sont insérées, le second à leur ordre d'apparition dans le texte.

Figure 2.1. L'analyse de la polyphonie linguistique chez O. Ducrot, dans la ScaPoLine et chez A. Rabatel.....	84
Figure 3.1. Articles de doctorants et de non-doctorants dans le corpus LET : poids en nombre d'articles.....	105
Figure 3.2. Articles de doctorants et de non-doctorants dans le corpus SCL : poids en nombre d'articles.....	105
Figure 3.3. Représentation des revues dans le corpus de LET et de SCL en nombre de numéros et en nombre d'articles.....	106
Figure 3.4. Exemple d'étiquetage d'une phrase dans Cordial Professionnel, cité dans le Manuel Utilisateur (Synapse Développement, 2002).	111
Figure 3.5. Modélisation des variables.....	119
Figure 3.6. Procédures statistiques mises en œuvre.....	120
Figure 3.7. Exemple d'histogramme dans sa forme élémentaire.....	122
Figure 3.8. Exemple de boîte à moustache dans sa forme élémentaire (N. Hunt, 1996).	124
Figure 3.9. Exemple de distribution en fréquence d'une variable continue.....	125
Figure 3.10. Liste des 97 variables Cordial.....	133
Figure 3.11. Liste des traits analysés manuellement.....	134
Figure 4.1. Distribution en fréquence du nombre de mots par article : comparaison disciplinaire.....	139
Figure 4.2a. Contribution du corps et des notes au volume total : articles de LET.....	140
Figure 4.2b. Contribution du corps et des notes au volume total : articles de SCL.....	140
Figure 4.3. Distribution en fréquence du volume des notes : comparaison disciplinaire.....	141
Figure 4.4a. Volume moyen des articles dans les revues de LET.....	143
Figure 4.4b. Volume moyen des articles dans les revues de SCL.....	143
Figure 4.5. L'usage des annexes dans les deux disciplines.....	150
Figure 4.6. Distribution en fréquence du nombre de notes : comparaison disciplinaire.....	151
Figure 4.7. L'usage des bibliographies dans les deux disciplines.....	153
Figure 4.8. Les stratégies d'exploitation des espaces textuels pour la gestion des références.....	154
Figure 4.9. Les introductions dans les deux disciplines : titrées (cas 1), identifiables (cas 2), non identifiables (cas 3).	159
Figure 4.10. Les conclusions dans les deux disciplines : titrées (cas 1), identifiables (cas 2), non identifiables (cas 3).....	160
Figure 4.11. Le titrage des sections dans les deux disciplines.....	160
Figure 4.12. La numérotation des sections dans les deux disciplines.....	161

Figure 4.13. L'usage des sous-parties dans les deux disciplines.	161
Figure 4.14. La structure des articles : continuum des modèles et comparaison disciplinaire.	163
Figure 4.15. L'usage des indices de structuration internes au paragraphe dans les deux disciplines.	165
Figure 4.16. Corrélations entre structure et formalismes : deux modèles de textualité opposés, des identités disciplinaires contrastées.	170
Figure 4.17. Distribution en fréquence du poids des exemples : comparaison disciplinaire.	173
Figure 4.18. Distribution en fréquence du poids des citations : comparaison disciplinaire... ..	174
Figure 5.1. Distribution en fréquence du nombre moyen de mots par phrase : comparaison disciplinaire.	189
Figure 5.2. Distribution en fréquence du nombre moyen de mots par phrase : comparaison des deux groupes d'auteurs.	191
Figure 5.3. Taux moyen de phrases comportant une subordonnée dans le corps, le corps sans les exemples ni les citations, et les notes : comparaison entre les groupes.	192
Figure 5.4. Des tendances contrastées dans l'usage des mots-outils et des mots signifiants.	195
Figure 5.5. Taux moyen des principaux temps verbaux à l'exclusion du présent : comparaison entre les groupes.	197
Figure 5.6. Une corrélation négative dans l'usage du présent, du passé simple et de l'imparfait en LET.	199
Figure 5.7. Des tendances contrastées dans l'usage des temps verbaux : bilan.	208
Figure 5.8. Taux moyen des personnes verbales : comparaison entre les groupes.	210
Figure 5.9. Le taux de <i>je</i> et de <i>nous</i> dans chaque article : comparaison entre les disciplines et les groupes d'auteurs.	211
Figure 5.10. Distribution en fréquence du nombre de ponctuations par article (fréquence relative) : comparaison disciplinaire.	214
Figure 5.11. Distribution en fréquence du nombre de ponctuations par article (fréquence relative) : comparaison des deux groupes d'auteurs.	215
Figure 5.12. Taux moyen de chaque type de ponctuation : comparaison entre les groupes.	216
Figure 5.13. Liste des 100 termes transdisciplinaires retenus.	245
Figure 5.14. Distribution en fréquence du nombre de termes transdisciplinaires par article (tous termes confondus) : comparaison disciplinaire.	247
Figure 5.15. Liste des 45 termes significativement plus fréquents dans les articles de SCL.	248
Figure 5.16. Les usages contrastés du lexique épistémique : l'effet du statut de l'auteur.	259
Figure 5.17. Les usages contrastés du lexique épistémique : l'effet du statut de l'auteur dans les articles de LET.	259
Figure 5.18. Les usages contrastés du lexique épistémique : l'effet du statut de l'auteur dans les articles de SCL.	260
Figure 6.1. Vue synoptique de l'analyse des styles et de la diversité interne du genre.	273
Figure 6.2. Styles d'auteur et lexique.	276
Figure 6.3. Liste des termes transdisciplinaires discriminant le style de A15.	285
Figure 6.4. Scores d'homogénéité et de distinctivité des styles en fonction des jeux de variables entrant dans la classification.	289
Figure 6.5. Singularité et homogénéité des styles d'auteur.	290
Figure 6.6. Classification des textes des différents auteurs dans DTM, à partir des variables TnT et du lexique transdisciplinaire.	291

Figure 6.7. Ellipses de confiance autour des 15 auteurs observés sur le premier plan factoriel	292
Figure 6.8. Ellipses de confiance autour des 15 auteurs observés et de la variante « autre » (tous auteurs du corpus « genre » confondus) sur le premier plan factoriel	293
Figure 6.9. Positionnement des variables sur les deux premiers axes factoriels – corpus « genre »	295
Figure 6.10. Positionnement des variables sur les deux premiers axes factoriels – corpus « auteurs »	297
Figure 7.1. <i>Il</i> et <i>Il</i> impersonnel dans les deux disciplines.	317
Figure 7.2. Les constructions les plus fréquentes du <i>il</i> impersonnel en contexte droit.....	319
Figure 7.3. Les autres constructions fréquentes du <i>il</i> impersonnel en contexte droit.	319
Figure 7.4. L’auto-désignation de l’auteur dans son texte	325
Figure 7.5. Tendances à la restriction et à l’abondance des marques de personne :	326
les formes de <i>je</i> et de <i>nous</i>	326
Figure 7.6. Modes de désignation des points de vue attestés dans le champ (articles de SCL).	346

Introduction

L'étude que nous présentons ici trouve son point de départ dans un questionnement sur l'écriture de recherche. Elle poursuit un double objectif : caractériser les textes et contribuer à la construction d'un cadre pour l'analyse des pratiques d'écriture, fondé sur la notion de genre et, secondairement, sur son approche énonciative et polyphonique, avec la notion de figure de l'auteur. Nous nous centrons sur le genre de l'article de recherche en français, dans deux disciplines de sciences humaines, les sciences du langage et les lettres. Nous décrivons les régularités et la variabilité observables dans un corpus d'articles et proposons ce faisant des pistes pour l'analyse des genres, des textes, et des unités linguistiques. La description voudrait aussi, à terme, nourrir une réflexion sur la didactisation de ce genre.

Notre interrogation sur les pratiques d'écriture de recherche s'est articulée autour de la notion de genre, qui représente une voie particulièrement stimulante, tant dans ses fondements théoriques que dans les études empiriques qu'elle appelle. Elle trouve avec la linguistique de corpus un outil adapté pour une description des textes dans leurs régularités et dans leur diversité, et ouvre des pistes pour aborder la langue dans ses usages textuels, en lien avec les pratiques où les textes sont produits et interprétés.

Exploratoire et empirique, notre investigation du genre multiplie les traits linguistiques pris en compte ; elle exploite les possibilités offertes par l'informatique pour mettre au jour les caractéristiques du genre par une approche quantitative et statistique. Celle-ci permet l'étude de la variation, et notamment des styles en tant qu'usages singuliers que chaque auteur fait d'un genre. L'observation qualitative est cependant essentielle, pour analyser les textes dans leur matérialité, construire des observables et interpréter les traits décrits en référence à leur fonctionnalité.

Cette méthodologie « quanti-qualitative », qui fait sans doute l'originalité de notre étude du genre, nous conduit à traiter à la fois de grandes masses de données numériques, et d'exemples tirés du corpus, représentatifs ou plus singuliers. Menée à partir d'un corpus principal de 220 articles, elle répond à la nécessité de concilier une description globale du genre et une description fine de ses caractéristiques, tout en mettant l'accent sur la diversité des profils de textes.

Un second axe de réflexion est venu, avec le genre, guider notre recherche, et concerne les approches énonciatives et leur intérêt pour l'analyse des textes. Sans prétendre

qu'il s'agisse de la seule entrée à privilégier, nous défendons l'idée que la problématique énonciative peut servir à décrire la dimension argumentative et rhétorique des genres. La question de la polyphonie linguistique, particulièrement travaillée et débattue à l'heure actuelle, nous a interpellée, pour mieux comprendre un enjeu des genres, déjà signalé par M. Bakhtine et dans l'approche aristotélicienne de la notion : l'idée que le genre impose un ethos spécifique, une image de son scripteur, ou encore, dans les termes que nous avons retenus, une figure de l'auteur.

Nous cherchons à asseoir linguistiquement la notion, qui doit permettre d'étudier ensemble et au niveau du texte les « marques de soi et de l'autre » analysées comme telles par les approches énonciatives. D'un point de vue théorique, nous discutons l'approche polyphonique et les niveaux d'analyse pour concilier la prolifération des points de vue et la construction d'un point de vue d'auteur cohérent. D'un point de vue empirique, nous proposons des pistes pour une analyse énonciative et polyphonique du genre qui mobilise la syntaxe, la sémantique et la pragmatique, et qui ouvre des perspectives pour aborder depuis les textes leur rôle central dans la définition des communautés et des pratiques où ils sont en usage.

Notre étude se centre sur le genre particulier de l'article de recherche, en tant que genre central dans la construction, l'évaluation et la diffusion des savoirs scientifiques. S'il ne s'agit pas d'un genre didactique, il pose cependant la question de savoir comment aider les doctorants à en maîtriser les normes, pour favoriser leur entrée dans le champ.

La description du genre représente un atout, pour affiner voire refonder ses caractéristiques supposées, et développer une conscientisation des pratiques. La didactisation du genre représente un horizon plutôt qu'un objectif de ce travail, mais la description des textes peut en effet servir de préalable pour s'interroger sur un modèle de référence du genre ; dans cette perspective, il convient de tenir compte de la spécificité des pratiques des néophytes, pour éviter de fonder l'objet d'enseignement et d'apprentissage sur le seul modèle expert.

L'écrit de recherche fait l'objet de nombreux travaux, dans des champs diversement désignés : celui des discours professionnels ou des discours spécialisés, celui du discours académique, bien exploré dans une perspective didactique notamment, et abordé sous deux aspects en contexte français, l'écrit de recherche et l'écrit dans l'enseignement supérieur. La dénomination retenue, celle de l'article en tant qu'écrit de recherche, met l'accent sur le domaine d'activité, et sur l'importance que joue l'écrit dans la recherche, par sa fonction cognitive ou heuristique. Nous reprenons également les termes souvent usités de discours et de genres scientifiques, qui envisagent le texte au titre de production symbolique et culturelle, concerné en l'occurrence par la production de savoir.

Dans ce contexte, la littérature existante montre que le français manque de descriptions qui portent sur les genres, sur des disciplines variées, et sur celles qui sont sociologiquement minorées, et qui n'ont pas à affronter au même degré la domination de l'anglais dans la communication scientifique. Le choix de l'article de recherche en sciences humaines place la réflexivité au cœur de la présente étude ; il s'explique prioritairement par l'importance de connaître un tant soit peu les domaines considérés pour pallier les difficultés d'analyse rencontrées dans le cas inverse. Il a en outre ceci d'intéressant que les disciplines qui ont affaire à des signes et du texte dans leurs objets posent de manière centrale le problème de la polyphonie dans l'écrit de recherche.

L'idée qu'on n'écrit pas la science de la même manière selon les disciplines est désormais bien établie. La comparaison disciplinaire représente une entrée importante en didactique, et fait écho aux travaux anglo-saxons sur l'English For Specific Purposes, où elle vise la mise au jour d'identités culturelles, avec les identités nationales et linguistiques. Elle

concerne généralement des disciplines éloignées, et notre approche de l'article en sciences du langage et en lettres se révèle à ce titre singulière. Si elle pose des difficultés méthodologiques, elle est à comprendre comme une heuristique : la comparaison disciplinaire vise à mettre en relief les caractéristiques des textes, et sert de point de départ à l'analyse de leur diversité, entre les disciplines et au sein de chacun des deux champs considérés. Elle rejoint en cela deux autres aspects de notre étude qui donnent prise sur la variabilité du genre, l'étude de la spécificité des textes des doctorants, et celle des styles d'auteurs dans l'article.

Les trois premiers chapitres s'intéressent à la construction de l'objet de recherche. La notion de genre fait l'objet de notre premier chapitre, qui propose d'analyser en quoi elle représente une entrée privilégiée en linguistique. A partir des écrits de M. Bakhtine, et en référence aux travaux plus récents sur les discours et les textes, nous partons de l'idée d'une interdépendance entre les textes et les pratiques sociales où ils sont produits et interprétés. Nous envisageons alors les genres du côté de la production verbale, pour analyser, avec la notion de dialogisme, en quoi ils conditionnent et régulent nos pratiques langagières. Les approches empiriques des genres et de l'article en particulier permettent au final de spécifier comment aborder le genre dans notre étude.

Dans les analyses programmatiques ou empiriques des genres et/ou des écrits de recherche, l'importance de notions comme l'ethos, le rôle, l'image du scripteur sert de point de départ au chapitre 2. Celui-ci introduit la notion de figure de l'auteur, en tant que figure du genre, et cherche à en définir les fondements en questionnant les approches énonciatives en linguistique. Les études sur la polyphonie linguistique retiennent notre attention, avec pour but de voir en quoi elles peuvent servir une analyse des textes et de la construction d'un point de vue d'auteur. Les questions qu'elles soulèvent fondent un programme vaste de recherche, dont la présente étude représente une première exploration.

Le chapitre 3 présente la problématique et la méthodologie de la recherche. L'étude du genre appelle une linguistique de corpus et une analyse à la fois quantitative et qualitative, et partiellement automatisée. Nous traitons des critères de sélection des textes, du contrôle de leur diversité, et des variables « discipline » et « statut de l'auteur ». La variable « style » qui intéresse également notre propos demande un protocole spécifique, détaillé pour cette raison dans le chapitre qui lui est consacré (chapitre 6.). Nous explicitons les ressorts de l'analyse dans ses différentes modalités et les choix des représentations graphiques, et terminons par la liste des traits retenus pour l'analyse des textes, à laquelle sont dédiés les quatre chapitres suivants.

Le chapitre 4 part du constat que tout texte est fait de plusieurs textes et de plusieurs systèmes sémiotiques, et propose une macro-analyse du genre de l'article en traitant de ses caractéristiques structurelles (le volume, l'exploitation du péri-texte, les plans de texte, les formalismes, les exemples et les citations). Il intéresse particulièrement la comparaison disciplinaire et la mise au jour de la diversité des modèles de textes, interprétés en termes de rapport au lecteur et à l'objet.

Le chapitre 5 fournit des éléments de caractérisation linguistique du genre en s'appuyant sur les résultats de l'analyse automatique de variables morpho-syntaxiques et lexicales. En regard des caractéristiques supposées du texte scientifique, il décrit la structure générale de la phrase, les temps verbaux, les marques de personne et de non-personne, se centre alors sur les ponctuations, notamment celles à valeur énonciative, et s'intéresse enfin

au lexique transdisciplinaire, spécifique à l'activité de recherche mais diversement mobilisé dans les deux champs considérés.

La description du genre dans ses régularités et sa variabilité linguistiques se prolonge au chapitre 6, qui s'intéresse aux styles, ou usages singuliers que chaque auteur fait de l'article de recherche. L'étude, menée en collaboration avec C. Poudat, requiert un corpus et une méthodologie spécifiques, et ne porte que sur l'article de sciences du langage. La caractérisation du style de 15 auteurs nous permet de décrire de quoi sont faits les styles et d'évaluer leurs degrés de distinctivité par rapport aux normes du genre. La variation stylistique du genre sert à discuter la diversité interne de ce dernier, et la question de sa stabilité.

Le chapitre 7 restreint le nombre de textes pris en compte et consiste essentiellement en des analyses qualitatives. Il cherche à cerner la subjectivité et la polyphonie caractéristiques du genre, et prend comme point de départ des marques traditionnellement décrites en linguistique de l'énonciation (pronoms *je* et *nous*, tours impersonnels, subjectivèmes, points de vue autres, etc.) ; il propose par ailleurs de les mettre en relation dans l'analyse de profils de textes, et d'entrées originales (la représentation de la circulation des discours, la légitimation de la recherche présentée), et fournit ainsi des pistes pour une analyse rhétorique fondée sur les aspects énonciatifs du genre.

En guise de bilan, nous résumons brièvement les apports de l'étude, notamment sur les identités disciplinaires observées dans les textes, en termes de figure de l'auteur, et revenons sur les difficultés posées par la spécificité des articles de doctorants dans la perspective d'une didactisation du genre.

Chapitre 1.

Aborder les pratiques d'écriture sous l'angle de la problématique des genres

Dans la science, au même titre que dans les autres domaines de l'activité humaine, les échanges langagiers s'inscrivent dans des genres. Héritée de la rhétorique antique, la notion perdure et se révèle aujourd'hui centrale en linguistique, où elle est convoquée depuis les années 1970 en référence notamment aux travaux de M. Bakhtine. La linguistique textuelle et la sémantique interprétative, l'analyse du discours et la rhétorique, la classification automatique de textes en traitement automatique des langues, et la didactique des langues et de l'écrit en particulier sont autant de champs qui revendiquent l'intérêt de la notion.

S'il ne s'agit pas de rendre compte de l'ensemble des approches qui ont traité des genres, il importe d'en définir les fondements, notamment la question de leur assise textuelle et celle de l'interdépendance entre les textes et les pratiques où ils sont produits et interprétés. L'analyse de la notion de genre du côté de la production verbale et de sa dimension dialogique permet de comprendre en quoi les genres conditionnent nos pratiques discursives, et de montrer leur intérêt pour l'analyse des textes dans leurs régularités et dans leur diversité, en tant qu'objets sociaux, constitutifs des communautés discursives et de l'action et l'activité humaines.

La problématique des genres telle que nous en explorons ici les fondements s'appuie sur les écrits jugés fondateurs de M. Bakhtine, et les confronte à d'autres approches plus actuelles pour avancer vers une définition opératoire des genres et déterminer comment analyser empiriquement un genre, et le genre de l'article en particulier.

Nous commencerons par une mise au point terminologique, qui pose les jalons de l'objectif poursuivi ici, celui d'une étude des pratiques discursives sous l'angle de la problématique des genres de textes.

1.1. Mise au point terminologique : genre, texte et discours

Un premier constat s'impose : les définitions de la notion de genre sont variées, ce qui reflète la diversité des approches qui convoquent cette notion, et la diversité de leurs fondements théoriques.

La dénomination de la notion n'est pas homogène. Ainsi coexistent les termes de « genres de discours » – depuis la tradition aristotélicienne des *genera dicendi* – et de « genres de textes ». J.M. Adam (1999) oppose genres et types de textes, quand J.P. Bronckart

(1996) oppose genres de textes et types de discours, alors que D. Maingueneau (1998, par exemple) traite quant à lui de types de textes et de genres de discours. Il ne s'agit pas de détailler ici les positionnements théoriques sous-tendant les dénominations proposées par chacun des auteurs cités. La diffusion des travaux de M. Bakhtine n'est pas exempte de ce problème : on rencontre dans son œuvre et celle de son cercle un certain nombre de fluctuations terminologiques, que l'on peut imputer à l'évolution interne des travaux du cercle mais aussi sans doute à des problèmes de traduction. Ainsi, dans les différentes traductions, quoique le syntagme « genre de discours » l'emporte, le choix entre « genres du discours » et « genres de texte » n'est pas tout à fait stabilisé.

1.1.1. Texte et discours

Cependant, loin de ne concerner que la question du genre, l'utilisation conjointe voire concurrente des deux termes constitue un fait ordinaire en linguistique. Cela oblige donc à en expliquer les ressorts afin de situer nos propres options terminologiques et conceptuelles.

Avant d'examiner le rapport entre texte et discours, il convient de souligner que l'importance du terme de discours depuis une trentaine d'années témoigne de ce qui est qualifié de « tournant » pris par la linguistique (François, 1980) ; ce tournant, marqué par le déclin des approches structuralistes et par l'influence des conceptions pragmatiques du langage, s'affirme sous la forme d'une première dichotomie, celle qui oppose langue et discours. Le discours est ainsi conçu comme une nouvelle dimension d'analyse, qui ne va pas sans rappeler le projet d'une linguistique de la parole défendu par F. de Saussure¹ : l'enjeu consiste à dépasser le cadre circonscrit par les approches structuralistes, centrées sur l'objet « langue », pour envisager la question de l'utilisation de la langue ; cette dernière fait l'objet d'approches très diverses, mais se présente comme le point commun d'une orientation pragmatique, intéressée par les interactions entre individus, d'une orientation énonciative, centrée sur l'actualisation de la langue en discours, et d'une orientation sociolinguistique et ethnographique, qui envisage les communautés de locuteurs dans leur dimension linguistique.

Parallèlement, une autre tendance se profile en linguistique ; en prenant pour objet d'étude la langue, les approches structuralistes ne se sont guère aventurées au-delà du cadre phrastique. Or, c'est justement dans la volonté de sortir de ce cadre que la *Textlinguistik*², qui apparaît vers la fin des années 1960 dans le domaine germanique et anglo-saxon, trouve ses fondements. La question des unités transphrastiques, qui définit ces approches nommées « grammaire de texte », puis « linguistique textuelle », s'associe à celle de la double structuration du texte par la langue et sa mise en discours.

Ces préoccupations concernent également l'analyse du discours, courant contemporain aux approches textuelles, mais qui, fortement influencé pour sa part par la sociologie, procède de manière diamétralement opposée en s'intéressant prioritairement aux conditions de production et d'interprétation du discours (Guespin, 1971 : 10) ; ce dernier est abordé en tant que dispositif communicatif relevant de ce qui est alors traité en termes de « formation discursive » (Foucault, 1969a, Haroche, Henry et Pêcheux, 1971). Explicitement althusserien au départ, le courant se donne pour objet l'analyse de ces formations discursives en tant qu'elles sont déterminées par des formations sociales idéologiques. L'analyse a donc pour but

¹ C'est ce que montrent ses *Écrits de linguistique générale* (2001), *a contrario*, comme le soulignent ses éditeurs S. Bouquet et R. Engler, de la réinterprétation de la pensée saussurienne fournie dans le *Cours de Linguistique générale* (1995, 1^{ère} édition 1916).

² Cf. par exemple les écrits de E. Coseriu (notamment 1981).

de rechercher dans les textes les traces d'une idéologie, et les interprète comme une prise de position sociale et politique³.

Avec ces deux courants se pose alors la question de l'opposition entre « texte » et « discours ». Pour préciser ce que sous-tend cette distinction terminologique, il convient, à l'instar d'un certain nombre de théoriciens en linguistique textuelle et en analyse du discours, de renoncer d'abord à une définition établie à partir de l'idée que la textualité renvoie à l'écrit ; dans cette optique, le choix du terme discours serait préféré pour rendre compte des pratiques de communication orales. Mais la textualité ne saurait être assimilée à la scripturalité : la question de l'organisation textuelle se pose aussi bien lorsqu'elle est produite dans l'ordre de l'oral que dans celui du scriptural.

Les chercheurs évoqués plus haut, J.M. Adam, J.P. Bronckart, et D. Maingueneau, abstraction faite des différences dans l'appareil terminologique et conceptuel qu'ils proposent, définissent le texte et le discours en accord avec la distinction fondatrice proposée par D. Slatka (1975) : celui-ci conceptualise le texte par opposition au discours comme objet formel abstrait de ses conditions de production et d'interprétation ; pour le dire autrement, le discours est conçu comme l'inclusion d'un texte dans son contexte⁴, c'est-à-dire dans ses conditions de production ou d'interprétation (Adam, 1999 : 39).

Le texte et le discours correspondent ainsi à deux perspectives différentes (Bronckart, 1996), ou encore à « deux modes d'appréhension des unités transphrastiques » (Maingueneau, 2001 : 186), donc à deux types d'approches d'un objet commun. Une approche de type interne et formel, de laquelle se réclame la linguistique textuelle de J.M. Adam (1999) notamment, fait du texte ainsi défini son objet théorique propre, tandis qu'une approche de type externe, attentive au cadre de l'énonciation, envisage le texte en lien avec la pratique sociale dont il est à la fois un produit et un des éléments structurants. On retrouve dans ces deux approches les deux manières de rendre intelligible la question du sens, que F. Rastier propose de désigner respectivement sous les termes de tradition logico-grammaticale et de tradition rhétorico-herméneutique (1987, 2001).

Finalement, c'est l'enjeu de la délimitation de la linguistique par rapport à ses « extérieurs », encore désignés en tant que ses « ailleurs » qui se perpétue dans le choix de l'un ou l'autre de ces paradigmes, l'approche textuelle et l'approche discursive. Or, dans la perspective du présent travail, il est important de soutenir le caractère complémentaire de ces deux approches. Ce dont il va être question au cours de cette partie, c'est que l'intérêt de la notion de genre, telle que la définit notamment M. Bakhtine, tient précisément au fait qu'elle permet d'explorer cette complémentarité. En d'autres termes, comme le souligne J.M. Adam (1999 : 83), « le genre relie ce que l'analyse textuelle parvient à décrire linguistiquement à ce que l'analyse des pratiques discursives a pour but d'appréhender socio-discursivement ».

Par ailleurs, le genre représente aussi, comme nous le verrons, ce niveau d'analyse où concilier les deux termes de la première dichotomie évoquée, la langue et son utilisation individuelle ; nous reviendrons sur ces aspects, qui sont autant de raisons pour lesquelles la notion de genre constitue selon nous une entrée privilégiée pour aborder les pratiques discursives, et plus particulièrement les pratiques d'écriture de l'article de recherche en sciences humaines.

³ Cf. D. Maingueneau, 1997.

⁴ L'opération de soustraction qui permet de passer du discours au texte, et, conjointement, la définition du discours comme inclusion du texte dans le contexte est tout à fait révélatrice quant à la conception ancillaire du contexte dans le modèle-code du sens, comme l'a bien montré G. Kleiber : dans cette optique, la phrase trouverait la complétude de son sens une fois « trempée » dans le contexte (1994 : 14).

1.1.2. Options terminologiques et conceptuelles

Avant tout, il reste, au terme de ce parcours définitoire des notions de texte et de discours, à préciser les acceptions retenues ici. Le terme de texte sera entendu dans la continuité des points de vue de F. Rastier (1996 : 19) et de J.P. Bronckart (1996 : 71), qui en dépit de leurs divergences, ont ceci en commun de considérer le texte comme un objet empirique, qui consiste en une suite linguistique attestée, et constitue le produit d'une pratique sociale.

Notre étude comportera une part importante d'analyse textuelle ; mais il s'agit, plus exactement, d'une étude de pratiques scripturales, que nous concevons en tant que pratiques discursives ; nous utiliserons ce syntagme ou celui de pratiques langagières pour désigner les pratiques de production et d'interprétation des textes. Ces pratiques ne se réduisent pas à ce qui est observable au niveau textuel, et sur ce point nous rejoignons donc la position de D. Slatka. Mais à la « formule d'inclusion du texte dans le champ plus vaste des pratiques discursives » (Adam, 1999 : 39), nous privilégions une problématique de leur articulation, en nous centrant sur la seule production verbale pour l'aborder par rapport aux textes en termes de processus et de produit. La production verbale – autrement dit le versant génétique des pratiques langagières qui intéresse plus particulièrement notre étude des pratiques d'écriture de l'article – ne relève pas en elle-même de la linguistique : elle est abordée par le linguiste du point de vue des textes produits. C'est avec la notion de genre que nous verrons comment relier les usages linguistiques tels qu'ils se manifestent au niveau des textes avec leur contexte social, avec leurs conditions de production, et avec l'interlocution et l'interdiscursivité qui définissent la production de tout texte.

Nous emploierons donc le terme de discours au sens de pratiques langagières verbales, conçues comme des pratiques sociales ; chaque groupe de pratique sociale correspondant à un discours, nous parlerons de discours scientifique comme d'un type de discours, au même titre que le discours juridique ou littéraire. La terminologie ainsi établie est en adéquation avec les propositions de F. Rastier (2001). Enfin, pour situer en première approche la notion de genre dans ce cadre, nous dirons avec F. Rastier que chaque type de discours recouvre un ensemble de genres, que chaque texte relève d'un genre, et que le genre est ce qui rattache un texte à un discours.⁵

Reste à traiter du point sur lequel nous avons introduit la discussion de ces problèmes terminologiques, la dénomination de la notion de genre : genre de texte ou genre de discours ? Précisément parce qu'un genre fonctionne comme instance de médiation entre le texte et le discours, il nous paraît quelque peu présomptueux de vouloir donner une solution définitive à ce débat terminologique. Cependant, parler de genres de discours ne va pas sans risques d'ambiguïté avec ce que nous avons défini en tant que types de discours. Nous préférons donc genres de textes, en insistant sur l'idée que le genre est le lieu où interroger les conditions sociales de production et d'interprétation des textes, comme nous allons le voir maintenant.

Notre examen de la notion de genre commencera par la mise en évidence de son intérêt pour l'approche des usages linguistiques dans leur lien avec les fonctionnements sociaux. C'est alors en situant la définition des genres du point de vue de l'acte d'énonciation que nous pourrions nous interroger sur la caractérisation du genre, et du genre de l'article de recherche en particulier.

⁵ A ces niveaux hiérarchiques supérieurs au texte, F. Rastier ajoute un niveau intermédiaire entre le genre et le discours : la médiation des « champs génériques », ou « groupe de genres qui contrastent voire rivalisent dans un champ pratique : par exemple, au sein du discours littéraire, à l'époque classique, le champ générique du théâtre se divisait en comédie et tragédie » (2001).

1.2. L'interdépendance entre usages linguistiques et fonctionnements sociaux : les genres comme entités socio-historiques et formelles

1.2.1. Pratiques discursives et formations sociales

Si, à la suite de E. Benveniste, on définit l'énonciation comme « la mise en fonctionnement de la langue par un acte individuel d'utilisation » (1970), l'utilisation de la langue s'effectue comme le souligne M. Bakhtine⁶ à l'intérieur de domaines d'activité. En désignant ces domaines d'activité sous les termes de « sphères d'usage du langage », M. Bakhtine cherche à mettre l'accent sur les relations d'interdépendance entre le domaine des activités humaines et celui des productions langagières : les domaines de l'activité humaine se rattachent toujours selon lui à l'utilisation du langage, laquelle ne peut être conçue indépendamment de ce cadre.

On retrouve ce parti pris dans les travaux de J.P. Bronckart (1985, 1996). Celui-ci s'inscrit dans le cadre de la psychologie de l'interactionnisme social, dont le projet est de saisir les actions humaines (en tant qu'unités psychologiques) dans leurs dimensions sociale et discursive constitutives. Pour J.P. Bronckart, l'activité humaine s'organise dans différentes formations sociales, qu'il définit en tant que groupes historiquement constitués dans des conditions variables. A ce niveau sociologique vient s'articuler la question du rôle du langage : J.P. Bronckart montre que dans l'espèce humaine, la coopération des individus dans l'activité est régulée et médiatisée par leurs interactions verbales » (1996, 1^{ère} partie). Le processus de sémiotisation à l'œuvre dans les formations sociales conduit alors J.P. Bronckart à parler de « formations socio-langagières », cette dénomination permettant nous semble-t-il d'insister sur le versant processuel de ce que M. Bakhtine (1977) appelle les sphères d'usage du langage.

C'est cette même idée qui se manifeste dans la notion de « discourse communities » ou communautés discursives, utilisée en particulier par J. Swales (1990) pour rendre compte de la constitution, *via* les pratiques communicatives, de communautés d'usages qui déterminent des communautés de locuteurs. Par rapport à l'analyse du discours évoquée *supra*, la perspective est renversée : l'idée est que les pratiques professionnelles, les communautés sociales et les institutions se construisent dans le langage (Branca-Rosoff, 1999 : 119).

Pour ces différents auteurs, et c'est aussi l'idée que nous défendons dans notre étude, l'enjeu consiste à affirmer que l'ordre social et l'ordre langagier sont mutuellement constitutifs, le langage se construisant dans et par des relations sociales, le social se construisant sous l'effet des interactions entre individus médiatisées par le langage.

C'est dans ce cadre, où s'affirme l'interdépendance entre l'activité humaine et la mise en fonctionnement de la langue que M. Bakhtine convoque la notion de genre ; elle permet de rendre compte du fait qu'en dépit du caractère historiquement singulier et non reproductible de l'(acte d')énonciation, l'énoncé ne serait pas infiniment variable :

« L'utilisation d'une langue s'effectue sous la forme d'énoncés concrets, uniques (oraux ou écrits) qui émanent des représentants de tel ou tel domaine de l'activité humaine. L'énoncé reflète les conditions spécifiques et les finalités de chacun de ces domaines [...] « tout énoncé pris isolément est, bien entendu, individuel, mais chaque sphère d'utilisation du langage élabore ses types relativement stables d'énoncés, et c'est ce que nous appellerons les genres du discours ». (1977 : 127)

⁶ Lorsque nous évoquons M. Bakhtine, nous nous référons principalement aux écrits suivants : Bakhtine/Volochinov, 1977, Bakhtine, 1984 et Todorov, 1981. Ce sont les travaux du Cercle de Bakhtine qui sont visés, mais nous les désignons pour des raisons de commodité comme le fait de M. Bakhtine.

En conformité avec les écrits de M. Bakhtine, J.P. Bronckart propose de définir les genres comme des « formes communicatives historiquement construites par diverses formations sociales en fonction de leurs intérêts et objectifs propres » (1996 : 46).

1.2.2. Le genre comme notion « bi-face »

En premier lieu, nous pouvons donc définir le genre, dans la continuité de ces points de vue, en tant qu'entité socio-historique et formelle ; il semble effectivement admis aujourd'hui que le genre, « notion bi-face » (Branca-Rosoff, 1999 : 116), se situe à l'intersection de propriétés sociales, ou plus précisément socio-historiques, et de propriétés formelles : élaborés par une formation sociale et réglés par ses intérêts et enjeux spécifiques, les énoncés présentent des caractéristiques stables, des proximités, reflétant leur fonction communicative et leurs conditions spécifiques de production. Dans cette optique, tout texte, toujours singulier, relève d'un genre, correspondant à un type d'interaction dans un domaine d'activité.

Le genre apparaît donc comme un lieu d'observation privilégié de l'articulation entre « la détermination externe de la parole par ses conditions de production, et l'organisation interne des discours produits dans ces conditions différentes » (Doline, 1999 : 28). Encore faut-il, pour faire du genre un objet observable, s'interroger sur les modalités de cette articulation. Cela suppose, d'abord, de définir plus en avant les aspects socio-historiques et formels du genre.

On ne saurait, en effet, se contenter d'une approche des conditions socio-historiques des pratiques discursives en termes de « types de situations données », de « situations de communications types » ou « récurrentes » ; quant aux caractéristiques relativement stables sur le plan textuel, les qualifier d'affinités ou, à l'instar de L. Wittgenstein (1961), d'« air de famille », n'est pas davantage satisfaisant et mérite de plus amples explications.

Les apports de Bakhtine à cet égard restent nous semble-t-il assez programmatiques. Associés à d'autres études actuelles de la notion de genre, ils permettent toutefois de formuler dès à présent deux séries de remarques, correspondant aux deux faces du genre.

Le genre dans sa dimension textuelle

En ce qui concerne d'abord les conventions génériques en tant que conventions textuelles, M. Bakhtine montre que les codifications génériques opèrent au niveau thématique, stylistique et compositionnel. Ces trois dimensions mises en évidence par M. Bakhtine ne vont pas sans évoquer la tripartition rhétorique de l'*inventio*, la *dispositio* et l'*elocutio* (Aristote, 1991). Ce qu'il paraît d'abord important de retenir, c'est que l'ensemble des niveaux de textualité répondrait à des normes du genre, et que, selon ce principe de détermination du local par le global, les composantes textuelles se font écho en regard du tout.

D'autre part, M. Bakhtine ne sépare pas la forme du contenu : sémantiquement, il définit le genre comme un modèle du monde que propose le texte. C'est aussi ce que dit M. Foucault (1969a) des formations discursives : générant des modalités particulières de mise en fonctionnement de la langue, elles façonnent ce faisant les connaissances des membres d'une même formation sociale ; en d'autres termes, les genres déterminent donc un « mode de penser-vivre au sein d'une communauté » – le terme de communauté étant entendu par l'auteur au sens de « communauté discursive » (C. Donahue, 2002).

F. Rastier (2001) développe cette idée en soulignant que c'est par les genres que s'opère la médiation sémiotique, celle qui articule le physique et le représentationnel ; à

l'encontre d'une approche qui restreint le champ de la réflexion sur le signe au palier de l'unité linguistique, la notion de genre s'avère une heuristique pour questionner ce que F. Rastier propose d'appeler la « semiosis textuelle » (2001 : 248), autrement dit un rapport normé entre signifiant et signifié au palier textuel, et pour appréhender cette semiosis textuelle dans sa dimension sociale : c'est à travers les genres que prend forme l'activité sémiotique, dont la fonction médiatrice tient à ce que le réel, ou le monde en soi, n'est jamais accessible directement mais au travers des représentations que les hommes s'en font, et qui s'élaborent dans et par le langage et les langues.

Les tenants d'une approche praxématique, et en particulier C. Détrie, P. Siblot et B. Vérine (2001 : 173-175, 187), parlent à cet égard de « logosphère », « littéralement univers du discours ou du langage », jouant le rôle de « maillage » ou de « grille » qui, construite à partir des expériences pratiques, met en ordre le monde et permet la connaissance, nécessairement indirecte, du réel.

Ainsi se manifeste l'intérêt qu'il y a à penser la notion de genre pour envisager dans une problématique intégratrice la production verbale avec ses dimensions sociale et cognitive, car c'est à travers ces formes discursives, socialement élaborées, que se construisent l'activité cognitive et les représentations du monde, autrement dit les « mondes représentés » pour reprendre une expression chère à J.P. Bronckart (1996).

Le genre dans sa dimension contextuelle

La seconde série de remarques a trait aux contraintes socio-historiques pesant sur les pratiques discursives. Pour M. Bakhtine, la relative stabilité des énoncés est déterminée par un certain nombre de facteurs dont les plus importants sont le destinataire (tel qu'il est conçu par le locuteur), les intentions de celui-ci, et plus largement la situation dans laquelle se déroule la communication : « qui parle à qui, voilà qui détermine le genre » selon M. Bakhtine (Todorov, 1981 : 288), insistant sur l'importance qu'ont la position sociale de l'un et l'autre des partenaires de l'échange, et la position hiérarchique de l'un par rapport à l'autre.

Les conventions génériques rapportent ainsi le texte à la situation de communication, laquelle est appréhendée par M. Bakhtine en lien avec les sphères d'usage du langage ; autrement dit, la situation est conçue par M. Bakhtine en accord avec l'approche de F. Rastier (2001), qui la définit en tant qu'occurrence d'une pratique sociale. Comme le souligne F. Rastier, le genre rattache donc un texte non seulement à une situation, mais aussi à une pratique sociale, laquelle, nous l'avons souligné au-dessus, correspond aussi à un mode de rapport au monde.

Distinguer ces niveaux avant d'envisager comment ils interagissent permet déjà de progresser un peu dans la définition des genres : si l'étude du genre suppose de coupler l'étude du texte à celle de son contexte, il est nécessaire de mettre en évidence ce que recouvre ce terme de contexte dont le caractère englobant dilue les enjeux ; avec la notion de genre, le contexte n'est plus réduit à un « extérieur » au statut fort problématique, il représente ce qui conditionne la production du texte.

Or, nous verrons que ce n'est qu'en inscrivant la définition des genres dans l'acte d'énonciation, i.e. de mise en fonctionnement du langage qu'il sera possible d'éclaircir et de rendre opératoire la notion de conditions de production langagière, et de mieux comprendre ce couplage entre fonctionnements linguistiques et sociaux qui caractérise le genre.

Mais auparavant, un rapide aperçu de la manière dont la relation entre les deux faces des genres a été abordée dans l'histoire s'avère nécessaire.

1.2.3. Aborder les relations entre propriétés sociales et formelles : approches taxinomistes versus empiriques des genres

L'hypothèse de l'existence de genres, autrement dit l'hypothèse selon laquelle des conditions de productions stables, adossées aux fonctionnements sociaux, laissent prévoir une régularité des énoncés produits ne date pas des écrits du cercle de Bakhtine ; dans la rhétorique antique, les *genera dicendi* sont définis comme des « pratiques qui visent à réguler la vie publique, auxquelles doivent correspondre des régularités internes » (Branca-Rosoff, 1996)⁷.

Comme le montre A. Compagnon (2001), les trois genres oratoires mis en valeur par la rhétorique vont servir au cours des siècles d'ébauche de classification générique de l'ensemble des textes littéraires. En s'étendant à l'ensemble des productions verbales, les perspectives classificatoires et typologiques perpétuent la quête d'une synthèse exhaustive. Or, toute taxinomie se heurte d'abord au problème des critères de classement à retenir ; ceux-ci sont pluriels et en étroite relation, comme nous aurons l'occasion d'y revenir.

On peut noter avec F. Rastier (1999 : 93) que pour rendre compte de l'hétérogénéité des textes, les tentatives typologiques introduisent une certaine flexibilité dans les classements. Dans cette optique, le genre est conçu comme un type ou un prototype dont les textes seraient des occurrences. Mais l'approche typologique n'échappe pas à la question de la validité des catégories descriptives : comme l'explique K. Canvat (1996 : 9), l'enjeu consiste à mesurer en quoi les occurrences sont plus ou moins typées, par la mise en relation des textes avec une « base de typologisation », c'est-à-dire un ensemble de critères de classement. Par ailleurs, la notion de type s'avère problématique pour décrire l'évolution historique des genres (Rastier, 2001 : 236) : ce ne peut être que dans ses occurrences que se définit le type.

Les difficultés que rencontrent les tentatives typologiques se traduisent par la multiplication des typologies proposées. Face à cette inflation, que J.L. Chiss (1987) n'hésite pas à qualifier de « malaise dans la classification », apparaît la nécessité de structurer cet espace de recherche. C'est ainsi que A. Petitjean (1989) entreprend une typologie des typologies, et que d'autres auteurs proposent plus modestement de faire le point sur l'abondante littérature en matière de types et genres de textes — ou de discours. On peut citer notamment les articles de K. Canvat (1996) et S. Branca-Rosoff (1999)⁸, qui présentent l'intérêt d'explicitier le statut des paramètres et des catégories sur lesquelles sont construites les différentes approches, et de les situer ainsi les unes par rapport aux autres.

Il reste que les limites des tentatives typologiques expliquent qu'elles tendent actuellement à être délaissées au profit de l'étude des genres sociaux empiriques, comme le montrent F. Rastier (2001) et S. Branca-Rosoff (1999) notamment.

F. Rastier (2001 : 238) souligne que ce changement d'optique consiste à opter, dans un sens prudentiel, pour une problématique descriptive, en renonçant donc à une problématique normative, qui projette sur les textes étudiés une grille établie *a priori*. Il s'agit d'étudier les caractéristiques des genres à une époque donnée, sans chercher à les subsumer sous des catégories générales. Les approches empiriques des genres sont ainsi en mesure d'intégrer la question de leur évolution, et de prendre en compte le fait que la relation entre les fonctionnements sociaux et les textes n'a rien de mécanique.

Bien que l'objet de M. Bakhtine ne se situe pas dans l'analyse détaillée des caractéristiques de chaque genre, notre intérêt pour son approche théorique de la notion tient

⁷ Cette hypothèse, par ailleurs, se retrouve de manière intuitive dans la conscience métalangagière des locuteurs : ceux-ci classifient spontanément leurs pratiques discursives en genres, de même que l'ensemble de leurs productions culturelles.

⁸ Ces articles introduisent respectivement les numéros 37/38 de la revue *Enjeux* et 87 de la revue *Langage et Société* consacrés à ces problématiques des types et genres textuels.

précisément pour beaucoup à ce qu'elle n'est pas exclusivement centrée sur l'idée de stabilité et sur les réseaux d'oppositions des genres en fonction des domaines d'activité. En abordant la question des genres en lien avec l'activité humaine, M. Bakhtine met l'accent sur la diversité des genres :

La richesse et la variété des genres du discours sont infinies car la variété virtuelle de l'activité humaine est inépuisable, et chaque sphère de cette activité comporte un répertoire des genres du discours qui va se différenciant et s'amplifiant à mesure que se développe et se complexifie la sphère donnée (Todorov, 1981 : 265).

M. Bakhtine apporte ainsi un éclairage nouveau aux impasses typologiques : de même que l'ensemble des tenants d'une approche empirique des genres, il thématise la difficulté, voire la vanité qu'il y a à définir des catégories *a priori* ; il montre que les genres, en raison de leur variété, en synchronie et en diachronie, sont indénombrables, que leurs frontières sont perméables, et qu'il est par conséquent impossible de les soumettre à une classification stable et exhaustive.

Cela ne remet pas pour autant en question l'intérêt de la notion de genre. La singularité de l'approche bakhtinienne de la notion relève d'un déplacement théorique : aux conceptions classificatoires et typologiques, M. Bakhtine oppose une définition du genre inscrite dans l'acte d'énonciation. Autrement dit, il introduit la dimension du sujet et de son activité, et c'est ce qui selon nous rend possible l'articulation – plutôt que la mise en relation mécanique – entre la définition socio-historique des genres et leurs caractéristiques textuelles.

1.3. Le genre comme principe de régulation des pratiques discursives

Tous nos échanges, y compris les échanges quotidiens les plus spontanés, s'inscrivent dans ces formes communicatives que sont les genres :

« Nous apprenons à mouler notre parole dans les formes du genre, et, entendant la parole d'autrui, nous savons d'emblée aux tout premiers mots, en pressentir le genre » (Bakhtine, cité par Todorov, 1981 : 285).

Il nous faut donc examiner plus en avant cette hypothèse de la structuration générique des pratiques de production et d'interprétation des textes. Dans cette optique, nous nous centrerons d'abord sur le rôle joué par les genres d'un point de vue linguistique, pour montrer ensuite que le genre fonctionne comme modèle guidant les pratiques discursives. La mise en évidence de l'interlocution et de l'interdiscursivité comme principes constitutifs des pratiques discursives permettra alors de comprendre ce qui explique que le sujet parlant se réfère à des modèles du genre.

1.3.1. L'incidence des genres sur la langue et son étude

Le rôle joué par les genres s'avère décisif du point de vue linguistique : comme le note F. Rastier,

il n'existe pas de texte, (ou même d'énoncé), qui puisse être produit par le seul système fonctionnel de la langue (au sens de mise en linguistique) ; en d'autres termes, la langue n'est pas le seul système sémiotique à l'œuvre dans une suite linguistique, car d'autres codifications sont à l'œuvre, le genre notamment » (1989 : 37).

Autrement dit, si les pratiques discursives se réalisent nécessairement à travers une langue naturelle, elles ne se réduisent pas aux contraintes linguistiques, et c'est ce qui explique l'importance prise dans la notion de genre dans les études en didactique des langues : la question des compétences langagières des apprenants ne peut être envisagée restrictivement en termes de maîtrise de la langue, mais demande à être traitée en termes de maîtrise des genres et d'acculturation aux genres (cf. notamment B. Schneuwly, 2001).

La notion de genre ouvre en fait à l'articulation entre la langue et la parole, aux sens saussuriens de ces termes ; elle explique les impasses auxquelles aboutissent les approches linguistiques dites « internes », qui identifient les unités de langue et décrivent leur fonctionnement indépendamment de leur utilisation effective.

Outre le fait que les caractéristiques linguistiques ainsi mises en évidence se limitent tendanciellement au cadre phrastique, intervient tôt ou tard la nécessité, assumée de manière plus ou moins explicite, de s'en remettre à des facteurs externes ou contextuels ; c'est particulièrement frappant pour tout ce qui touche à la question du sens : comme le montre G. Kleiber (1994), le contexte sert de « roue de secours » aux approches sémantiques en linguistiques qui cherchent à pallier les faiblesses d'un modèle-code du sens⁹. Pour dépasser les limites de telles approches, G. Kleiber souligne la nécessité de renoncer à la relation d'opposition et de complémentarité entre les objets linguistiques et les objets externes, au profit d'une problématisation de leur articulation. Le genre apparaît comme le niveau où construire cette problématique : c'est aux contraintes génériques qu'il est fait référence lorsqu'un jugement d'inacceptabilité se trouve réfuté par un contre exemple avancé en référence à des « circonstances particulières ».

Si la notion de genre mérite toute l'attention des linguistes, c'est parce que tous les niveaux de la textualité, sur le plan du signifiant comme du signifié, varient avec les genres : le lexique, l'acceptabilité d'un énoncé, la syntaxe et les structures textuelles sont contraintes par les genres (Rastier, 2001 : 230), de même que ces unités de palier inférieur au texte que constituent les séquences, telles que les définit J.M. Adam (1999).

Nous pouvons dire avec F. Rastier que, comme il n'existe pas de pratique générale non spécialisée, le concept de langue générale reste normatif (2001 : 228) ; c'est le genre qui constitue « l'instance historique majeure d'actualisation et de normalisation de la langue » (ibid. : 272). Il s'agit donc, et c'est dans cette perspective que se situe le présent travail, de s'orienter vers l'étude des textes en usage, en reliant les textes aux pratiques où ils sont produits et interprétés.

Cela dit, un tel changement de paradigme ne va pas de soi. Comme le souligne J.P. Bronckart (1996 : 145), passer des unités de la langue à leur usage en discours pose problème, car les méthodes d'analyse ne sauraient être transposées d'un objet à l'autre. Or, il nous semble que ce qui rend possible l'étude de ces usages, c'est le fait qu'ils ne sont pas, pour reprendre ce que dit M. Bakhtine, infiniment variables, ou, en d'autres termes, qu'en dépit de la variabilité de ces usages, il existe des principes qui les régulent.

Nous allons montrer maintenant en quoi le genre constitue cette instance de régulation des pratiques discursives qui permet de les observer dans leur unité et leur diversité.

⁹ Ce modèle-code conçoit le sens en tant que produit compositionnel des significations afférentes aux signes linguistiques.

1.3.2. Le rôle des genres du point de vue du sujet et de ses pratiques discursives : les genres comme modèles textuels

L'hypothèse de la régulation des pratiques de production et d'interprétation des textes par les genres consiste à concevoir, à l'instar de M. Bakhtine, que les genres « organisent notre parole de la même façon que l'organisent les formes syntaxiques » (Todorov, 1981 : 285). Elle permet à M. Bakhtine de souligner que « si les genres du discours n'existaient pas et qu'il nous faille les créer pour la première fois, l'échange verbal serait impossible » (ibid.). Ainsi, dans les termes de J.P. Bronckart, les genres, en tant que formes communicatives élaborées par l'activité des générations antérieures, constituent un « réservoir de modèles textuels auquel tout agent d'une action langagière doit puiser » (1996 : 117).

Des modèles pour l'interprétation

La notion de genre ouvre des perspectives prometteuses pour rendre intelligible ce qui se joue lors de l'interprétation. La sémantique interprétative développe aujourd'hui l'hypothèse formulée dans les écrits du cercle de M. Bakhtine selon laquelle de toute parole on pressent le genre. La problématique de l'interprétation se fonde sur la reconnaissance de *La part du lecteur* (Visoz, 1989) ou plus largement du sujet interprète pour éviter de ne se référer qu'à l'écrit. L'enjeu consiste alors à s'interroger sur les modalités de la construction du sens par le sujet interprète, à partir des indices matériels attestés dans le texte.

Dans ce cadre, le genre s'apparente à ce que H.R. Jauss (1978), se donnant comme objet l'esthétique de la réception, nomme un « horizon d'attente » : c'est au travers des genres que nous interprétons les textes oraux et écrits, en ce sens que la conception générique avec laquelle l'interprète aborde le texte guide son interprétation ; les genres, pour reprendre les termes de F. Rastier (2001), ont une incidence sur les parcours interprétatifs d'actualisation des sèmes.

Si nous nous référons alors à la réflexion sur le fonctionnement du signe élaborée par C.S. Peirce (1978), et au triangle sémiotique qu'il propose, nous pouvons envisager les genres comme des interprétants, c'est-à-dire des images mentales qui vont déterminer la signification.

Il reste cependant à mettre l'accent sur le fait que, si le genre fonctionne comme un modèle de reconnaissance qui oriente la compréhension, il ne saurait être conçu comme s'imposant *a priori* et de manière figée et définitive ; autrement dit, comme le souligne A. Compagnon (2003), le genre ne fonctionne pas en tant que préjugé : la construction du sens par le sujet passe par des négociations, à partir des indices attestés dans le texte, et les hypothèses que se fait le sujet sur l'appartenance générique d'un texte sont révisées au cours du processus d'interprétation. Concevoir le genre comme modèle guidant le processus d'interprétation suppose donc de prendre acte des détours qui définissent les cheminements interprétatifs, et de mettre l'accent sur le fait que ce modèle varie au cours de ce processus.

Nous retrouvons ainsi un trait fondamental de la définition du signe chez C.S. Peirce (1978) : la signification selon lui se stabilise sous l'effet d'une chaîne d'interprétants qui est en principe infinie. Non seulement les éléments du texte, mais également les entours du texte constituent autant de clés pour l'interprétation ; c'est en ce sens que peut être assigné au genre ce rôle de médiation dans l'interprétation, et en tenant compte du double principe de détermination du local sur le global et du global sur le local (Rastier, 2001).

L'expérience des genres joue ainsi un rôle décisif pour l'interprétation, car c'est cette compétence, mobilisée par l'interprète, qui détermine les attentes génériques ; à ce titre, nous voyons à nouveau s'affirmer l'importance de cette dimension dans une perspective didactique. Façonnée dans ses pratiques des textes, cette connaissance que le sujet se forge des genres de textes en usage intéresse donc également au plus près la question de la production verbale.

Des modèles pour la production

Du point de vue de la production verbale, les échanges langagiers sont à comprendre en référence aux « actions langagières », et c'est la situation d'action intériorisée qui influe sur la production d'un texte empirique (Bronckart, 1996 : 94), mais mise en interface avec les genres de textes (op. cité : 110). Que la littérature mette l'accent sur le « vouloir-dire » du locuteur (cf. par exemple Bakhtine) ou sur une norme de comportement dans une situation donnée (cf. par exemple Berkenkotter et Huckin, 1995), l'idée est toujours là que le locuteur recourt aux formes communicatives en usage dans les formations sociales, et plus spécifiquement dans la pratique en cours, en fonction de l'expérience qu'il a des genres.

Cependant, de même qu'en matière d'interprétation, le modèle de production verbale que constitue le genre est à comprendre de manière dynamique et évolutive, et non en tant que modèle figé dans son unicité ; il convient en effet de se garder de renouer avec une approche classificatoire, normative et prescriptive des genres, déjà évoquée, qui caractérise des œuvres modèles en tant que normes à intérioriser et à imiter, et qui ce faisant passe sous silence l'activité du sujet parlant. Si M. Bakhtine défend l'idée qu'énonçant, nous reproduisons des types d'énoncés, il insiste aussi sur la singularité de l'acte d'énonciation, et, partant, sur le fait que l'énoncé, en sa qualité même d'énoncé, est non reproductible.

La définition du genre en tant que modèle du point de vue des pratiques de production et d'interprétation des textes permet donc de s'interroger plus en avant sur la question de la stabilité du genre.

Stabilité des genres et prescriptivité du modèle

Le genre se définit à la fois par sa stabilité, laquelle justifie que l'on parle de genre, et par les variations inhérentes à l'activité énonciative des sujets dans l'interaction verbale.

Cette approche du genre, et corrélativement du système des genres à laquelle invitent les écrits de M. Bakhtine montre que la stabilité d'un genre, autrement dit son unité, procède à la fois d'un principe d'identité, de conformité, ou encore de reproduction et d'un principe d'écart, de déplacement, de variation. Dans une dimension historique, ce double principe constitutif des genres fait clairement apparaître en quoi ceux-ci se laissent mal analyser en tant que classes, comme en tant que types dont l'actualisation dans les textes serait à traiter en termes d'écarts. L'analyse des genres appelle semble-t-il à développer une conception réticulaire des textes et des pratiques où ils sont en usage.

La question des limites d'un genre, et des frontières entre les genres, engage alors à la prudence : l'instabilité des genres rend vaine la question de savoir jusqu'où les variations sont caractéristiques d'un genre ou constituent à l'inverse un trait discriminatoire. En situant la problématique des genres du point de vue des pratiques discursives, nous soulignerons, après S. Branca-Rosoff (1999, 116-117) notamment, que l'indexation d'un texte à un genre suppose une activité d'interprétation : c'est l'évaluation sociale des pratiques, par le sujet interprète ou plus largement par les membres de la « communauté discursive » (Swales, 1990), qui préside

au jugement d'appartenance (Branca-Rosoff, *ibid.*), décidant de considérer que les variations sont autorisées ou non par le genre.

La stabilité du genre est donc à aborder, à l'instar de M. Bakhtine, en termes de prescriptivité pour le locuteur. Dans cette optique, on peut dire qu'il existe des genres plus ou moins codifiés voire stéréotypés, peu propices à la créativité du locuteur – M. Bakhtine donne à titre d'exemple le cas du formulaire administratif. D'autres sont davantage ouverts, laissant au locuteur une part d'individualité plus grande.

Cette gradation de la prescriptivité des genres sur un axe qui va de la reproduction à la singularité de leurs usages évoque la proposition de H. Boyer (1988) de penser l'ordre scriptural comme étant régulé par deux grands principes opposés, et par les tensions textuelles qu'ils engendrent : le principe de scription, ou de conformité, et le principe d'écriture, ou de créativité. Sans traiter de la question des genres, H. Boyer montre bien comment le principe d'écriture peut transgresser le principe de scription, et comment les deux principes interagissent dans le discours journalistique ; toutefois, il nous semble que son approche tend à constituer le pôle de l'écriture comme un pôle libéré de toute contrainte.

Nous défendons au contraire l'idée que le degré de prescriptivité d'un genre n'est jamais nul, ce qui revient à défendre la thèse que tout texte relève d'un genre, que le genre régule les pratiques discursives et que dans tout genre coexistent des « procédés canoniques obligatoires » et des « procédés libres » (Tomachevski, cité par Adam, 1999 : 93). La validité d'une telle bipartition resterait cependant à évaluer : la stabilité du genre implique certes des procédés canoniques mais qui doivent cependant être eux-mêmes soumis à une variabilité du genre.

Le déjà-là des genres : un obstacle à la création verbale ?

Du point de vue de leur stabilité et de leur prescriptivité pour le locuteur, les formes du genre seraient donc plus souples que les formes linguistiques, mais il faut, avec M. Bakhtine, mettre l'accent sur le fait qu'elles n'en sont pas moins normatives pour l'individu parlant, dans la mesure où elles relèvent du « déjà-là » et que ce n'est donc pas le locuteur qui les crée (Todorov, 1981 :287). La reconnaissance de ces formes communicatives socialement et historiquement construites dans lesquelles se moule l'échange verbal implique donc de renoncer à ce que J. Authier-Revuz (1984) appelle « l'illusion subjective de la parole » : le discours ne peut être restrictivement conçu comme un acte individuel, consistant en une combinaison absolument libre des formes de langue.

C'est peut-être cette idée sous-jacente à la notion de genre, tant il est vrai qu'elle déplace l'image que se fait la doxa du sujet et de son dire, qui explique les résistances qu'elle a pu rencontrer, en particulier dans le domaine de la littérature. Si là n'est pas l'objet de notre propos, il n'est toutefois pas inintéressant de se pencher sur cette lutte contre les genres, dont T. Todorov (1978), ou encore A. Compagnon (2003) notamment, expliquent les ressorts : la théorie littéraire du vingtième siècle soutient, au nom de la créativité, qu'avec la « modernité », les genres auraient disparu ; c'est cette thèse que défend R. Barthes (1973) dans *Plaisir du texte* : il conteste les genres qu'il voit comme des contraintes périmées, car l'œuvre selon lui échapperait par définition aux genres, et créerait son propre genre. A la prescriptivité du genre est donc opposé le principe d'écriture tel que le définit H. Boyer (1988) comme nous l'avons évoqué ci-dessus, en tant que pôle de la production verbale libéré de toute contrainte.

Il est vrai que, contrairement à d'autres types de discours plus codifiés, l'histoire de la littérature est faite d'innovations, qui sont le fruit d'individus ou d'écoles, et que c'est le déplacement des genres qui semble définir les avant-gardismes comme les grandes œuvres.

Mais conclure pour autant à la disparition des genres serait trop hâtif ; comme le note à juste titre T. Todorov (1978), dont les écrits, avec ceux de G. Genette notamment, plaident en faveur de la réintroduction du genre dans la théorie littéraire, affirmer que l'œuvre déplace le genre ne rend pas caduque la notion : la transgression en tant que telle suppose des lois, et marquer un écart par rapport à des normes du genre, c'est aussi maîtriser ces normes et les révéler.

C'est précisément ce qu'il nous paraît important de retenir de ce détour par le discours littéraire et sa théorisation : cette idée nous semble acceptable tant en matière de littérature que pour ce qui concerne le discours scientifique et l'article de recherche en particulier ; bien qu'il soit clairement admis pour la science qu'elle repose contrairement à la littérature sur des modèles d'écritures, les pratiques langagières dans leur ensemble sont soumises à cette problématique des normes et de la variation.

Or, la notion de genre, appréhendée du point de vue de l'acte d'énonciation, ne fait pas obstacle à une conception des textes sous l'angle de leur diversité ; au contraire, elle permet de rendre compte de la dimension évolutive des genres, et d'aborder ainsi le système des genres dans sa continuelle transformation : les genres viennent d'autres genres, un nouveau genre est toujours la transformation d'un ou de plusieurs genres anciens, par déplacement, par combinaison, par inversion (Todorov, 1978 : 44-60). Dans cette optique, si dans certains domaines d'activité, comme la littérature, les textes se définissent par l'hybridation et l'hétérogénéité des genres, il n'en reste pas moins vrai qu'aucun texte n'est hors de toute norme générique.

Il convient donc de concéder, à ceux pour qui les genres sont morts, que les genres littéraires hérités de l'histoire sont dépassés, comme le sont les approches classificatoires et typologiques, normatives et prescriptives de la notion de genre. Mais selon nous, la validité de la notion de genre ne se limite pas aux genres les plus conventionnels, et inversement, elle ne se limite pas aux textes littéraires : nous défendons l'importance des genres en tant que principes dynamiques de production et d'interprétation de tout texte. Nous ne saurions en effet souscrire aux conceptions de la création verbale qu'implique le refus des genres : ce serait soutenir que la création ne serait l'œuvre que du sujet créateur dans sa singularité ; or, l'idée de ce sujet créateur premier, donnant naissance à univers sémiotique *ex nihilo*, ne va pas sans toucher au mythe de l'incommunicabilité ; elle est contraire au principe consubstantiel du langage de conventions socialement et historiquement constituées.

1.3.3. Aux fondements des genres : l'hétérogénéité constitutive du discours et du sujet

Loin de ne concerner que le domaine de la littérature, cette tendance à opposer créativité et modèle, et à réduire la créativité à la part de l'individu dans sa singularité repose sur une idée simplifiée de la communication, et soulève un ensemble de problèmes, ayant trait à ses dimensions langagière, sociale et cognitive. C'est cette approche de la communication, héritée des modèles techniques de la transmission d'informations, qu'il s'agit de dépasser, pour comprendre, dans le cadre d'une approche dialogique des pratiques langagières, en quoi la création verbale n'est pas incompatible avec la notion de genre, ou en quoi la créativité et le modèle du genre ne sont pas deux principes antinomiques.

De la communication à l'interaction

Malgré quelques adaptations du modèle technique de la communication comme transmission de signaux porteurs d'informations (Shannon et Weaver), le modèle proposé par R. Jakobson

(1963) n'en demeure pas moins problématique : ce que dénonce C. Kerbrat-Orecchioni (1980) notamment, c'est que ce modèle pose deux locuteurs face-à-face, échangeant des informations sur le mode de la transparence, au moyen d'un code neutre et extérieur. Cette conception instrumentale de l'échange, cautionnée par la philosophie idéaliste, suppose des représentations cognitives préalables, codifiées dans un second temps lors de leur formulation ; cette codification rendrait plus ou moins possible les échanges, selon que l'on s'attache à voir dans les langues un « mauvais outil » (Henry, 1977), déformant la pensée préexistante et mettant en doute le caractère partageable des représentations du monde, ou au contraire un bon instrument : cela implique alors de faire abstraction des limites que présente cet outil pour garantir la transparence entre les mots et la pensée et entre les représentations d'un locuteur et de l'autre, d'où la mise à l'écart de la variation linguistique, de l'ambiguïté, de la polysémie.

Or, cette problématique de la communication, centrée sur ce qui devrait la rendre possible, à savoir une langue commune, se renverse si l'on admet que la langue ne peut être appréhendée que dans sa mise en fonctionnement à des fins communicatives qui trouvent leurs motivations dans les activités sociales par lesquelles les humains organisent leur rapport au « milieu », c'est-à-dire au monde.

A cette conception de la communication, reposant sur une approche monologique des productions verbales qui va de pair avec le postulat d'un sujet psychologique universel et autonome, il s'agit d'opposer une approche dialogique : en continuité avec les points de vue développés par M. Bakhtine et par les tenants du dialogisme bakhtinien, nous nous rallions à l'idée que le langage et la pensée sont interindividuels avant d'être individuels, et que le sujet, en tant qu'être psychologique et social, n'existe comme tel que sous l'effet des interactions langagières, qui sont des relations sociales. Seul un tel positionnement épistémologique, défendu également par J.P. Bronckart dans le cadre de l'interactionnisme socio-discursif (1996) permet de comprendre le rôle joué par les genres du point de vue de l'acte d'énonciation.

La production verbale est dialogique

L'approche dialogique du discours chez M. Bakhtine repose sur un déplacement de ce que l'on désigne ordinairement sous le terme de dialogue¹⁰, à savoir un échange entre individus placés en face-à-face. Dans cette optique, la plupart des énoncés se présentent comme monologiques.

Mais selon M. Bakhtine, tout échange verbal de quelque type qu'il soit est fondamentalement dialogique. Bien qu'abstrait de ses conditions de production, l'énoncé apparaisse comme monologique, cet isolement n'a rien de naturel : il représente un choix théorique, et un parti pris idéologique, que Bakhtine appelle à dépasser, en posant l'interaction verbale comme réalité première des pratiques langagières. Comme le souligne M. Bakhtine, « il ne saurait y avoir d'énoncé isolé » : « [un énoncé] n'est jamais le premier, jamais le dernier, il n'est qu'un maillon dans la chaîne complexe d'autres énoncés, et ne peut être étudié hors de cette chaîne » (1981 : 335).

C'est ainsi qu'est introduit par M. Bakhtine le concept de dialogisme : il réfère aux relations que tout énoncé entretient avec les énoncés produits antérieurement dans une sphère

¹⁰ En accord sur ce point avec un certain nombre d'auteurs, J. Bres, dans l'article « dialogisme » de *Termes et concepts pour l'analyse de discours* (Détrie et al., 2001 : 83), propose de réserver le terme « dialogique » à l'acception bakhtinienne, en le distinguant de « ce qui se manifeste comme dialogue externe » : le « dialogal ».

de l'échange, et avec le discours à venir que pourraient produire les destinataires, lesquels adoptent face à tout énoncé ce que M. Bakhtine appelle une « attitude responsive active ». Dans la perspective adoptée par M. Bakhtine, et que nous défendons à sa suite pour aborder la question de la production verbale, tout discours est donc doublement dialogique, les relations dialogiques entretenues avec les autres énoncés recouvrant deux aspects : le dialogisme interdiscursif et le dialogisme interlocutif.

Le premier aspect du dialogisme se situe dans les relations dialogiques interdiscursives : elles tiennent à ce que tout énoncé répond à des énoncés antérieurs, dans le sens où le discours entre en interaction avec des discours précédemment tenus sur le même objet, et convoque des mots déjà utilisés ; on peut citer M. Bakhtine :

L'objet a déjà, pour ainsi dire, été parlé, controversé, éclairé et jugé diversement, il est le lieu où se croisent, se rencontrent et se séparent des points de vue différents, des visions du monde, des tendances. [...] Un locuteur n'est pas l'Adam biblique, face à des objets vierges, non encore désignés, qu'il est le premier à nommer (Todorov, 1981 : 299).

Le dialogisme interdiscursif apparaît en ce sens comme un équivalent de ce que certains auteurs désignent en termes d'intertextualité, notamment J. Kristeva (1969, cf. aussi Piégay-Gros, 1996), qui s'inscrit dans la lignée du point de vue défendu par le cercle de M. Bakhtine et considère l'intertextualité comme propriété constitutive de tout texte.

Mais le dialogisme du discours dans l'approche bakhtinienne comporte un autre aspect, celui des relations dialogiques interlocutives : elles concernent parmi l'ensemble des discours avec lesquels le discours entre en interaction, le discours de l'interlocuteur. En effet, le locuteur, porteur comme le note également E. Benvensite d'une « intention d'influencer l'autre en quelque manière » (1970), cherche à anticiper sur la compréhension responsive de son interlocuteur, de sorte que ce rapport au discours de l'autre est constitutif de la production verbale : « tandis que j'élabore mon énoncé, je tends, d'une part, à déterminer cette réponse de façon active, d'autre part, je tends à la présumer et cette réponse présumée, à son tour, agit sur mon énoncé (je pare des objections que je prévois, je marque des restrictions, etc.) » (Bakhtine, cité par Todorov, 1981 : 303-304) ; En d'autres termes, « le discours est déterminé par la réplique non encore dite, mais sollicitée et déjà prévue » (Bakhtine, 1978 : 103).

En raison donc du dialogisme interdiscursif avec le déjà-là des discours préexistants, et parce que le discours, adressé, procède de la relation entre interlocuteurs, celui-ci ne peut être conçu comme le seul fait du locuteur. Autrement dit, il s'agit de dépasser la définition benvenistienne de l'énonciation en tant qu'acte individuel, au profit d'une « mise en communauté de l'énonciation » défendue par F. Jacques (1983). C'est l'idée à laquelle souscrit également J. Authier-Revuz (1982), qui propose de qualifier d'hétérogénéité constitutive du discours ce double principe constitutif des pratiques discursives que sont le dialogisme interdiscursif et interlocutif. Comme nous allons le voir, cette hétérogénéité constitutive du discours va de pair dans la réflexion de J. Authier-Revuz avec celle du sujet. S'inscrivant dans le cadre des approches énonciatives en linguistique, celle-ci développe par ailleurs les implications de ces principes au niveau des textes produits, en termes notamment d'hétérogénéité montrée.

L'hétérogénéité constitutive du sujet

L'approche dialogique des pratiques langagières, qui met l'accent sur les dimensions interdiscursive et interlocutive constitutives de tout discours, implique donc un dépassement d'une conception du sujet en tant qu'être autonome à la source du sens. C'est précisément

cette conception qui est mise en cause dans les critiques adressées à la pragmatique, par J. Authier-Revuz (1995) et C. Kerbrat-Orecchioni (1980) notamment.

Si la pragmatique nous paraît présenter ceci d'intéressant qu'elle aborde le langage en tant qu'action, et du point de vue de ses usages, l'accent porté sur la valeur illocutoire des énoncés conduit à doter le locuteur d'une intention initiale ; comme le montre R. Vion (1998 : 190), ce lien établi entre l'illocutoire et l'intentionnalité est très net chez J.L. Austin et R. Searle notamment. Or, nous dirons que si « dire, c'est faire », selon la célèbre formule de J.L. Austin (1962), le dire ne peut être imputable à la seule responsabilité d'un locuteur, mais implique un rapport à l'autre, qu'il contribue aussi à façonner.

Autrement dit, l'approche pragmatique du langage se doit d'intégrer sa dimension sociale, laquelle ne se réduit pas à l'interaction en face-à-face, souvent privilégiée dans les analyses conversationnelles développées en référence à la théorie des actes de langage. Par ailleurs, il convient de souligner, avec J. Authier-Revuz et C. Kerbrat-Orecchioni, que les approches énonciatives n'échappent pas non plus à cette conception discutable du sujet que sous-tendent les approches pragmatiques.

Or, le sujet ne se constitue comme tel que dans la relation avec ses « environnements » : à l'encontre d'une problématique psychologique d'un sujet seul face à un monde vierge, ou d'un sujet saisi dans son unité, en tant qu'entité homogène et extérieure au langage, il faut souligner, à l'instar de E. Benveniste, que le langage n'est pas un instrument qui permettrait de mettre en mots ce que le sujet a préalablement conçu, mais que « c'est dans et par le langage que le sujet se constitue comme tel » (1966 : 259-260). Il faut ajouter cependant qu'il s'agit aussi de concevoir le langage en lien avec l'expérience pratique des sujets dans leur rapport au monde.

Nous défendons donc l'idée que le sujet se construit individuellement en se socialisant par les interactions qu'il entretient, et que dans cette socialisation se joue tout à la fois une rencontre avec ce que J.P. Bronckart (1996) appelle les préconstruits socio-historiques, produits de l'activité des générations précédentes, et une appropriation individuelle de ces préconstruits. C'est sur une telle conception de la socialisation que s'accordent des approches aussi diverses que la théorie de l'agir communicationnel de J. Habermas (1973), l'interactionnisme social initié par L.S. Vygotsky (1934) – deux des fondements théoriques sur lesquels s'appuient les travaux contemporains de J.P. Bronckart, mais aussi l'approche du développement cognitif de l'enfant chez J.S. Bruner (1983), ou encore la théorie psycho-sociologique du sujet dans le cadre de l'école de Chicago et des travaux de G.H. Mead en particulier (1934).

Ces auteurs ont en commun de contester l'idée que le développement langagier serait sous la dépendance d'un fonctionnement cognitif qui s'élaborerait « dans un pur rapport au monde (...), un rapport vierge de toute histoire » (Bronckart, 1996 : 108), et militent en faveur d'une approche constructiviste, selon laquelle c'est dans les activités collectives, médiatisées par le langage, que se développent l'individu et sa pensée. C'est également cette thèse qui sous-tend la notion de logosphère, déjà évoquée, qui est proposée par les tenants d'une approche praxématique du langage (cf. notamment Détrie et al., 2001), ainsi que l'hypothèse de E. Sapir et B.L. Whorf, selon laquelle l'expérience que les hommes se font de la réalité est structurée par le langage (cf. l'article de Siblot qui lui est consacré dans Détrie et al., op. cité).

Il est remarquable de constater à quel point la mise en cause d'un sujet autonome, porteur d'intentions et d'une conscience souveraine, est transversale dans les sciences humaines et sociales ; plus précisément, elle se retrouve en sociologie, en psychologie sociale, en philosophie et enfin en linguistique, à partir du moment toutefois où cette dernière s'affranchit d'une focalisation structuraliste sur l'objet « langue », au sens saussurien du

terme, mais aussi d'une problématique générativiste, centrée sur les compétences d'un locuteur-auditeur idéal et universel.

« Co-acteur » et non simplement acteur (Vion, 1998 : 191), le sujet n'existe comme tel que dans son rapport à l'autre, ce que nous traiterons, en référence à J. Authier-Revuz (1982), en termes d'hétérogénéité constitutive du sujet. Comme le montre cet auteur, ce parti pris théorique se rallie aussi à la structuration clivée d'un sujet déterminé par l'inconscient, postulée par l'approche psychanalytique de S. Freud, en particulier telle qu'elle est poursuivie par J. Lacan. Bien que la problématique psychanalytique, et son objet, l'inconscient, n'intéressent pas directement notre propos, elle permet d'envisager la question du sujet en tant qu'être de langage en lien avec cette instance clivée en même et autre que constitue le sujet dans l'approche freudienne et lacanienne.

A nouveau donc, nous nous accordons avec M. Bakhtine à reconnaître « l'intersubjectivité comme logique antérieure à la subjectivité » (T. Todorov, 1981 : 51) : la subjectivité procède de l'intersubjectivité. Pour M. Bakhtine, l'individu porte en lui un « horizon social », par lequel il existe comme sujet. Ce rapport constitutif de l'individu à l'autre est bien mis en évidence par P. Bourdieu, qui conçoit le sujet en tant qu'« agent social » (par exemple, 1994), dans le cadre d'une approche objectiviste en sociologie. A l'instar de cet auteur, qui avec *La Misère du monde* (1990) oriente ses recherches vers une approche résolument subjectiviste, tout en intégrant le point de vue objectiviste qui définit sa sociologie, nous dirons que le sujet est à la fois agent et acteur social. C'est également en ces termes que se construit la réflexion sur l'action langagière chez J.P. Bronckart (1996).

Le sujet, comme le groupe social, se construisent en effet comme le souligne M. Bakhtine dans une dialectique entre une « force centrifuge », autrement dit l'hétérogénéité, et une « force centripète », qu'il aborde en termes d'« interorientation sociale ». L'identité trouve bien là ses deux principes constitutifs, le principe d'ipséité et d'altérité, ainsi nommés par P. Ricœur (1990). Finalement, il nous semble que ce parti pris commun aux différents auteurs que nous venons d'évoquer constitue une alternative intéressante pour éviter les pièges respectifs de ce que R. Vion (1998 : 200) pointe comme « deux problématiques simplistes » : celle d'un sujet « sous-socialisé et surpuissant » agissant de manière autonome, consciente et volontaire, et celle d'un sujet « sur-socialisé », assujetti aux normes sociales, dont on nie qu'il agit au profit de l'idée qu'il est agi.

Ce n'est donc pas en se saisissant d'emblée de sa seule unité, mais dans une dialectique entre unité et hétérogénéité qu'il paraît opportun d'appréhender la subjectivité, ce qui revient à dire, dans les termes de P. Ricœur, que c'est la mise en couple de l'altérité et de l'ipséité qui est constitutive de l'ipséité elle-même (1990 : 13).

1.3.4. Retour à la question de la production verbale : les genres comme modèles de textes, contraintes, et compétence.

Le principe explicatif de l'adoption par le locuteur de modèles du genre : intertexte et évaluation sociale

Mettre en évidence l'hétérogénéité du sujet et du discours permet de prendre acte du rôle joué par les genres du point de vue de la production verbale : la mise en fonctionnement du langage est à concevoir dans une dialectique entre les contraintes socio-historiques et langagières, et la singularité de l'acte d'énonciation, ouvrant un espace de décision possible pour le sujet. Se trouvent ainsi mises en évidence les limites de l'approche des productions écrites sous l'angle dichotomique d'un principe de scription et d'un principe d'écriture,

comme nous l'avons évoqué (Boyer, 1988). De manière intéressante, celui-ci montre que le principe de scription correspond au « territoire du Groupe, de la Communauté », tandis que le principe d'écriture ancrerait le discours dans le « territoire du Sujet » (op.cité : 18). Cela étant, ces deux pôles nous semblent reproduire les deux problématiques réductrices d'un sujet alternativement non conscient ou conscient, sur-socialisé ou sous-socialisé que met en cause R. Vion (1998 : 200), comme nous l'avons précédemment souligné.

Opter pour une approche des productions verbales sous l'angle de la problématique des genres du discours favorise la prise en considération du processus de socialisation à l'œuvre dans la mise en fonctionnement du langage. Les genres constituent comme le montre J.P. Bronckart (1996) ces formes communicatives élaborées par l'activité des générations précédentes, indexées à des pratiques sociales, et synchroniquement disponibles dans ce qu'il appelle l'intertexte, en tant que modèles plus ou moins stabilisés ; ce sont des instruments de la socialisation du sujet, laquelle suppose une rencontre avec les préconstruits socio-historiques de nature langagière, que le sujet adopte tout en les adaptant : en raison de la singularité de l'acte d'énonciation, le locuteur produit une configuration individuelle des modèles textuels auxquels il se réfère nécessairement.

Nous dirons donc, en accord en cela avec F. Rastier (2001) et J.P. Bronckart (1996), que les genres représentent une réalité intertextuelle, laquelle permet d'expliquer certaines affinités entre les textes. Mais il faut noter aussi, outre l'importance de ce rapport à l'intertexte dans la production verbale, celle de l'évaluation sociale ; comme le note F. Rastier, « relativement à leur genre, les performances sémiotiques sont sanctionnées comme convenantes ou non, qu'il s'agisse d'écrire des rapports d'activité ou de danser le pasodoble » (Rastier et Pincemin, 1999).

Cette évaluation repose sur des attentes génériques, c'est-à-dire sur des modèles du genre. Or, une approche dialogique des pratiques discursives ne peut qu'insister sur le fait que cette évaluation est déterminante dans la dimension interlocutive de la production verbale. Cette dimension de l'évaluation sociale mérite toute notre attention pour ce qui a trait à l'article de recherche : si les textes font toujours l'objet d'une évaluation sociale, celle-ci se trouve être institutionnalisée dans certaines revues avec comités de lecture et de rédaction ; l'expertise des articles soumis à la publication consiste bien en une évaluation, laquelle s'effectue à l'aune de modèles du genre. Les instances évaluatives jouent un rôle primordial du point de vue des normes du genre : un texte considéré comme s'écartant trop des normes du genre peut se voir refuser la publication, ou donner lieu à des demandes de correction ; la correction peut donc être définie en tant que procédure sociale de régulation du genre. Il nous faudra donc veiller dans le cadre de notre étude du genre de l'article à prendre acte du rôle joué par cette dimension constitutive de la production d'un article.

Non seulement le modèle du genre, en tant que modèle préconstruit relativement prescriptif pour le locuteur, n'est pas incompatible avec la création verbale, mais il en constitue une composante essentielle : l'adoption par le locuteur des modèles socio-historiquement élaborés que sont les genres trouve son principe explicatif dans cette double dimension sociale constitutive de la production verbale, celle de l'interdiscours et celle de l'interlocution.

Le genre apparaît donc comme ce niveau d'analyse des pratiques discursives où appréhender dans leur articulation les dimensions sociale, langagière et cognitive. En effet, les genres assurent ce que F. Rastier nomme la « médiation symbolique », celle qui articule l'individuel et le social, outre leur fonction de médiation sémiotique dont nous avons traité en amont (2001 : 272). Le genre représentant ce « lieu sémiotique de l'intersubjectivité en tant qu'elle est médiatisée par la Loi » (F. Rastier, op. cité : 273), l'approche des pratiques

discursives sous l'angle de la problématique de leur régulation par les genres permet non seulement de concilier mais aussi d'intégrer les deux territoires dont traite H. Boyer (1988), celui de l'individu et celui du groupe, tout en envisageant le langage en lien avec l'expérience des sujets dans leur rapport au monde.

Les genres comme contraintes de la production verbale

Il reste cependant à examiner la question de la différence entre les genres du point de vue de leur prescriptivité pour le locuteur : le cas des genres littéraires permet indéniablement de reconnaître que certains genres sont ouverts à la variation, tandis que d'autres, plus conventionnels, le sont moins. C'est cette différence, nous l'avons dit, qui conduit H. Boyer (1988) à discriminer les productions écrites selon qu'elles relèvent d'un principe de scription ou d'un principe d'écriture. Si selon nous le principe d'écriture suppose des contraintes génériques, au même titre que, comme le note H. Boyer, dans le principe de scription peut s'ouvrir un espace de décision du sujet, force est de reconnaître que la prescriptivité d'un genre varie avec chaque genre.

Cette question de la prescriptivité et des contraintes se pose tant au niveau intra-générique qu'au niveau inter-générique : elle doit permettre de rendre compte de la diversité observable des textes relevant du même genre, ainsi que de la différence des genres en termes de degré de prescriptivité.

Pour ce qui concerne l'article de recherche en sciences humaines, les études sur l'article ont mis en évidence qu'il est bâti sur un certain nombre de modèles canoniques et qu'il comporte un ensemble de routines scripturales (De Nuchèze, 1998) ; ce genre « institutionnalisé » dans l'optique de D. Maingueneau (1995) est donc décrit comme extrêmement contraint. Cela étant, tout lecteur d'articles ne peut que constater la diversité des textes relevant de ce genre.

Le genre comme nous l'avons souligné se constitue sous l'effet d'un double principe de stabilité et de variations inhérentes à la singularité des actes d'énonciation. Ainsi, l'étude de textes du même genre doit permettre de caractériser une unité du genre au niveau textuel. En même temps, l'étude du genre doit permettre en retour d'analyser la singularité des textes, et en particulier l'usage singulier que chaque auteur fait d'un genre et que nous qualifierons, à l'instar de F. Rastier (2001), de « style »¹¹.

La question de la prescriptivité d'un genre appelle alors à examiner celle des facteurs de variabilité. Selon M. Bakhtine, et à sa suite bon nombre d'auteurs, ce point faisant l'objet d'un relatif consensus, le facteur principal de variation est lié aux conditions mêmes d'énonciation : c'est la situation de communication qui détermine ce qu'H. Boyer (1988) traite en termes de « rapport contrainte-crédation » ; cette idée de contrainte demande toutefois à être précisée.

Il peut s'agir d'une part de l'analyser du point de vue des caractéristiques des textes, pour voir lesquelles sont particulièrement contraintes ; on peut penser *a priori* que les caractéristiques contraintes résultent tout aussi bien de prescriptions négatives, autrement dit à des procédés tendanciellement prohibés, ou de prescriptions positives, encourageant le recours à tel ou tel tour. C. Doline (1999) est d'avis que les normes du genre stipulent ce qu'il

¹¹ F. Rastier suggère notamment d'étudier en quoi le style, en tant qu'usage singulier du genre, peut être défini comme le propre d'un auteur, auquel cas il transcenderait les différents genres dont use cet auteur, ou en quoi, hypothèse en faveur de laquelle opte l'auteur, le style dépend des normes du genre dont l'auteur fait un usage singulier. Nous traitons de cette question des styles dans le genre de l'article de recherche au chapitre 6.

ne faut pas faire, davantage que ce qu'il faut faire, et il nous reviendra en traitant des normes du genre dans le cas de l'article de recherche de tenter de mieux cerner cette question. Cela étant, c'est vraisemblablement aussi une comparaison inter-générique qui doit pouvoir permettre de déterminer des caractéristiques positives et négatives des genres les uns par rapport aux autres.

Par ailleurs, si dans certaines situations, la conformité aux modèles du genre est de mise, tandis que dans d'autres, une marge de manœuvre est autorisée, la notion de contrainte demande à être envisagée en référence au sujet et à son activité ; un genre est plus ou moins prescriptif selon que le locuteur se sent contraint ou non de se conformer à tout ou partie de ses caractéristiques, ou, pour le dire autrement, selon qu'il s'estime autorisé ou non à déplacer certaines des règles en usage. Cela peut résulter des modèles de textes eux-mêmes, et des contraintes qu'ils manifestent, mais d'autres facteurs doivent vraisemblablement pouvoir intervenir, notamment l'image que le scripteur a de lui-même : sa légitimité dans le champ, objective ou telle qu'il la perçoit, détermine sans doute en partie l'autorisation qu'il peut se donner de déplacer les contraintes du genre, ou au contraire la nécessité qu'il perçoit de se conformer à ses caractéristiques les plus canoniques. L'analyse de la variabilité du genre en fonction des styles, en tant qu'usages singuliers du genre, et en fonction du statut social du scripteur dans le champ pourrait permettre d'étayer une réflexion de ce type sur les contraintes du genre.

Cela dit, il faut insister à nouveau aussi sur le rôle de l'évaluation sociale relativement à cette question des contraintes pesant sur un genre. Nous avons montré à propos des limites d'un genre et des frontières entre les genres que poser l'appartenance d'un texte à un genre suppose une activité d'interprétation : c'est donc l'évaluation sociale des pratiques textuelles qui décide en dernière instance si les variations sont ou non autorisées par le genre. En matière de production scientifique, et plus spécifiquement d'article de recherche, les instances telles que les comités de lecture et de rédaction représentent un facteur essentiel des contraintes du genre et de son degré de prescriptivité. Moins rigides probablement que pour le cas du formulaire administratif par exemple, les modèles d'écriture qui définissent les attentes génériques exercent cependant comme nous l'avons souligné une forte contrainte sur le genre, en décidant de ce qui peut être publié en tant qu'article de recherche, et en allant jusqu'à impliquer que le scripteur retravaille son article dans le sens des demandes qui lui sont faites par l'évaluateur.

L'évaluation sociale, en tant que contrainte du genre, rejoint en vertu du dialogisme la dimension interlocutive et interdiscursive de la production verbale. La contrainte est d'écrire avec le déjà-là des normes du genre et les attentes sociales, et avec les évaluations, et c'est donc dans son rapport à l'autre que se joue pour le locuteur l'adoption et l'adaptation des modèles de textes dont il a fait l'expérience et du texte évaluatif. Ainsi, l'évaluation sociale est en elle-même contraignante, mais le genre est contraint également par la manière dont se joue, dans la production du texte, l'interprétation par le scripteur des attentes sociales quant aux modèles du genre.

Les contraintes externes au texte ou la question du contexte

Ce double statut externe et interne des contraintes liées à l'évaluation sociale d'un genre constitue un aspect fondamental de la réflexion à mener sur la situation de communication, le contexte ou les conditions de production, qui caractérisent un genre dans sa dimension socio-historique : nous avons dit en effet qu'on ne saurait se satisfaire de l'idée selon laquelle un genre de texte correspond à une « situation de communication donnée », ou à un type de contexte.

D'emblée se manifestent les limites des termes de « situation » et de « contexte ». Ceux-ci en effet occultent le rapport du locuteur à ce contexte, et l'effet de ce rapport sur les pratiques discursives, et posent le problème de ne pas problématiser le rapport entre le texte et un tel « extérieur » : la dimension sociale notamment n'est pas qu'une composante externe aux textes, car les textes sont en eux-mêmes des objets sociaux.

Nous nous appuyerons sur les apports de divers travaux relatifs à cette question du contexte ou de la situation de communication, pour examiner ce faisant deux problèmes qui font obstacle à la construction du contexte en tant qu'observable pour le linguiste : le problème de son statut d'extérieur au texte, et celui de son caractère non réitérable ou de sa singularité.

Les approches énonciatives, qui par delà leur diversité ont en commun l'étude des énoncés rapportés à l'acte d'énonciation dont ils sont le produit, se sont naturellement posées cette question de la situation de communication. A l'instar de l'acte d'énonciation lui-même, la situation de communication, c'est-à-dire le contexte effectif de l'acte d'énonciation, est nécessairement unique, non réitérable, et par là même difficilement observable. Mais, fondées sur l'hypothèse que tout dans l'énonciation n'est pas individuel, chaotique, mais qu'il existe un invariant à travers la multiplicité des actes d'énonciation, les approches énonciatives cherchent à théoriser la situation de communication. Elles vont ainsi l'aborder en termes de situation d'énonciation, circonscrite en tant que système de trois coordonnées qui définit toute situation de communication : un temps, un lieu et des actants.

Cependant, il importe de noter qu'un certain nombre d'auteurs ont bien mis en évidence les limites de cette conception physique de la situation d'énonciation, pour rendre compte des enjeux de l'acte d'énonciation. Comment en effet créditer l'idée que la situation est commune aux actants de l'échange ? Quand bien même ceux-ci interagiraient en face à face, il est clair que chacun se représente l'autre et lui-même selon un jeu complexes d'images, comme l'a montré notamment M. Pêcheux (1969).

Dans la reformulation du schéma de la communication qu'elle propose, C. Kerbrat-Orecchioni (1980) revient sur ces images en montrant que les données dites situationnelles ne sont pertinentes que sous la forme d'« images », de représentations que les sujets énonciateurs s'en construisent, et que ce sont ces images, et non la situation en elle-même qui contraignent la production verbale. Elle détaille ces images que se font les actants de l'échange en ces termes : images d'eux-mêmes, de leur discours, du support de leur discours, de la langue qu'ils utilisent, du destinataire, de la réalité sociale et physique.

Sérier ainsi les images ne va pas sans problèmes, car celles-ci se construisent en étroite interrelation : les représentations de la réalité sociale déterminent les autres, lesquelles contribuent cependant à les façonner ; les images du destinataire déterminent celles que l'on a de soi et de ses pratiques ; mais réciproquement, les images de soi dépendent de ce qu'on pense de l'image que se fait l'autre de soi, de sorte que ces images dans leur ensemble se bousculent selon un enchevêtrement complexe. La question de savoir si le contexte peut être théorisé surgit alors à nouveau.

Le parti pris des approches énonciatives, (telles qu'elles ont été initiées en particulier par E. Benveniste, 1966, 1974), consistant à traiter en termes de situation d'énonciation le contexte physique des échanges, répond partiellement au problème de l'unicité de la situation de communication effective, et de la singularité de chaque acte d'énonciation ; cette réponse partielle est rendue possible par la réduction du contexte à sa dimension physique, et l'exclusion corrélatrice de certaines dimensions contextuelles au statut plus problématique, telles que la dimension sociale, culturelle voire institutionnelle des échanges.

Cette série d'exclusions constitue la première raison pour laquelle réduire le contexte des échanges à leur dimension physique paraît peu satisfaisant ; la seconde raison, de toute importance selon nous, tient précisément à l'extériorité de la situation ainsi conçue : le pouvoir explicatif des trois coordonnées physiques en termes d'effets sur les textes risque fort d'être limité ; il permet cependant de rendre compte de types différents de textes en fonction de leur ancrage dans leur situation d'énonciation au moyen des déictiques, ou de leur autonomisation. En contrepartie, ces unités déictiques, ainsi discriminées pour leur référence dite situationnelle, légitiment l'approche du contexte sous l'angle des trois coordonnées physiques de la situation d'énonciation.

Les approches sociolinguistiques, qui se développent autour des années 1960-1970 ont également à affronter ce problème de l'articulation entre les usages linguistiques et la dimension sociale de ces usages, car tel est bien leur objet propre. B. Bernstein (1975) inaugure les travaux visant cette mise en relation, en montrant les différences d'usages linguistiques chez les enfants en fonction de leur classe sociale. Or, une grande partie des critiques formulées à l'égard de sa théorie du handicap linguistique et cognitif tiennent précisément au caractère réducteur du dispositif mis en place pour rendre compte des modes de corrélation entre usages linguistiques et contexte social. En effet, la situation de communication en particulier n'est pas prise en compte, et les variables linguistiques sont elles-mêmes sujettes à caution ; en outre, les indicateurs sociologiques se limitent à l'appartenance de classe (ouvrière *versus* bourgeoise), laquelle ne permet pas d'appréhender dans sa complexité la socialisation des enfants. C'est à l'encontre d'une vision mécaniste de la détermination des usages de la langue par des causes sociales externes qu'ont en commun de s'orienter les travaux de W. Labov (1976, 1978). Ses différentes enquêtes sur la variation sociale et la variation stylistique ou situationnelle montrent l'importance d'associer aux variables linguistiques et sociales l'évaluation subjective des variables, autrement dit de tenir compte des représentations que les locuteurs se font des usages du groupe auquel ils s'identifient.

Des travaux de W. Labov, nous retiendrons précisément ce fait que les variables objectivées, saisies d'un point de vue externe par l'observateur, sont analysées en lien avec la manière dont le locuteur se représente les usages linguistiques et la dimension sociale dans leur articulation. Cette démarche mise en œuvre dans les enquêtes de W. Labov corrobore d'une part l'hypothèse que la situation telle que se la représente le locuteur, c'est-à-dire avant tout en tant que pratique sociale, détermine les usages linguistiques ; elle montre d'autre part que les usages linguistiques jouent un rôle déterminant dans les représentations que se font les locuteurs de la situation.

On peut alors rapprocher cette idée de l'importance des genres comme contrainte de la production verbale, au titre de modèles intériorisés par le locuteur dans sa pratique des textes en usage. Par ailleurs, le traitement causaliste du contexte ne suffit pas à rendre compte de la complexité des liens entre le linguistique et le non linguistique : le non linguistique exerce une contrainte sur le linguistique, qui fonctionne lui-même comme un élément structurant le contexte situationnel et social.

Vers une problématisation de l'articulation texte/ contexte : le genre

La question que pose la sociolinguistique, celle du mode de mise en relation de la dimension linguistique avec la dimension sociale constitue un problème épineux en linguistique. Ce problème, qui explique notamment les difficultés de statut institutionnel qu'a pu rencontrer la

sociolinguistique par rapport à la linguistique, est celui du rapport entre la linguistique et ses extérieurs, ou entre la linguistique et l'extralinguistique.

Or, ce principe dichotomique d'extériorité et d'intériorité, censé rendre compte du problème des observables en linguistique, nuit à une problématisation de l'articulation entre les deux termes ; c'est ce que montre G. Kleiber (1994) à propos du problème du sens, comme nous l'avons évoqué plus haut.

Le modèle externaliste du contexte et l'explication causaliste des effets du contexte sur le texte ne vont pas sans rappeler les approches classificatoires des genres : en cherchant à mettre à jour des systèmes de correspondances bi-univoques entre les textes et les fonctionnements sociaux, celles-ci traduisent comme nous l'avons dit une vision trop mécaniste du lien entre les deux dimensions, laquelle ne permet pas de rendre compte des pratiques textuelles effectives. La question du couplage entre l'étude des textes et l'étude de leur dimension sociale et historique demande à prendre comme point de départ l'idée que les textes sont en eux-mêmes des objets sociaux et historiques.

Problématiser l'articulation entre le texte et ses extérieurs oblige en premier lieu à aborder les extérieurs au texte en tant qu'éléments constitutifs de sa production, à l'instar de ce que propose J.P. Bronckart (1985) avec la notion de « pertinence contextuelle » : les éléments pertinents du contexte sont ceux qui conditionnent effectivement la production d'un texte. C'est la raison pour laquelle nous préférons aux termes de contexte et de situation de communication celui de conditions socio-historiques de production.

De là, à l'instar des propositions de J.P. Bronckart (op.cité) nous aborderons les conditions socio-historiques de production d'un texte en tant que telles dans la mesure où elles exercent une influence observable sur les textes produits. C'est donc l'étude des textes qui permet de décider de la pertinence des éléments contextuels. En ce sens, les conditions socio-historiques de production d'un texte constituent bien un objet linguistique.

Se pose alors la question du statut des éléments contextuels à retenir pour tester leur pertinence par l'étude des textes. Les travaux de G. Kleiber (1994) plaident en faveur de la reconnaissance du statut cognitif du contexte, à l'encontre notamment d'une conception de la situation de communication en termes d'environnement physique des échanges. Les approches énonciatives et sociolinguistiques que nous avons évoquées, et tout particulièrement les travaux de W. Labov (1976, 1978), appellent à une prise en compte du statut double que revêt le contexte : il s'agit d'associer à la description d'éléments contextuels au statut externe, objectivables à l'aide de paramètres d'ordre sociologique, l'étude de la manière dont les locuteurs interprètent la situation de communication et la pratique sociale où ils sont engagés.

Pour J.P. Bronckart également (1996), les représentations que se font les locuteurs de la situation constituent une dimension essentielle de la production verbale. Mais il ajoute à cette dimension les représentations du genre : selon lui, la production verbale est conditionnée par ces deux ordres de représentations que constituent les représentations de la situation et les représentations du genre :

la production de tout texte peut être définie comme la mise en interface des représentations relatives à une situation d'action langagière singulière avec les représentations relatives aux propriétés déjà-là de l'intertextualité (1996 : 336).

Ces deux ordres de représentation, comme le note à juste titre l'auteur, sont relativement inextricables : les représentations du genre consistent indissociablement en des modèles de textes et de situation sociale. Nous ferons à notre tour cette hypothèse que les représentations du genre exercent une forte contrainte sur la production verbale et sur les textes produits, hypothèse que suggère également F. Rastier, qui note que

si l'on cesse de rechercher des causes sociales externes, les conditions de production d'un texte consistent notamment dans les normes de discours et de genre (2001 : 246).

Cette idée que les représentations du genre conditionnent la production de tout texte rejoint le principe, mis en évidence par M. Bakhtine, du double dialogisme constitutif de la production verbale. En effet, les représentations du genre du texte à produire que se fait le locuteur consistent en des modèles textuels indexés à des pratiques sociales, que le sujet s'est forgé dans sa pratique des textes en usage. Ces modèles définissent la dimension intertextuelle de la production verbale, abordée en termes de dialogisme interdiscursif chez M. Bakhtine. D'autre part, comme nous l'avons dit en ce qui concerne l'évaluation des articles soumis à la publication, les représentations du genre s'élaborent en fonction des représentations que le sujet se fait des attentes sociales du genre. C'est la dimension interlocutive de la production verbale qui se trouve mise en évidence ici : non seulement la manière dont le locuteur se représente son destinataire est importante, mais également la manière dont il se représente les attentes de son destinataire en termes de modèles textuels.

A l'encontre de l'idée que les textes sont conditionnés par un extérieur au statut incertain relégué hors des bornes de la linguistique, nous dirons qu'en vertu du dialogisme interdiscursif et interlocutif constitutif de la production verbale, ce sont notamment des normes du genre, consistant en des modèles de texte définis en lien avec des discours (ou des pratiques) qui conditionnent la production de tout texte. Objectives et institutionnalisées, mais aussi subjectives, intériorisées comme telles par le locuteur, les contraintes du genre déterminent la forme que prennent les textes produits ; en d'autres termes aussi, ce sont les normes du genre en tant que modèles textuels indexés à des pratiques sociales qui définissent les contraintes socio-historiques d'un genre.

Le genre se présente donc de nouveau comme un point de départ particulièrement opératoire dans la caractérisation des textes : si l'on ne tient pas compte de cette instance des genres, on risque comme nous l'avons dit de passer à côté d'une instance de régulation de la langue, hors de laquelle la quête d'un système de la langue est mise en péril. D'autre part, on passe à côté aussi d'une instance de régulation des pratiques textuelles, et d'une approche qui tienne compte de leur dimension sociale. L'idée de point de départ montre cependant que l'objectif reste de caractériser les textes dans leur diversité ; si le genre représente une entrée particulièrement intéressante pour rendre compte de la stabilité des usages textuels et de leur variabilité, il importe de rendre compte non seulement de ce qui fait cette stabilité (ce qui est commun aux textes, ou ce qui fait que leur diversité se structure de manière relativement stable), mais aussi de ce qui fait la variabilité (quelles caractéristiques textuelles sont concernées ? En quoi peut-on rendre raison de cette variabilité à partir d'autres entrées que le genre¹², par exemple comme nous le verrons dans notre étude, l'usage singulier que chaque auteur en fait, le statut social du scripteur, etc.).

Le point de vue des usagers du genre pourrait également représenter une entrée intéressante ; il ne s'agit pas cependant de chercher dans ce point de vue les normes du genre elle-même, qui restent textuelles. La question se pose à ce propos de déterminer dans quelle mesure les modèles du genre sont conscientisés : si tout locuteur se réfère aux modèles en sa connaissance dans la production d'un texte, rien cependant ne permet d'affirmer qu'il est apte à verbaliser les normes auxquelles il se conforme dans sa pratique. Une analyse du point de vue qu'ont les usagers sur le genre est utile à rendre compte de cette question de la conscientisation des pratiques ; elle pourrait en particulier montrer comment cette

¹² Si le genre est un point de départ, la prise en compte d'autres entrées pour rendre compte de la diversité des textes ne doit cependant donc pas ignorer la question des genres.

conscientisation varie avec un degré d'expertise, ou d'acculturation à un genre. En outre, comme nous l'avons suggéré avec la sociolinguistique, une telle étude donnant la parole aux usagers peut permettre de spécifier la question de la légitimité du locuteur et de son statut social, non plus seulement en termes objectifs, mais dans la manière dont le locuteur lui-même se représente en tant que tel, i.e. l'image de lui-même qu'il a en tant que locuteur d'un genre donné.

1.3.5. Conscientisation des modèles du genre et compétence générique

Si les pratiques de production verbale sont régulées par les modèles textuels que constituent les genres, tels que se les représente le locuteur, il est vrai également que les usagers classifient spontanément les productions verbales en genres ; mais comme l'a bien montré P. Bourdieu notamment (1994), les sujets sont souvent peu conscients de leurs pratiques, car, s'ils se représentent ces pratiques, ce qui a statut de représentation n'est pas nécessairement conscientisé. Le genre relèverait des savoir faire, tant du côté de l'interprétation que de la production, et les verbalisations des usagers sur les genres permettraient quant à elles de déterminer des degrés divers de conscientisation, et en cela aussi d'expertise vis-à-vis du genre.

Dans le cas de l'article de recherche en particulier, l'évaluateur de l'article doit être celui qui repère dans la matérialité des textes en quoi ceux-ci sont homologues aux attentes génériques ou demandent à être retravaillés pour mieux s'y conformer. Une conscientisation des normes du genre permet de mieux identifier ce qui fait éventuellement problème, et de proposer des solutions adaptées pour y remédier.

Les savoirs sur le genre, développés dans le cadre d'entreprises descriptives sont nécessaires pour que le genre soit constitué en objet d'enseignement (Schneuwly, 2001). A défaut d'analyses empiriques des genres, leur enseignement peut comporter un risque de confusion voire de décalage entre ce qui relève de caractéristiques supposées du genre (des normes du genre au sens de prescriptions verbalisées en tant que telles) et de caractéristiques effectives (les normes du genre observables dans les textes).

La question de la conscientisation des pratiques langagières représente un objectif majeur en didactique : l'idée est que c'est elle qui rend possible une amélioration des textes en fonction de leur objectif communicatif et d'abord aussi un repérage, dans l'acte de lecture, des caractéristiques d'un genre.

Pour analyser les ressorts de la compétence générique, le modèle de la compétence scripturale proposé par M. Dabène (1987, 1991) pour rendre compte des compétences que mobilise le locuteur dans ses activités langagières de production et d'interprétation d'écrits nous semble intéressant : M. Dabène (1991) montre que la compétence scripturale est structurée par trois composantes, que sont les savoirs, les savoir faire, et les représentations. Bien que M. Dabène ne traite pas exclusivement dans ce modèle la question de la dimension générique des pratiques scripturales, ces trois composantes peuvent par hypothèse être considérées comme essentielles en matière de genre : la compétence générique se situerait bien à ces trois niveaux, le genre ayant statut de représentation pour les usagers, fonctionnant au niveau des savoir faire, et pouvant également faire l'objet de savoirs de la part des usagers.

Le modèle de la compétence scripturale proposé par M. Dabène permet ainsi d'avancer dans une perspective d'analyse du genre. Si le genre repose sur la caractérisation des textes, la question des compétences génériques pourrait appeler à tenir compte des variations observables dans les textes en fonction des usagers, et des savoirs qu'ils ont du genre. Dans une telle optique, il convient cependant de se garder d'une hiérarchisation des

caractéristiques des textes en termes d'expertise¹³ ; cela revient aussi à dire comme le souligne l'auteur du modèle qu'il importe de ne pas analyser la question des compétences langagières en fonction d'un scripteur-lecteur idéal (1978 : 40). Par ailleurs, et c'est là encore un point sur lequel insiste M. Dabène, il importe de ne pas rapporter la diversité saisie à des variables externes de manière mécanique. Les différences entre les usagers dans leurs pratiques de l'écrit doivent selon lui être prises en compte « dans le modèle explicatif de la compétence, comme une variable intrinsèque et non hiérarchisée » (ibid.).

Une analyse du genre fondée les seuls textes ne permet pas à elle seule de résoudre la question des compétences génériques, mais a ceci d'intéressant par rapport à cette question qu'elle permet de les décrire dans la matérialité des textes. Si les textes attestent d'effets de variables externes et notamment sociologiques, une analyse prenant comme point de départ le genre doit aussi montrer l'irréductible diversité des textes : leurs caractéristiques ne se laissent vraisemblablement pas toutes analyser en termes de degré d'expertise, et le facteur sociologique n'épuise pas cette diversité car d'autres facteurs doivent entrer en jeu.

En abordant la question de la compétence scripturale en termes de compétence générique du point de vue de la production verbale, nous nous démarquons de la démarche de M. Dabène, mais réinvestir ce modèle nous semble toutefois favoriser l'analyse de la diversité intrinsèque au genre de l'article. S'il ne s'agit pas dans notre optique de mettre au jour des profils de compétence scripturale, la réflexion sur ces profils chez M. Dabène peut nous semble-t-il éclairer la question de profils diversifiés de textes, et en particulier l'impossibilité de hiérarchiser de tels profils, et la nécessité de les envisager de manière multi-critériée, en fonction de variables externes elles-mêmes multiples

Elle montre aussi l'intérêt qu'il pourrait y avoir, à terme, à tester l'hypothèse qu'il existe des différences dans la conscientisation des pratiques d'écriture du genre de l'article, témoignant de degrés divers de maîtrise du genre. Il serait alors important d'observer en quoi les différences observées en termes de savoirs déclaratifs sur le genre font écho à des différences du point de vue des textes produits, ce qui permettrait de fournir une prise sur la question des relations entre les savoir et les savoir faire dans les pratiques d'écriture.

L'approche théorique de la notion de genre permet de vérifier la pertinence théorique de la notion, en particulier dans une perspective didactique où le genre se présente actuellement comme une entrée centrale pour l'enseignement et les apprentissages langagiers. Elle permet également de progresser vers une définition opératoire du genre, et d'étayer la question qui oriente notre étude, celle de la caractérisation du genre, et du genre de l'article en particulier. Elle ouvre des pistes qui dépassent les objectifs de notre analyse, mais permettent d'envisager une prolongation à celle-ci, avec notamment la prise en compte conjointe des textes et des verbalisations des scripteurs sur leurs pratiques. Pour avancer encore dans la problématique des genres, pour l'analyse de l'article de recherche en particulier, un dernier aspect de la réflexion bakhtinienne sur la notion mérite notre attention : la distinction entre genres premiers et genres seconds. Largement réexaminé à sa suite, elle paraît *a priori* discutable, mais, mise en perspective avec d'autres notions, elle fournit à titre programmatique des pistes de réflexion intéressantes pour l'analyse des textes en tant qu'actions.

¹³ Ce qui rejoint le problème posé en sociolinguistique en particulier de variétés hautes ou basses de la langue.

1.4. La secondarité de certains genres : des textes à l'action

Selon M. Bakhtine, l'article de recherche, qui relève du discours scientifique, ferait partie à ce titre des genres seconds, ou complexes, qu'il distingue des genres premiers, qualifiés aussi de simples.

Les genres seconds du discours (roman, théâtre, discours scientifique, idéologique) apparaissent dans les circonstances d'un échange culturel (principalement écrit), artistique, scientifique, sociopolitique, plus complexe et relativement plus évolué » (Todorov, 1981 : 267).

Les genres premiers, en revanche, sont décrits comme le fruit de l'échange quotidien.

Il faut se défendre en premier lieu d'interpréter cette proposition de M. Bakhtine comme une nouvelle tentative taxinomiste : bien que la distinction bakhtinienne fonctionne comme un principe qui permet d'ordonner l'hétérogénéité des genres, l'enjeu ne consiste pas à décider d'une frontière entre genres premiers et seconds, entreprise qu'il serait peu réaliste de mener. Au contraire, M. Bakhtine met l'accent sur le continuum qui existe entre genres premiers et seconds du point de vue de leur constitution.

Ainsi, ce qu'il nous paraît important de retenir, c'est moins peut-être la distinction elle-même que l'idée de secondarité : les genres seconds recourent à des formes de discours déjà établies pour les intégrer dans une nouvelle instance discursive, avec de nouvelles fins communicatives ; notons également que la relation entre les deux genres n'est pas univoque : le fait de disposer de genres seconds modifie aussi les échanges les plus quotidiens.

M. Bakhtine, prenant pour exemple un de ses objets d'étude privilégiés, le roman, développe cette idée de secondarité :

« Au cours du processus de leur formation, [l]es genres seconds absorbent et transmutent les genres premiers (simples) de toutes sortes, qui se sont constitués dans les circonstances d'un échange verbal spontané. Les genres premiers, en devenant composantes des genres seconds, s'y transforment et se dotent d'une caractéristique particulière : ils perdent leur rapport immédiat aux énoncés existants et au réel des énoncés d'autrui ». (Todorov, 1981 : 267).

Cette dernière remarque appelle quelques précisions: les genres premiers, en devenant composantes des genres seconds, ne perdent bien entendu pas tout enjeu communicatif, mais se dotent d'une nouvelle visée communicative, en se trouvant inscrit dans une nouvelle instance discursive qui leur confère un autre statut.

La secondarité tient donc à cet emprunt aux genres premiers dans des échanges qui participent de ce que P. Bourdieu, dans une perspective sociologique, traite en termes de « champ » culturel (1979, 1994). La perte d'immédiateté dans le rapport au réel que pointe M. Bakhtine peut alors être interprétée en lien avec ce que dit P. Bourdieu de ces champs : ils induisent une certaine autonomisation du discours par rapport au contexte socio-historique de sa production.

Cette propriété des champs de la production culturelle pointée par P. Bourdieu fait écho aux caractéristiques du discours scientifique : un de ses traits linguistiques réside dans cette autonomisation, facteur d'objectivation. B. Schneuwly développe cette caractéristique des genres seconds et ses conséquences d'un point de vue linguistique : en raison de leur autonomisation par rapport au contexte de leur production, les genres seconds, « *non contrôlés directement en situation* », appellent des « *mécanismes de contrôle autres et plus puissants* » pour répondre à la nécessité de clôture interne du texte (1994 :163). B. Schneuwly insiste ainsi sur l'idée que ce sont des moyens linguistiques, assurant un mode de référence à

un contexte linguistiquement crée, qui assurent la cohésion interne des genres seconds et leur autonomie par rapport au contexte.

En ce sens, les genres seconds ne vont pas sans évoquer ce que E. Benveniste (1874) traite en termes d'énonciation historique, pour définir les énoncés coupés de leur situation d'énonciation : en raison de cette propriété, ces énoncés sont marqués notamment par une absence tendancielle de référence de type déictique (ou référence relative aux paramètres physiques de la situation d'énonciation que sont les acteurs, le temps et le lieu). Leur clôture interne est garantie alors par un système complexe de reprises anaphoriques, et notamment par ce que D. Maingueneau (1994) traite en termes de deixis interne.

L'article de recherche, en tant que genre second, a donc ceci de particulier qu'il entretient un rapport médiatisé par le champ culturel et non un rapport immédiat ou direct avec sa situation de production, et avec l'expérience de la vie quotidienne. Une telle caractéristique peut intéresser une dimension souvent convoquée dans les analyses programmatiques des genres, et diversement traitée en termes d'ethos ou de rôles discursifs (Maingueneau, 1992, Branca-Rosoff, 1999), ou encore d'image de l'auteur ou de personnage du locuteur (Bakhtine, 1984).

En vertu de cette propriété d'autonomisation, la personne instanciée par le *je* dans le texte, à savoir l'auteur de l'article, ne coïnciderait pas avec le sujet de la vie quotidienne ; la question de savoir « qui parle ? », centrale comme le souligne M. Bakhtine dans la définition d'un genre, montre avec cette idée d'autonomisation l'intérêt d'être traitée en termes de « rôles », et suggère l'intérêt qu'il y aurait, à terme, à envisager dans leur continuité les approches en analyse du discours que D. Maingueneau (1992) en particulier distingue selon qu'elles s'intéressent à retrouver dans les textes la trace d'un rôle prédéfini par les formations sociales, ou à les analyser sous l'angle des genres et des rôles qu'ils induisent.

Pour revenir à la distinction entre genres premiers et seconds, et avancer dans la question de la secondarité et de ses possibles implications pour l'analyse d'un genre et du genre de l'article en particulier, soulignons que D. Maingueneau (1995) propose une autre distinction, relativement conforme nous semble-t-il à celle formulée par M. Bakhtine : elle oppose les genres conversationnels et les genres qu'il nomme routinisés ou institutionnalisés.

Notons d'abord que terme de conversation est ambigu et mérite selon nous quelques explications : ce qui est appelé conversation dans une société donnée correspond bien, comme le montre F. Grossmann (1996 : 10), à un genre extrêmement codifié, proche en cela des genres seconds.

Toutefois, ce que D. Maingueneau traite en termes de genres conversationnels intègre le type d'échanges que prennent pour objet les analyses pragmatiques conversationnelles : il s'agit, le plus souvent, d'interactions quotidiennes en face-à-face, qui à ce titre se satisfont de la description que donne M. Bakhtine des genres premiers. On retrouve donc dans la définition que donne D. Maingueneau de ces genres le critère sur lequel est fondée de manière prépondérante la différence entre genres premiers et seconds chez M. Bakhtine, à savoir le rapport entre le texte et le contexte de sa production.

En référence à la distinction de D. Maingueneau, l'article de recherche correspond à un genre institutionnalisé, en raison de son inscription dans le champ scientifique, et notamment du dispositif éditorial impliqué dans sa production. La notion de routines dans l'approche de D. Maingueneau permet de compléter ce que nous avons dit du genre de l'article en tant que genre second : les genres institutionnalisés se caractérisent par un ensemble de routines, ce sur quoi s'accordent un certain nombre d'auteurs à propos du discours scientifique ; comme le montre D. Maingueneau, la routinisation concerne en premier chef les rôles discursifs, qui dans ces genres sont pré-établis, très codifiés, et, partant,

relativement stables, à la différence de ce qui se joue dans le cadre de genres conversationnels.

La stabilité du genre, question bien étudiée par M. Bakhtine, est donc clairement rattachée à son caractère institutionnel, de même que la possibilité d'associer à un genre un rôle endossé par celui qui en est l'auteur. Nous serons amenée à développer cette idée en ce qui concerne la question de l'auteur de l'article, qui doit investir par hypothèse un rôle discursif fortement conventionnel, celui de représentant du champ scientifique.

Cela étant, un autre élément de discussion de la distinction proposée par M. Bakhtine mérite notre attention : le fait que les genres seconds sont décrits comme relevant principalement de l'écrit.

En premier lieu, nous avons dit, en référence à B. Schneuwly (1994), que les genres seconds se définiraient par des mécanismes de contrôle interne, ce qui effectivement rejoint une caractéristique de l'écrit : l'écrit entretient un lien privilégié avec les situations d'énonciation différée, lesquelles impliquent de la part du locuteur un contrôle important de son texte. Ce contrôle consiste notamment en une capacité d'auto-évaluation du caractère compréhensible du texte, capacité qui est également qualifiée de manière quelque peu vague de « distance ».

D'autre part, les travaux de J. Goody (1979, 1993), mais aussi de D.R. Olson (1998), mettent en évidence que l'écrit dans notre société constitue un instrument privilégié de ce que M. Bakhtine appelle la transmutation des genres premiers dans les genres seconds (Todorov, 1981 :287) ; si les traditions orales sont davantage marquées par un ancrage dans l'expérience, la trace écrite ouvre en revanche à une distanciation d'ordre cognitif, et rend possible la construction des savoirs scientifiques, propre à l'activité de la science telle qu'elle se développe dans les sociétés fondées sur l'écrit.

Cependant, il est nécessaire de rester prudent face à de telles pistes interprétatives, car le risque est de faire de la différence entre écrit et oral le critère majeur de distinction entre genres premiers et genres seconds : l'oral n'est pas le seul fait des échanges quotidiens, de même que ces derniers n'excluent pas l'écrit, et inversement en ce qui concerne les genres seconds. En d'autres termes, la différence entre oral et écrit n'est pas le seul élément discriminant les conditions de production des genres premiers de celle des genres seconds.

S'il n'est pas possible de se satisfaire d'une définition des genres premiers et seconds sur la base de l'opposition entre oral et écrit, nous avons souligné toutefois l'intérêt que présente l'idée de distanciation et d'autonomisation comme caractéristique des genres seconds. Plutôt que le corrélat de l'ordre scriptural, JP. Bronckart (1996 : 63) propose de voir dans cette distance des genres seconds un trait définitoire des actions langagières, par opposition aux actions non langagières. Selon lui, les genres premiers seraient structurés par une action non langagière, alors que les genres seconds s'en détacheraient pour faire l'objet d'une structuration autonome, spécifiquement linguistique : ils constitueraient de véritables actions langagières.

Cette redéfinition de la distinction bakhtinienne nous semble intéressante pour ce qui a trait à l'article de recherche, dont les enjeux en termes d'action sont d'ordre communicationnels : l'activité scientifique n'existe comme telle qu'au travers des genres de texte en usage dans la communauté discursive que constitue la communauté scientifique, comme le revendique en particulier J. Swales (1990), comme nombre d'approches intéressées en analyse du discours par la question des genres.

Par ailleurs, définir les genres seconds en tant qu'actions langagières permet, avec l'idée que les genres seconds empruntent aux genres premiers, de souligner que dans l'article, peuvent se trouver reconfigurés un certain nombre de discours structurés par des actions non

langagières : l'article peut rendre compte de certains échanges entre collègues, ou encore du dispositif méthodologique mis en œuvre pour mener à bien une observation, ou une analyse. C'est cette propriété de l'article que pointe A. Greimas à propos du discours scientifique en général :

Ce ne sont pas les procédures réelles mises en œuvre par le savant (si tant est qu'elles soient jamais accessibles), mais des élaborations secondaires, des simulacres de procédures observées par un sujet qui s'est dédoublé et qui met en scène la quête du savoir (cité par E. Nonnon, 2002)

Si nous souscrivons à cette idée que le discours scientifique ne correspond pas à la réalité de la science en train de se faire, mais à une représentation langagière de cette réalité, il faut noter aussi que ce n'est pas le seul fait de ce type de discours : il s'agit fondamentalement d'une propriété du langage que de constituer une représentation, ou une mise en scène de la réalité, et par là de configurer et de structurer cette réalité, comme nous avons déjà eu l'occasion de le dire en particulier avec la notion de logosphère dans l'approche praxématique (Détrie et al., 2001).

La notion de mise en scène nous conduit alors à revenir à la question de l'auteur et des rôles que nous avons évoqués *supra* : le sujet scripteur mettrait en scène dans l'article un « double » de lui-même – pour reprendre l'idée de dédoublement dans la citation donnée par E. Nonnon (2002) – ou en d'autres termes une image du chercheur, autonomisée par rapport au chercheur que nous qualifierions d'effectif, qui est le producteur empirique de l'article.

La secondarité de certains genres, telle qu'elle a été mise en lumière par M. Bakhtine, et retravaillée par J.P. Bronckart (1996) en lien avec ce qu'il appelle action langagière et non langagière suggère l'intérêt d'une analyse qui prendrait comme point de départ les genres pour une analyse des textes en lien avec l'« agir » (Bronckart, *ibid.*) : elle montre les textes en tant qu'actions, au service de pratiques ou d'activités¹⁴, et dont l'orientation collective détermine des rôles stables, consubstantiels des genres.

La secondarité et ses implications en termes d'action langagière et de textualité nous semble par ailleurs ouvrir des pistes intéressantes dans une perspective didactique : dans les discours sur les genres scolaires et universitaires sont souvent invoquées les notions d'« élaboration seconde » ou de « distance », tant pour définir ces genres qu'en tant que lieu majeur de difficultés pour les apprenants. Bien entendu, le terme de distance est fortement galvaudé, comme le montre très justement B. Daunay (2002)¹⁵, dénonçant le manque de fondements d'un tel critère d'évaluation des productions écrites en contexte didactique, plus précisément en ce qui concerne le commentaire composé.

Cependant, il n'est pas anodin de voir ce critère sans cesse mis en avant dans les recherches en didactique comme dans les manuels à vocation didactique ; s'il renvoie à la place de l'écrit et aux représentations que l'on s'en fait dans notre société, aborder cette notion de distance sous l'angle de la question de l'auteur et du rôle qu'il endosse dans son texte pourrait faire avancer la réflexion didactique sur les attentes en matière de productions écrites et sur les difficultés des apprenants. Bien souvent, ce qui leur est reproché derrière le constat de leur manque de distance, c'est, comme l'explique E. Bautier (1998), le fait que c'est leur « moi » qui s'exprime dans leurs textes ; or, une telle posture ne correspond pas aux attentes génériques : il est attendu de la part des apprenants qu'ils adoptent une posture auctoriale détachée, ou autonomisée par rapport à ce « moi » du sujet expérientiel.

¹⁴ S'agissant de pistes d'analyse largement programmatiques, nous ne discuterons pas ici des fluctuations terminologiques attestées dans la littérature ni des définitions spécifiques données à chacune de ces notions.

¹⁵ Terme qu'il emploie lui-même dans son titre « Le lecteur distant ».

Au terme de cette discussion de la distinction proposée par M. Bakhtine entre genres premiers et genres seconds, et des échos qu'elle a trouvés chez B. Schneuwly, J.P. Bronckart et D. Maingueneau notamment, diverses pistes d'analyse s'ouvrent à titre de perspective pour une analyse des textes fondée sur la notion de genre. Les genres offrent une prise sur la problématique de l'action et ce que D. Maingueneau (1992) traite en termes de « rôles discursifs » se présente comme une dimension essentielle des genres. Alors que, y compris dans l'approche bakhtinienne, l'idée est de relier les textes aux pratiques où ils sont en usage par l'intermédiaire des genres, on voit se profiler ici l'idée d'une analyse des textes en tant qu'actions, qui envisage les acteurs du champ en termes de rôles caractéristiques du genre.

Nous allons voir au prochain chapitre que dans les analyses empiriques ou grammaticales, intéressées par les genres ou par les écrits de recherche, la notion de rôle intervient, et s'associe à d'autres notions qui posent de manière centrale la question de l'image de soi du scripteur dans son texte, ou en d'autres termes d'une figure de l'auteur. Nous tenterons de montrer l'intérêt d'une telle entrée pour l'analyse du genre de l'article, sans prétendre qu'elle soit suffisante à elle seule, et chercherons à déterminer ses fondements linguistiques, et plus spécifiquement énonciatifs. Auparavant, le bilan de ce chapitre poursuit l'objectif, amorcé dans le parcours théorique de la notion, d'une définition opératoire des genres ; il s'agit de s'intéresser aux approches existantes dans la littérature, qui visent une caractérisation empirique des genres, et du genre de l'article de recherche en particulier.

1.5. En guise de bilan : la caractérisation empirique des genres, et du genre de l'article de recherche

L'importance des genres de textes dans les pratiques langagières en fait un lieu d'entrée privilégié pour analyser ces dernières dans leurs régularités tout autant que dans leur diversité. Aborder les pratiques langagières sous l'angle de la problématique des genres revient, à la suite de F. Rastier (1996), à poser les textes en tant qu'objet empirique de la linguistique. Si tous les niveaux de la textualité varient avec les genres et que tous sont *a priori* pertinents pour caractériser un genre en particulier, l'étude d'un genre ne saurait s'établir au seul niveau du mot ou de la phrase ; du moins les caractéristiques linguistiques de ces derniers doivent-elles être envisagées en regard des textes et de leurs enjeux.

Les termes employés dans un texte littéraire ou dans un texte spécialisé ne sont sans doute pas littéraires ou spécialisés en eux-mêmes, mais leur usage dans ces textes invite à les lire de la sorte, et peut par suite déterminer le caractère littéraire ou spécialisé de certains de leurs emplois. Pour prendre un autre exemple, la pratique de la citation peut constituer en elle-même un objet d'étude, mais s'il s'agit de décrire les genres universitaires pour accompagner les étudiants entrant dans l'écrit de recherche, l'intérêt pour la citation se justifie par sa spécificité et sa fonctionnalité dans ces écrits.

Nombre d'études s'attachent ainsi à certaines caractéristiques d'un genre ou d'un discours. A propos de l'écrit scientifique, on peut penser notamment, en contexte français, aux travaux du Cediscor, centre de recherches sur les discours ordinaires et spécialisés¹⁶, relevant du laboratoire Syled (Systèmes Linguistiques, Énonciation et Discursivité) de l'Université Paris 3, et rattaché au Ceditec (Centre d'études des discours, images, écrits, communications) de l'Université de Paris 12. Le Cediscor s'intéresse aux discours de transmission des connaissances (discours de recherche, de vulgarisation, d'enseignement, de formation, discours des médias, discours professionnels), à travers des objets linguistiques susceptibles de rendre compte de ces discours comme de leurs réseaux de relations, notamment avec des travaux sur la reformulation menés par le Syled dans les années 1990.

¹⁶ Cf. <http://www.univ-paris12.fr/www/labos/ceditec/cediscor.html>

Les études sur les genres et les discours peuvent se centrer sur un phénomène linguistique, mis en regard de sa fonctionnalité dans le genre ou le discours concerné ; privilégier un niveau de description a cependant pour conséquence de ne pas rendre compte du genre dans son ensemble¹⁷. Une autre perspective dans l'analyse des genres vise au contraire à multiplier les dimensions prises en compte. L'objectif est alors plus nettement de caractériser un discours ou un genre, de manière multi-critériée. La caractérisation se veut généralement différentielle : il s'agit de mieux faire ressortir le propre d'un genre par contraste avec d'autres et de décrire un genre en termes de traits communs aux textes et de traits singuliers. L'idée est que le genre se définit par « un faisceau de critères » (Rastier, 2001 : 253), et l'analyse des textes va trouver dans la linguistique de corpus et dans l'analyse statistique de corrélations des outils particulièrement adaptés pour l'exploration des textes et la définition de genres.

Nous ne fournissons pas ici un aperçu de l'ensemble des travaux relatifs à un genre, ni de l'ensemble des travaux relatifs au genre de l'article de recherche. A partir des éléments que nous venons de mentionner quant aux manières d'aborder un genre, nous proposons d'évoquer quelques approches empiriques des genres qui nous semblent pouvoir intéresser la nôtre. Nous progressons ainsi vers la démarche retenue, qui consiste à prendre le genre comme point de départ, et à multiplier les dimensions du texte prises en compte pour caractériser les textes dans leurs traits communs et dans leur diversité.

L'analyse multi-critériée des textes ne va pas d'abord sans évoquer les travaux de D. Biber (1988, 1989 et 1993). Il est d'usage de considérer ces travaux comme pionniers dans l'exploitation des statistiques multidimensionnelle pour l'analyse de types de textes¹⁸. L'ouvrage fondateur de 1988, *Variation across Speech and Writing*, présente une approche inductive qui vise à établir une typologie textuelle sur la base de critères linguistiques. En d'autres termes, les types de textes sont considérés comme des agglomérats de traits linguistiques et il s'agit de rendre compte de la répartition et des co-occurrences de traits linguistiques pour voir lesquels discriminent et caractérisent des types de textes. L'objectif est d'inclure un grand nombre de caractéristiques linguistiques, représentant l'éventail des possibilités fonctionnelles de l'anglais (Biber, 1988 : 211). 67 traits linguistiques, de type morpho-syntaxiques, sont retenus, selon en 16 catégories (marqueurs de temps et d'aspect, pronoms, questions, passifs, modaux, coordination, négation, adverbes et locutions adverbiales de temps et de lieu, etc.).

Ces traits sont étudiés dans des textes issus de deux grands corpus pour l'anglais, le « Landcaster-Oslo/Bergen Corpus of British English » (ou LOB¹⁹, de 1000000 mots et divers genres de l'écrit) et le « London-Lund corpus of Spoken English »²⁰ (87 textes d'anglais parlé). Les textes sont normalisés pour être de taille comparable (1000 mots), ce qui induit une forme de biais dans l'analyse, dans la mesure où, plutôt que d'étudier les textes, elle consiste au final à étudier des séquences qui correspondent à des débuts de textes. Le corpus

¹⁷ De telles approches peuvent cependant permettre la mise au jour de types de textes sur la base du critère choisi, comme c'est le cas par exemple avec les approches systémiques du genre, développées sur la base de la grammaire fonctionnelle (Halliday, 1985). Il s'agit de rendre compte de l'organisation du système sémantique d'un texte (sa texture) à partir de la contribution sémantique des parties d'un texte, et en premier chef de l'unité phrase. On note donc une extension au texte des théories et des méthodes de l'analyse syntaxique, par exemple pour l'étude de *The Language of Science* (Halliday, 2004).

¹⁸ Il faut souligner que d'autres études menées à la même époque se situent dans une optique tout à fait comparable, en particulier les travaux de J.P. Bronckart sur *Le fonctionnement des discours* (1985) et la mise en évidence de types linguistiques de textes sur la base d'un certain nombre de variables linguistiques. Ces types linguistiques sont mis en perspective avec 4 architypes énonciatifs de textes.

¹⁹ Cf. http://clwww.essex.ac.uk/w3c/corpus_ling/content/corpora/list/private/LOB/lob.html

²⁰ Cf. J. Svartvik (1990).

fait également l'objet d'un étiquetage automatique manuellement vérifié pour rendre possible l'étude des 67 traits. Il est ensuite soumis à une analyse factorielle (qui détermine des facteurs linguistiques de différenciation des textes, i.e. des groupements de traits linguistiques corrélés). Avec une entreprise de classification automatique, l'étude aboutit à la mise en évidence de huit types de textes ainsi désignés : « intimate interpersonal interaction », « informational interaction », « scientific exposition », « learned exposition », « imaginative fiction », « general narrative exposition », « situated reportage », « involved persuasion ».

Ceux-ci sont définis par rapport à 5 dimensions fonctionnelles à l'aune desquelles D. Biber interprète les facteurs linguistiques de différenciation des textes : le caractère impliqué ou informationnel de la production, l'orientation narrative ou non narrative, la référence explicite ou relative à la situation d'énonciation, la visée persuasive explicite et le style abstrait. Dans ce cadre, le texte académique est décrit comme un texte informatif, non-narratif, abstrait et explicite et se situerait enfin à un degré moyen en ce qui concerne la persuasion.

L'originalité de la démarche de D. Biber tient à l'exploitation des possibilités nouvelles offertes par le développement de l'informatique. La numérisation des corpus, les outils d'étiquetage automatiques et le recours aux statistiques rendent possible la mise en œuvre d'une démarche inductive. L'intérêt est de caractériser des types de textes sur la base de corrélations de traits, et de s'intéresser à des données linguistiques autres que le lexique. Les types de textes mis en évidence correspondent pour partie à une approche inductive des usages langagiers : ils corroborent l'idée d'une différence linguistique entre la fiction et l'exposé scientifique. Certaines distinctions permettent par ailleurs d'affiner ou de refonder ces intuitions, notamment celle au sein de l'interaction, entre une interaction à visée informative et une interaction à visée interpersonnelle.

Au final, D. Biber prend pour objet la caractérisation linguistique de registres, en tant qu'emplois du langage définis situationnellement ou fonctionnellement, mais non des genres : il discute lui-même la proximité des deux notions (1995) mais montre que le genre relève de catégorisations sociales, alors que les registres qu'il entend mettre en évidence sont exclusivement fondés sur des variables linguistiques interprétées en termes de dimensions fonctionnelles et situationnelles. Si le caractère inductif de la démarche en fait tout l'intérêt, elle en signale aussi les limites. D'une part, le fait que les outils de linguistique de corpus permettent l'analyse de catégories morpho-syntaxiques ne doit pas masquer que de telles catégories n'ont rien de naturel, et que leur étude doit résulter de l'hypothèse qu'elles sont pertinentes pour la mise en évidence de types de textes, comme le souligne S. Branca-Rosoff (1999 :15-16). Par ailleurs, il serait intéressant de confronter la typologie de textes obtenue au final avec un classement initial, car le choix des textes entrant dans le corpus ne peut là encore être considéré comme allant de soi.

Les travaux de D. Biber ont permis d'initier une démarche reprise dans nombre d'entreprises de profilage de textes : il s'agit de dresser des typologies textuelles de manière inductive, à partir de constellations de traits linguistiques ; les types de textes sont ainsi définis comme des « agglomérats fonctionnellement cohérents de traits linguistiques » (Habert, 1999). Les méthodes typologiques inductives se présentent comme une entrée particulièrement intéressante pour l'analyse des genres, mais demandent à être repensées au vu de la spécificité de l'objet traité.

J.P. Bronckart (1996 :77) ou encore F. Rastier (2001) s'accordent sur cette idée, thématisée par D. Biber avec la notion de registre, que le genre ne peut être défini sur la base des seuls critères linguistiques ou, plus largement, sur « des critères définis indépendamment des genres (comme oral vs écrit, public vs privé, etc.) » (Rastier et Malrieu, 2001).

Le point de départ de l'analyse d'un genre ne peut se situer dans du texte tout-venant. Le corpus doit être constitué et structuré en référence aux genres, et l'étude inductive des propriétés linguistiques des textes doit alors être confrontée à ce classement initial. Le parti pris, explicité en particulier dans D. Malrieu et F. Rastier (2001) est donc doublement déductif et inductif : si le corpus est classé en genre, il ne s'agit pas simplement de mener une démarche déductive puisque « la linguistique de corpus [...] doit permettre de refonder ou d'affiner les distinctions intuitives et empiriques des genres », en particulier au niveau morpho-syntaxique, qui s'avère effectivement productif pour la discrimination de genres (les auteurs donnent entre autres exemples les manuels et les brochures commerciales au sein du domaine des textes techniques). Cela étant, l'analyse inductive des caractéristiques des textes est une entrée pour l'analyse, mais c'est l'interaction normée de composantes textuelles qui définit le genre, et non le seul niveau morpho-syntaxique que la linguistique de corpus permet d'étudier empiriquement et en grande partie automatiquement.

L'étude menée par D. Malrieu et F. Rastier (op. cité), sur 2500 textes classés par genres, et analysés à l'aide de 251 variables morpho-syntaxiques suggère tout l'intérêt de ce type de démarche pour la comparaison inter-générique, et la caractérisation d'un genre en particulier. L'étude de C. Poudat (2006a) s'inscrit dans ce cadre, mais se centre sur l'article scientifique en linguistique en proposant d'une part de le situer par rapport à d'autres genres, d'autre part d'en décrire les propriétés, au niveau morpho-syntaxique, au niveau de la structure des textes et enfin au niveau lexical et thématique. Les caractéristiques des textes sont mises en évidence statistiquement en termes de caractéristiques centrales *versus* marginales (les variables qui définissent le genre par leur rareté voire leur exclusion tendancielle), et les analyses statistiques permettent également de cerner les axes d'organisation interne du genre (les facteurs linguistiques qui structurent sa diversité), les configurations optatives (avec notamment la question des exemples dans l'article), et les facteurs de variabilité du genre (la langue, les styles d'auteur ou usages singuliers du genre²¹, la revue ou le numéro etc.).

De telles études doivent au final permettre de mieux décrire le fonctionnement linguistique des textes et des réseaux de textes dans lesquels ils entrent. Elles montrent aussi la complexité de la question de l'induction et de la déduction dans la démarche d'analyse des genres : si le niveau morpho-syntaxique se prête à des analyses inductives en raison de l'opérationnalité des outils d'annotation à ce niveau (Véronis, 2000), il est peu envisageable de penser pouvoir aborder un genre indépendamment de toute hypothèse sur ses caractéristiques. Dès lors que l'on tente d'affiner l'annotation automatique, on est conduit à faire des choix qui se justifient par les spécificités du genre et du discours auxquels on a affaire. Ainsi C. Poudat (2006a) propose-t-elle notamment de recourir à un étiqueteur entraîné qui permette de distinguer le *il* anaphorique et le *il* impersonnel, parce que ce dernier représente une caractéristique *a priori* importante dans le texte scientifique, et cette importance est vérifiée par l'examen des textes. Si une étude du genre de l'article doit permettre, comme le soulignent D. Malrieu et F. Rastier à propos des genres de texte en général, d'affiner ou de refonder les caractéristiques supposées du genre, les choix d'analyse ne vont pas sans référence à de telles caractéristiques supposées, qui relèvent comme nous avons eu l'occasion de l'évoquer de la compétence générique de tout locuteur.

Il n'en reste pas moins que seul le va-et-vient entre la définition préalable des genres et l'analyse des propriétés linguistiques des textes peut offrir une prise sur la caractérisation

²¹ Selon les propositions de F. Rastier (2001), qui définit les styles de manière relative aux genres, comme l'usage singulier que chaque auteur fait de normes génériques, que ce soit en y adhérant ou en s'en écartant. Cf. aussi le chapitre 6 de la présente étude, consacré à l'analyse de la variation stylistique de l'article de recherche (en sciences du langage uniquement).

différentielle des genres et des discours et sur la description linguistique d'un genre en particulier. Nous retiendrons l'importance de se baser sur un corpus de textes génériquement structuré pour décrire les genres. Par ailleurs, il nous faut revenir sur ce que nous avons dit de l'étude de C. Poudat (2006a), qui propose de décrire le genre de l'article de revue en linguistique dans une étude contrastive français-anglais. La comparaison inter-générique s'associe donc à une comparaison intra-générique, l'idée étant, au-delà de la différence schématique entre ces deux paradigmes, de mettre en évidence les caractéristiques du genre du double point de vue de ses régularités et de sa variabilité : les deux ordres de comparaison servent un même objectif de description linguistique des textes. La description des genres est à comprendre comme une description probabiliste, à la manière de ce que revendique B. Paltridge (1997) : la diversité des textes est essentielle, et ce sont les propriétés des textes qui définissent un genre, dont la stabilité ne saurait être restreinte aux seuls traits communs aux textes.

La dimension contrastive est particulièrement bien représentée dans les études existantes sur les genres scientifiques. Son importance est à comprendre par rapport aux motivations qui déterminent l'intérêt pour les genres scientifiques et plus largement pour les genres professionnels, en particulier dans le contexte anglo-saxon. Que ce soit dans le domaine de l'English For Specific Purposes ou dans les travaux sur la communication professionnelle écrite aux Etats-Unis (par exemple, Bazerman, 1988, Berkenkotter et Huckin, 1995, Bizell, 1992), l'objectif descriptif se veut un préalable à un objectif d'ordre pédagogique. Il s'agit dans cette perspective de décrire les caractéristiques rhétoriques et/ou communicatives des textes pour mieux comprendre comment fonctionne un champ spécifique, i.e. la communauté et l'identité professionnelles telles qu'elles se structurent dans l'usage des textes, et ce dans une perspective de formation des « nouveaux entrants », pour reprendre l'expression utilisée par P. Bourdieu (2001 : 95) à propos des doctorants dans le cas du champ scientifique.

La dimension contrastive est notamment développée dans le courant de l'English For Specific Purposes, et dans sa composante dédiée aux écrits dits académiques, l'English For Academic Purposes. La nécessité de former des étudiants à la communication spécialisée en anglais implique de tenir compte de différences culturelles, d'ordre linguistique, qui peuvent expliquer les difficultés de certains étudiants à produire un texte conforme aux exigences académiques anglo-saxonnes. Le genre est considéré comme un principe fondateur de la « communauté de discours » et est défini en termes d'objectifs communicatifs (Swales, 1990 : 158) ; en conformité avec ce que nous avons dit du genre jusqu'ici, J. Swales y voit un principe de régulation qui détermine la structure des textes, et contraint le choix des contenus et le style. Les communautés de discours sont conçues comme étant d'ordre socio-rhétorique et se caractérisent par les objectifs rassemblant leurs membres, objectifs tels qu'ils sont supérieurs à ceux de la socialisation et de la solidarité, par opposition selon J. Swales aux communautés de parole. L'enjeu consiste en particulier à s'intéresser à la construction et à la communication du savoir, et à son expertise.

Une telle approche rejoint ainsi l'idée d'une sociologie de la science qui ne soit pas seulement fondée sur la structure du champ, mais sur les procédures socio-rhétoriques d'adhésion et de persuasion : celles-ci sont aux fondements des consensus et des conflits relatifs aux paradigmes, et qui caractérisent l'évolution et les révolutions scientifiques selon T. Kuhn (1983). Par ailleurs, la dimension sociologique intervient aussi dans la nécessité de définir les relations sociales entre les membres de la communauté scientifique, les jeux de pouvoir, et la manière dont il est possible de se faire une place dans ce champ. Un des axes d'analyse rejoint alors la question des rôles endossés par le scripteur dans son texte, que nous avons évoquée *supra* à propos de la secondarité de certains genres, et nous verrons dans le

prochain chapitre l'importance prise par une telle question dans les analyses des genres et dans celles de l'écrit de recherche ou du discours académique.

La place de l'anglais à l'échelle mondiale, et dans le discours scientifique en particulier explique alors que nombre d'études entendent offrir des éléments de comparaison entre les standards anglo-saxons de l'« academic discourse », et les standards propres à des communautés plus locales (par exemple, en français, en allemand, en espagnol pour l'Europe, mais c'est vrai aussi en Asie, etc.). Les descriptions textuelles fournies dans ce type d'approches sont linguistiques ou rhétoriques, et sont mises en regard d'enjeux sociaux et culturels. On notera en particulier l'importance prise par les études sur les marques de personne, car il s'agit là d'un trait du discours académique particulièrement sensible aux variations culturelles. D'autres études entendent plutôt rendre compte de mouvements rhétoriques ou argumentatifs, par exemple, l'analyse proposée par J. Swales des introductions d'article (cf. au chapitre 7. de la présente étude pour davantage de précisions).

Plus largement, il s'agit en général de décrire certains phénomènes, ou certaines sections des textes (les introductions, les résumés²² si on ne les considère pas comme des genres en eux-mêmes) plutôt que des corrélations de traits textuels caractéristiques d'un genre comme dans l'étude de C. Poudat (2006a) dont nous avons expliqué les grandes lignes ; les phénomènes retenus le sont éventuellement au titre d'illustration de principes généraux du discours scientifique ; ainsi, les pronoms personnels doivent-ils permettre de dépasser l'idée d'un discours scientifique objectivant, et s'ils ne permettent pas en eux-mêmes d'en dire beaucoup sur l'ethos caractéristique du discours scientifique, une entrée comme l'évaluation peut contribuer à un tel objectif (cf. par exemple Hyland, 2000).

La perspective contrastive concerne comme nous l'avons dit les langues en jeu, mais plus largement l'idée de cultures nationales, voire internationales ; elle permet l'identification d'un modèle anglo-saxon dans l'écrit académique, modèle qui se caractérise en particulier par le recours à des principes dits « reader-friendly » : l'article guide le lecteur et lui donne le sentiment de participer à la construction du contenu (titrage, linéarisation du contenu, questions rhétoriques, etc.). Dans d'autres cultures à l'inverse, la digression ne serait pas prohibée, l'auteur serait plus effacé, etc. (cf. notamment M. Clyne, 1998, sur les différences entre les textes académiques anglais et allemands au niveau de la structure des textes). Le contraste concerne aussi les domaines et les disciplines, en conformité avec le parti pris descriptif à visée applicative qui caractérise le courant de l'English For Specific Purposes : l'idée est qu'on n'écrit pas la science de la même manière dans les différents champs.

Il nous faut alors mentionner le projet KIAP (Cultural Identity in Academic Discourse), dirigé par K. Fløttum à l'Université de Bergen²³, et mené de janvier 2002 à juin 2006. Ce projet se donne explicitement pour objet la comparaison culturelle dans le discours académique, et envisage l'idée de différences culturelles sous deux angles : celui des différences linguistiques et nationales, et celui des différences disciplinaires ; il s'agit de déterminer en quoi les identités culturelles observables dans l'article sont plutôt des identités nationales ou disciplinaires. Par rapport au courant de l'English For Specific Purposes, on relève donc ici une centration nette sur un seul genre de texte, l'article.

Celui-ci est décrit dans trois langues, l'anglais, le français et le norvégien, et dans trois disciplines, l'économie, la linguistique et la médecine. Neuf possibilités structurent donc le corpus de 450 textes²⁴ en fonction de la langue et de la discipline de l'article. Les trois langues (et les trois identités nationales) représentent trois degrés sur une échelle qui va d'une culture scientifique qui se caractérise par sa dimension internationale (l'anglais) à une culture

²² Cf. Melander, Swales et Fredrickson dans Duszak, 1997.

²³ Cf. <http://www.uib.no/kiap>

²⁴ Articles publiés entre 1992 et 2003 dans des revues à comités, et représentant un total de 3 millions de mots.

scientifique locale (le norvégien). De manière intermédiaire, l'article français est vraisemblablement lu par un lectorat non exclusivement natif, mais ne présente bien entendu pas des enjeux de diffusion comparables à ceux de l'anglais. Les trois disciplines correspondent quant à elles à trois pôles qui structurent les sciences, avec le pôle des sciences dures représenté par la médecine, le pôle des sciences humaines avec la linguistique et le pôle des sciences sociales avec l'économie.

L'originalité du projet tient d'abord à son ampleur, puisque c'est une description systématique de l'article dans les trois langues et les trois disciplines qui est visée. Elle repose par ailleurs sur le choix des phénomènes linguistiques pris en considération. Tout en se basant sur l'hypothèse d'une rhétorique du discours académique et en revisitant certains phénomènes déjà décrits dans le courant de l'English For Specific Purposes (le métadiscours, les pronoms personnels), l'entrée est très nettement linguistique (plutôt que rhétorique ou socio-rhétorique) : sont retenus des faits linguistiques liés à l'argumentation, et dont le choix semble pouvoir être compris en référence aux approches énonciatives en linguistique, et plus spécifiquement encore aux approches polyphoniques. C'est la question des marques de soi et de l'autre qui retient l'attention des chercheurs du projet KIAP, autour de trois questions prises comme point de départ :

- comment le scripteur de l'article se manifeste dans son texte,
- comment le travail d'autres chercheurs est pris en compte dans l'article,
- et comment l'attitude de l'auteur s'exprime à travers la présentation et la promotion de sa propre recherche, sur un axe épistémique (le degré de certitude dans les assertions, par exemple avec les *hedges*, ou marques de nuance) et sur un axe évaluatif (un exemple est donné avec « this approach is highly problematic ») (cf. notamment Fløttum, K., Gjesdal, A., Kinn, T., Vold, E.T., Dahl, T., 2006 et Fløttum, K., Dahl, T., Kinn, T., à paraître 2006).

Plus spécifiquement, les traits décrits sont les suivants : les pronoms de première personne et les pronoms indéfinis en fonction sujet (typiquement, le *on* français), les verbes combinés à ces pronoms, les marqueurs de la modalité épistémique, les connecteurs argumentatifs, les balises méta-textuelles, les constructions « reader-oriented » du type *let me/let us* (+ infinitif), les constructions polyphoniques (en particulier la négation polémique, les connecteurs d'opposition, la concession) et les références bibliographiques.

Au final²⁵, l'hypothèse générale d'une influence plus forte de la discipline que de la langue semble validée au niveau des faits étudiés. S'il reste difficile selon les auteurs de mettre en évidence des profils uniformes, ce qui est là nous semble-t-il tout à l'honneur du projet, ils montrent que des ethos différents se manifestent, selon que le chercheur s'efface plus ou moins de son texte, selon qu'il met l'accent sur une recherche en cours ou sur les savoirs produits, et selon qu'il argumente plus ou moins directement, et de manière plus ou moins polémique à l'égard des travaux du champ auquel il se réfère. La question de l'effacement et des modes de marquage de la personne dans son texte induit cependant des différences nationales, pointées aussi dans d'autres travaux comme nous l'avons dit, avec l'usage fort de la première personne *I* chez l'auteur anglo-saxon, l'usage d'un *nous* collectif chez le norvégien et l'usage du *on* chez l'auteur français.

Ainsi, il semble possible de dire que le projet KIAP fournit des pistes particulièrement intéressantes pour une analyse linguistique de la question des rôles discursifs ou de l'ethos caractéristiques d'un genre à proprement parler²⁶. La comparaison entre les langues et les cultures l'emporte cependant sur la mise en évidence de la diversité des textes interne à une langue et un champ considéré ; tel qu'il est traité dans l'étude de C. Poudat (2006a), cet aspect montre l'intérêt d'une étude fondée sur des corrélations de traits linguistiques, et sur des

²⁵ Cf. les textes cités (Fløttum, et al., 2006 et Fløttum, et al., à paraître 2006).

²⁶ Plutôt que d'extraits de texte d'une part ou d'analyses de textes menées sans référence aux genres d'autre part.

analyses factorielles en particulier, qui décrivent les textes en montrant comment leur diversité peut se structurer en fonction de critères linguistiques de différenciation.

Ce rapide aperçu des approches empiriques des genres montre comment elles corroborent et prolongent leur analyse programmatique, qui caractérise notamment les travaux de M. Bakhtine. A l'encontre d'une perspective taxinomiste, elles doivent leur intérêt à la description linguistique des genres qu'elles fournissent, mais indiquent aussi que les seules propriétés linguistiques ne suffisent pas à une analyse des genres. Le genre doit servir de point de départ dans la caractérisation des textes et dans la constitution du corpus, et l'analyse des textes peut alors permettre d'affiner voire de refonder les classements *a priori* des genres, leurs caractéristiques supposées (les savoirs sociaux circulant sur un genre, et qui révèlent une conscientisation partielle de leurs normes empiriques), et de rendre compte au final de la diversité des textes.

Dans cette optique, la caractérisation différentielle des textes se présente comme un outil heuristique et l'analyse des textes doit permettre de déterminer en retour la pertinence de facteurs potentiels de variation du genre (en particulier la langue, la discipline, le style ou usage singulier que chaque auteur fait du genre, etc.). La linguistique de corpus et en particulier les analyses statistiques peuvent ainsi permettre de répondre à l'objectif descriptif, ou du moins de fournir une prise sur les textes, à affiner par des observations plus qualitatives. Par ailleurs, la question des traits linguistiques à retenir est diversement résolue dans les différentes approches, selon qu'elles entendent multiplier les dimensions prises en compte sans hypothèse préalable sur les caractéristiques du genre²⁷, ou les restreindre à des aspects considérés *a priori* comme pertinents.

Notre étude du genre de l'article a été rattachée au projet KIAP au titre de projet satellite ; dans le champ des travaux existants sur l'article, et dont nous venons d'esquisser ici quelques traits, notre étude poursuit en effet des objectifs communs avec le projet KIAP, au niveau de la comparaison interdisciplinaire d'une part, et des traits linguistiques analysés d'autre part ; les aspects énonciatifs nous semblent en effet présenter un intérêt spécifique pour la description d'un genre, en particulier en regard de notions comme le rôle ou l'ethos, dont nous avons montré qu'elles sont jugées importantes dans les analyses programmatiques des genres. Nous allons développer dans le chapitre suivant cette piste, explorée dans le projet KIAP, d'une analyse énonciative des marques de soi et de l'autre, offrant une assise linguistique à la caractérisation d'un ethos, ou, dans notre perspective, d'une figure de l'auteur en tant que figure du genre. En d'autres termes, si nous convenons avec les travaux évoqués de la dimension rhétorique du genre de l'article, celle-ci doit se construire linguistiquement, et c'est cette hypothèse notamment que nous proposons d'étayer dans notre analyse.

Cela étant, par rapport aux travaux existants sur les genres, l'objectif ici est de caractériser un genre dans son ensemble plutôt que de restreindre *a priori* les phénomènes linguistiques retenus. Nous entendons prendre en compte avec les aspects énonciatifs d'autres traits linguistiques, en particulier au niveau de la structure des textes et au niveau morpho-syntaxique et lexical, ce qui rapproche alors notre étude de celle menée par C. Poudat (2006a) sur le genre de l'article en linguistique.

Nous visons par ailleurs à cerner la variabilité du genre de l'article en retenant, avec la discipline, d'autres facteurs de variation. Nous nous intéresserons ainsi aux styles d'auteur, dans le chapitre 6 qui s'appuie sur une étude que nous avons réalisée en collaboration avec C. Poudat pour les besoins respectifs de nos thèses. Par ailleurs, la question didactique de

²⁷ Bien que, comme nous l'avons dit, tout analyste dispose, comme tout locuteur, d'hypothèses sur les caractéristiques supposées d'un genre, et que l'automatisation de l'analyse ne doive pas masquer le fait que les traits linguistiques retenus correspondent à des catégories construites.

l'acculturation à un genre nous invite à cerner des profils diversifiés de compétence générique, en traitant de la spécificité des textes en fonction du statut du scripteur. Plus spécifiquement, comme nous l'expliquerons dans le chapitre 3, consacré à la problématique et à la méthodologie de notre recherche, nous proposons de voir en quoi les scripteurs doctorants font un usage spécifique du genre, mais il convient de souligner que si leurs textes sont publiés, c'est qu'ils ont été expertisés comme conformes au genre.

L'idée au final est donc de caractériser les textes en prenant comme point de départ le genre, à partir d'un corpus génériquement homogène, et de mobiliser dans ce cadre d'autres entrées pour rendre compte des régularités observables entre les textes, tout en mettant l'accent sur leur diversité. Les traits linguistiques retenus doivent dans la mesure du possible être mis en relation entre eux et analysés en référence à leurs fonctions dans les textes ; en cela, ils peuvent montrer en retour l'intérêt d'être étudiés en tant que caractéristiques d'un genre, pour pallier les limites d'une description de la « langue générale ».

Chapitre 2.

La figure de l'auteur comme figure du genre. Pour une analyse énonciative et polyphonique des textes.

Quels traits retenir pour caractériser un genre ? Nous avons vu à propos de la notion de genre que celui-ci varie avec tous les niveaux de la textualité et que tous sont donc *a priori* pertinents pour la caractérisation d'un genre en particulier. De fait, c'est l'étude des textes qui doit permettre de dire ce qui caractérise le genre, et dans cette optique, il s'agit de multiplier les aspects pris en compte, en s'intéressant à la fois à des traits structurels, sémiographiques, phraséologiques, énonciatifs et lexicaux. L'analyse multi-critériée des textes doit également permettre la mise au jour de profils diversifiés d'article sur la base de corrélations de traits.

Cela étant, nous voudrions explorer dans cette partie, à titre programmatique, des pistes pour une analyse énonciative et polyphonique des genres ; s'il ne s'agit pas de revendiquer une entrée unique pour l'analyse des genres, qui serait contraire à leur définition, nous proposons d'examiner en quoi les approches énonciatives peuvent servir leur description et plus spécifiquement celle de l'article de recherche. Qu'est-ce que les approches énonciatives peuvent permettre de dire d'un genre ? A quelles conditions, et comment peuvent-elles fournir une prise sur la textualité, au-delà de phénomènes locaux (par exemple, les déictiques de personne), en eux-mêmes très peu productifs pour la caractérisation d'un genre ?

Nous prenons comme point de départ le constat que la littérature existante sur les genres d'une part et sur l'écrit de recherche et les discours académiques d'autre part fait intervenir certaines notions récurrentes qui ont partie liée avec la dimension énonciative des textes ; elle suggère ainsi l'intérêt de cette dernière, dans une perspective de description, doublée d'une perspective de didactisation de l'article de recherche.

Nous nous tournons alors vers les approches énonciatives, pour voir quel cadre elles peuvent permettre d'élaborer pour l'analyse des textes et du genre de l'article de recherche alors que leur objet est l'énoncé et ses marques énonciatives. Nous esquissons d'abord les traits principaux de la problématique énonciative, autour des notions d'énoncé, d'énonciation et d'énonciation représentée, et nous intéressons alors à la polyphonie linguistique, avec pour objectif d'envisager son analyse au niveau du texte.

A titre de piste pour l'analyse des genres, ce second chapitre introduit la notion de figure de l'auteur, en tant qu'instance qui dans le texte s'y présente comme son responsable. Déterminée par le genre, cette figure se constitue linguistiquement dans la multiplicité des

points de vue, i.e. dans la dimension polyphonique des textes, et correspond en des termes interprétatifs à un principe d'économie dans leur prolifération. Elle vise à asseoir dans la matérialité du texte l'idée qu'il comporte une représentation du « rôle » investi par celui qui en est à l'origine, autrement dit une représentation du scripteur en tant que sujet social et praxéologique concerné par la construction du savoir.

2.1. En filigrane dans les analyses des genres et de l'écrit de recherche, l'intérêt de la dimension énonciative des textes

La notion de genre a pour intérêt une approche des textes dans leur dimension sociale et praxéologique. Si elle intéresse la description linguistique car elle doit permettre d'envisager la constitution d'un système de la langue dans ses usages textuels, elle représente également une entrée importante dans le cadre d'approches communicatives ; comme nous l'avons vu au chapitre 1, le genre favorise une problématisation de l'articulation entre les textes et le contexte de leur production, car c'est par l'entremise des genres que se structurent les communautés de discours.

Aussi l'ethnographie de la communication, telle que la définissent J.J. Gumperz et D.H. Hymes (1964, 1972) se donne-t-elle comme programme la caractérisation d'une compétence communicative sur la base des propriétés des productions verbales, par rapport au contexte dans lequel elles sont utilisées et évaluées. Cette perspective se manifeste nettement dans les travaux anglo-saxons sur la communication professionnelle et sur l'English For Specific Purposes que nous avons évoqués à propos des études sur les genres académiques, et qui ont pour spécificité leur perspective didactique.

En contexte européen, elle caractérise le « tour ethnolinguistique » pris par l'analyse du discours¹ : les genres textuels sont considérés comme un objet ethnolinguistique qui doit permettre de « lire [le linguistique] à travers l'*ethno-* d'un groupe qu'il présuppose et constitue » (Maingueneau, 1992 : 114). Il s'agit de rendre compte de la régulation des interactions sociales, et de la manière dont les pratiques langagières font une communauté *via* des rituels sociolangagiers qui légitiment les textes.

Dans le cadre des approches communicatives et ethnolinguistiques, certaines entrées sont privilégiées pour la caractérisation des genres, écrits en particulier, et plus spécifiquement encore des genres dits professionnels ou institutionnels, et notamment les genres scientifiques. Il ne s'agit pas tant d'étudier l'ensemble des propriétés linguistiques d'un genre que de se centrer sur celles qui peuvent dire quelque chose quant aux manières de dire et de faire propres à une communauté, et socialement régulées dans les genres qui y sont en usage. En dépit de l'irréductible diversité des approches, des notions communes sont convoquées : l'ethos, la légitimation, l'autorité, les places, les rôles, etc.

Très schématiquement, il semble possible de dire que, du côté anglo-saxon, ces notions sont introduites en référence à une conception rhétorique ou socio-rhétorique des textes² et que du côté européen, ce sont les approches énonciatives qui sont appelées à rendre compte de la dimension ethnolinguistique de la textualité³. Si traiter communément de ces

¹ Compte tenu de l'extrême plasticité de la désignation « analyse du discours », on pourrait parler ici plus spécifiquement encore de contexte français, s'agissant de signaler le passage entre une école française d'analyse du discours intéressée par les formations idéologiques (qui présupposent l'existence de formations sociales) à la prise en compte des genres et à l'analyse des conditions d'existence, dans les textes, de communautés discursives (Beacco, 1992, Maingueneau, 1992, Charaudeau et Maingueneau, 2002).

² Cf. les travaux cités à la fin du chapitre 1 (notamment Bazerman, 1988, Berkenkotter et Huckin, 1995, Bizell, 1992, Campbell, 1975, Swales, 1990).

³ Cf. J.C. Beacco (1992) et les différentes contributions de ce numéro 105 de la revue *Langages* consacré à l'*Ethnolinguistique de l'écrit*, ou encore S. Moirand (1988) et D. Maingueneau et F. Cossuta (1995).

diverses notions ne va pas sans gommer leurs spécificités, il semble possible de dire qu'elles ont toutes trait à la manière dont se constitue dans le texte une image de soi du scripteur, image qui sert de garant dans la construction des objets du discours, et légitime ainsi les contenus assertés (Amossy, 1999, Maingueneau, 2002, Swales, 1990).

L'objectif de description d'un genre peut se doubler comme nous l'avons vu *supra* d'un objectif didactique. C'est le cas des approches anglo-saxonnes sur la communication professionnelle et le discours académique. La perspective de didactisation d'un genre contraint la pertinence des descripteurs à retenir : la question qui se pose est de déterminer ce qu'il est essentiel que les scripteurs sachent faire pour produire un texte conforme aux attentes du genre.

Là encore, tous les niveaux de la textualité sont *a priori* concernés ; mais il importe de déterminer ce qui, dans les attentes, est déclaré le plus décisif, et de s'attacher à la description des pratiques du public visé pour cerner les éléments qui dans les textes et leur production peuvent être utiles à une réflexion sur l'accompagnement des scripteurs. L'objectif, d'ordre praxéologique est de fonder l'intervention didactique sur les difficultés même du public ciblé.

Concernant l'écrit de recherche et le discours académique, les travaux existants n'accordent pas tous aux genres une place centrale, mais se caractérisent par des constats et des propositions communes. Si nombre d'études sur les écrits de recherche et le discours académique entreprises dans une visée didactique se centrent sur leur structure ou leur lexique, la tendance actuelle semble indiquer un intérêt partagé pour des problématiques linguistiques d'ordre énonciatif, telles que les pronoms personnels (par exemple Loffler-Laurian, 1980, Fløttum et al., 2006, Tukia, 1996, Vassileva, 1998, etc.) et autres marques de soi du scripteur (Hyland, 2005), les citations (Boch et Grossmann, 2001a, Kara, 2004, Rabatel, 2002, Rosier, 2002, Swales, 1990) ou le discours rapporté, abordé en termes de « discours d'autrui » (Boch et Grossmann, eds., 2001, Guibert, 2001, Pollet et Piette, 2002, Reuter, 2001) pour intégrer nombre de phénomènes proches (notamment la modalisation en discours second du type *selon X*, etc.), ou enfin, de manière plus large encore, la question du dialogisme (Guibert, 2002) ou de la polyphonie énonciative (Delamotte-Legrand, 1996, 2002, Donahue, 2002, Fabre-Cols, 2001, Laborde-Milaa, 2002, Fløttum, 2005).

Ces aspects linguistiques sont pris pour indice de difficultés traitées en termes de posture (Rabatel, 2002), de position (Daunay, 2002), de positionnement ou d'identité énonciative (Guibert, 2001), d'investissement du scripteur dans son texte (Delcambre et Laborde-Milaa, 2002), de point de vue (Nonnon, 1998), d'ethos (Bizell, 1992, Maingueneau, 2002), de persona et d'autorité (Becker, 2004, Hyland, 2005), ou encore de construction du sujet textuel (Donahue, 2002) et d'image de soi du scripteur dans son texte (Delcambre et Reuter, 2002) et la liste est bien entendu non exhaustive.

Les analyses des difficultés des étudiants et des caractéristiques des genres auxquels ils sont confrontés convergent sur l'importance de ces aspects, dans les travaux européens visant l'élaboration d'une didactique de l'écrit dans l'enseignement supérieur¹ comme dans le domaine anglo-saxon avec l'English For Specific Purposes, ou même indépendamment de ce champ, par exemple dans la réflexion que propose le sociologue H.S. Becker (2004) sur l'écriture des sciences sociales, à partir de son expérience d'encadrement de mémoires.

Que l'objectif de description des pratiques langagières intègre ou non une visée applicative d'interventionnisme en didactique de l'écrit, le caractère récurrent des diverses notions que l'on a évoquées montre leur importance dans l'étude des genres, en particulier

¹ Cf. notamment le n° 24 de la revue *Lidil* (Boch et Grossmann, 2001), les n° 53 et 54 de la revue *Enjeux* (Pollet et Boch, 2002), le n° 29 de la revue *Spirale* (Delcambre et Jovenet, 2002), les n° 113-114 (Delcambre et Reuter, 2002) et 121-122 (Laborde-Milaa, Boch et Reuter, 2004) de la revue *Pratiques*.

dans le cas des genres institutionnalisés comme celui de l'article de recherche. Une question transversale parcourt ces différentes notions, celle du texte en tant qu'objet à la fois linguistique et ethnographique, dans sa relation avec la communauté discursive qui conditionne sa production et qu'il contribue à structurer.

A un premier niveau, les notions convoquées ont trait à la manifestation du scripteur dans son texte et plus largement à la gestion des deux dimensions que l'on peut désigner dans les termes de K. Fløttum (2005) comme « the self and the others », et qui renvoient à deux aspects traités par les approches énonciatives en linguistique, la subjectivité langagière (Kerbrat-Orecchioni, 1980), et la problématique de l'« autre en discours » (Bres et al., 1999), autrement dit à la polyphonie linguistique ou au dialogisme de tout énoncé.

Des compétences techniques sont ainsi visées : par exemple, comment gérer les références, introduire une citation, parler en son nom propre tout en s'effaçant pour produire un discours objectivant, etc. Mais les travaux existants montrent que les enjeux formels se doublent d'enjeux culturels, ne serait-ce qu'avec l'usage du *je*, du *nous* ou du *on* dans le texte scientifique. Ils mettent également en avant le rôle joué par ces aspects linguistiques dans la construction du savoir (par exemple, Boch et Grossmann, eds., 2001).

Par ailleurs, comme le soulignent K. Fløttum et al. (2006), et les travaux visant une didactique de l'écrit dans l'enseignement supérieur évoqués *supra*, la manifestation du scripteur ne se réduit pas à ces marques personnelles et celle des « autres » dépasse largement la question des formes de citation et de discours rapporté. Au-delà donc de ces phénomènes bien circonscrits et très fréquemment décrits dans les études sur les genres académiques, la question qui se pose linguistiquement est celle de la construction d'un point de vue problématisé, à la fois cohérent et complexe, personnel et documenté. L'enjeu se situe au niveau du texte pris dans son ensemble et les phénomènes pointés déterminent le problème du positionnement, de l'image de soi du scripteur dans son texte, ou de l'ethos spécifique des genres concernés.

A un second niveau, ces contraintes linguistiques, jugées centrales dans les genres considérés, sont le socle sur lequel se construit ce que nous conviendrons d'appeler avec R. Amossy (2000) leur « dimension argumentative » : on peut considérer à sa suite que si certains genres ont un but argumentatif avoué (i.e. une visée argumentative dans les termes de l'auteur), « tout texte tente d'influencer l'autre, de lui faire voir ou penser le monde d'une certaine façon, d'orienter le regard ou de susciter des interrogations » (Baron, 2006, en référence à Amossy et Maingueneau, eds., 2004).

Il s'agit donc de s'intéresser au dire et, indissociablement, à la légitimation du dire (Beacco, 1992), autrement dit à ce qui fait que le dire se présente comme convaincant et bien-fondé, et peut en cela amener le lecteur à modifier ses propres représentations, selon la définition de l'argumentation proposée par J.B. Grize (1990). Dans cette perspective, les analyses linguistiques sont appelées à éclairer la manière dont se constitue cette dimension argumentative, et avec elle le fonctionnement rhétorique des genres¹, et des genres académiques et scientifiques en particulier².

A un dernier niveau, la dimension rhétorique des textes pose la question de leur contexte socio-institutionnel : les notions d'ethos, d'identité, de postures, de rôles, rendent compte d'une tendance à envisager ce que disent les textes en termes de pratiques et de socialisation dans les mondes professionnels. C'est tout particulièrement vrai comme nous avons eu l'occasion de l'évoquer à la fin du chapitre 1 de l'analyse des genres académiques

¹ Cf. notamment R. Amossy, 1999, 2000, 2006.

² Cf. notamment D. Maingueneau, 2002, K. Fløttum, 2005.

dans une perspective communicative, et plus largement des travaux sur les discours spécialisés et professionnels¹. Si l'on s'en tient aux genres académiques, l'idée est d'aider les étudiants à se faire une place dans le champ académique, et l'analyse des textes s'associe alors à la prise en compte des images que les scripteurs se font d'eux-mêmes ; on le voit par exemple chez H.S. Becker (2004) qui évoque l'insécurité de ses étudiants dans la rédaction scientifique à partir de son expérience d'enseignement, ou encore avec les travaux de I. Delcambre et Y. Reuter sur les images du scripteur (2002), qui mènent une analyse de ces images dans les textes ainsi que des entretiens avec des étudiants pour s'intéresser aux images déclarées des scripteurs et à leur rapport à l'écrit et aux savoirs universitaires.

En résumé, ce rapide aperçu d'une littérature extrêmement foisonnante sur les genres d'une part et les écrits de recherche et le discours académique d'autre part montre qu'au-delà de la diversité des phénomènes traités et des notions convoquées, la question de l'image de soi du scripteur dans son texte, en tant qu'image déterminée par le genre, autrement dit l'éthos, se présente comme une question centrale. Elle l'est à double titre, en tant que caractéristique importante des textes à produire et en tant que lieu majeur de difficultés des étudiants, et conséquemment, cible privilégiée d'un interventionnisme didactique.

Dans la perspective qui est la nôtre, il s'agit d'une part d'envisager que l'image de soi du scripteur dans son texte est déterminée par le genre. D'autre part, si cette idée d'une image du scripteur constitutive du genre présente une assise linguistique, en termes notamment de manifestation de la subjectivité du scripteur dans son texte, il s'agit de mieux cerner encore comment se construit dans le texte une telle image, et les approches énonciatives nous semblent représenter une voie particulièrement intéressante. C'est pour ces deux raisons que nous parlerons de figure de l'auteur, en référence au titre éponyme de l'ouvrage de M. Couturier (1995) : la figure de l'auteur est l'image de soi du scripteur telle qu'elle est déterminée par un genre ; elle spécifie la notion d'éthos parce qu'elle interroge, comme elle, la manifestation du scripteur dans son texte (Amossy, 1999 : 11), mais à l'aune d'une approche polyphonique des textes comme nous le verrons.

Une entrée pour l'analyse des genres désignée en termes d'éthos mobilise des champs extrêmement divers, celui de la linguistique, celui de la rhétorique, celui de la pragmatique et celui de la sociologie des champs (Amossy, 1999). Cela étant, le problème transversal d'une relation entre les textes et les pratiques, qui affleure dans l'ensemble des propositions descriptives et didactiques que nous avons évoquées, qu'elles soient empiriquement étayées ou qu'elles demeurent programmatiques, appelle à notre sens une certaine prudence.

La question qui se pose est de savoir ce que les textes permettent de dire sur le contexte social où ils sont en usage. Les notions de posture, de position, d'identité peuvent tendre à une assimilation entre le niveau du texte et un extérieur au texte, d'ordre social et praxéologique, alors que cette assimilation est loin d'aller de soi ; si le texte comporte une dimension sociale et praxéologique, et sert la communauté du discours, rien cependant n'autorise à définir l'acteur social d'un champ à partir des attributs de l'image du scripteur dans son texte. Le danger de l'assimilation se manifeste aussi dans une analyse des textes en termes de stratégies, et dans une conception que D. Maingueneau qualifie de « cynique », et qui consiste à voir dans des caractéristiques textuelles un « invouable trafic destiné à l'autopromotion du scripteur » (1992 : 116).

Avec les notions d'énoncé et d'énonciation, la problématique énonciative a ceci d'intéressant qu'elle doit fournir des pistes pour interroger le lien entre le texte et l'acte où s'origine sa production ; elle peut permettre de mieux définir ce qu'on entend par « image », à

¹ Avec les travaux anglo-saxons cités à la fin du chapitre 1., cf. en contexte français J.M. Colletta, 1994, S. Moirand, 1990, J.C. Beacco et D. Lehmann, 1990.

l'encontre de l'idée que les textes sont le simple reflet d'une réalité extérieure, d'ordre sociologique ou praxéologique.

Par ailleurs, la nécessité avancée de mieux asseoir linguistiquement l'image du scripteur dans son texte vise aussi à envisager cette image dans un cadre qui intègre les différentes manifestations de soi et de l'autre dans les textes. C'est ce qui ressort comme nous l'avons vu d'un grand nombre de travaux en didactique, et qui est explicitement défendu à ce titre par K. Fløttum (2005). D'abord, il importe d'éviter de restreindre la manifestation du scripteur dans son texte aux indices les plus couramment convoqués (typiquement, les pronoms personnels). Il s'agit en outre de faire place aux manifestations de l'« autre » ; les problèmes d'images du scripteur reposent aussi comme nous l'avons souligné sur la gestion des « discours autres » et plus largement sur la dimension polyphonique des textes. L'objectif est d'éviter de restreindre cette dimension aux formes les plus attendues de la présence de voix autres dans l'écrit de recherche (typiquement, la citation, ou les références bibliographiques). Les approches énonciatives les plus récentes, qui portent sur la polyphonie linguistique, offrent nous semble-t-il une prise pour une analyse de la construction d'une image du scripteur, analyse qui ne se limite ni aux marques les plus évidentes de soi ni aux marques de l'autre traditionnellement décrites, et qui intègre les deux dimensions de soi et de l'autre.

Les approches énonciatives en linguistique vont à présent retenir notre attention :

1. d'abord, il s'agit d'interroger l'idée d'« image » ou de « figure » dans le texte d'un extérieur au texte,
2. d'autre part, il s'agit d'envisager la construction d'une telle figure à partir d'un cadre qui intègre les différentes manifestations de soi et de l'autre dans les textes.

Nous commencerons donc par esquisser les grands traits de la problématique énonciative, pour mieux cerner à partir des notions d'énoncé, d'énonciation et d'énonciation représentée la question de l'image que donne l'énoncé de son énonciation ou celle de l'image que le texte donne d'un extérieur au texte, et qui définit comme nous l'avons dit la notion de figure de l'auteur. Si les approches existantes qui se réclament d'une linguistique de l'énonciation sont extrêmement diverses, nous retiendrons qu'elles tendent actuellement à accorder une place centrale à la polyphonie linguistique, et c'est à cette problématique que nous nous intéresserons alors, avec pour objectif de voir en quoi une approche polyphonique des textes peut servir l'analyse d'une figure de l'auteur en tant que figure constitutive du genre.

2.2. Énoncé, énonciation et énonciation représentée dans l'énoncé

La définition linguistique de l'ethos, de l'image de soi du scripteur, ou encore, dans les termes que nous avons retenus, de la figure de l'auteur demande d'abord à s'interroger sur ce qui fonde l'idée d'image ou de figure, dans le texte, d'une instance qui en est à l'origine, ou qui le prend en charge énonciativement, autrement dit d'une responsabilité de l'énonciation de ce texte et des contenus qui y sont assertés. Les notions d'énoncé, d'énonciation effective et d'énonciation représentée doivent permettre de mieux comprendre cette idée d'« image » ou de « figure », en la distinguant du scripteur effectif à l'origine de la production du texte. Les grandes lignes des approches énonciatives esquissées ici ont pour objectif de mieux cerner leur objet, et la manière dont il se constitue avec ces trois notions. Elles fournissent ainsi une première ébauche du cadre à partir duquel analyser une figure de l'auteur dans l'article de recherche.

2.2.1. L'énonciation comme mise en fonctionnement de la langue

C'est sur la distinction entre énonciation et énoncé que les approches énonciatives trouvent leur fondement : leur objet, l'énonciation, est abordé sous l'angle de son produit, l'énoncé, ou en termes d'opérations qui y conduisent. Ce fondement n'est cependant que secondaire, et pourrait-on dire, méthodologique, par rapport à l'objectif premier, celui de s'intéresser aux usages de la langue.

Cet objectif ne concerne pas les seules approches énonciatives ; l'intérêt pour les usages caractérise plus largement une conception pragmatique du langage, telle que L. Wittgenstein s'en fait le défenseur dans ce qu'on appelle sa seconde philosophie. Cette philosophie du langage ordinaire développée dans les années 1930 entend dépasser les problèmes posés par la conception logique en philosophie analytique. Elle repose sur un double principe, l'idée que le langage sert à agir en relation aux autres, et qu'il faut étudier le langage dans ses usages ordinaires.

Ce parti pris décisif en philosophie du langage l'est aussi en linguistique où il sert de fondement aux approches énonciatives d'une part, et, de manière sans doute plus explicite encore, à la pragmatique, en tant que courant linguistique anglo-saxon initié par les travaux de J.L. Austin et la théorie des actes de langage qu'il développe dans *How to do things with words* (1962). Plus largement, on soulignera que cette conception pragmatique en linguistique participe de ce que F. Rastier (2001) désigne comme une tradition rhéto-herméneutique, qui s'oppose à la tradition logico-grammaticale en posant que le sens se définit dans les usages.

En dépit de leur grande diversité, les approches énonciatives en linguistique ont ceci de commun qu'elles se fondent sur cette position pragmatique en se donnant comme objectif d'étudier les usages effectifs de la langue. Cet intérêt pour les usages de la langue s'inscrit dans ce que l'on a pu concevoir comme un « tournant » pris par la linguistique dans les années 1970 (François, 1980), tournant qui tient à la remise en question de l'opposition entre langue et parole.

Si l'intérêt pour les usages de la langue peut être interprété comme une rupture avec la conception structuraliste, dominante jusqu'alors en linguistique, l'analyse rétrospective de la discipline mériterait que l'on s'attarde aussi sur les éléments de continuité, ne serait-ce que parce qu'on trouve trace de cet objectif d'analyse des usages de la langue chez des structuralistes comme C. Bally (1909) ou R. Jakobson (1973). Le premier entend d'ailleurs fonder une théorie de l'énonciation, mais c'est à E. Benveniste que reviendra le titre de père fondateur de ces approches énonciatives, dont l'émergence est précisément à comprendre en substituant à l'idée de rupture paradigmatique celle de continuité. Les travaux pionniers sur les déictiques ou embrayeurs en attestent : il s'agit pour des structuralistes comme R. Jakobson de rendre compte du système de la langue, mais de renoncer à l'immanence du signe en montrant que certaines unités ne sauraient être décrites abstraction faite de leur utilisation.

Aussi la position défendue par E. Benveniste (1966, 1974) témoigne-t-elle de cette continuité très nette entre une perspective structurale et une perspective énonciative attentive aux usages de la langue. Il ne s'agit pas tout à fait alors d'étudier le « discours » en lui-même : à l'instar de la position de F. de Saussure telle qu'on l'envisage nouvellement depuis la redécouverte de certains de ses écrits (cf. notamment Bouquet, 1999), l'objectif affiché dans les textes de E. Benveniste se situe dans la problématisation de l'articulation entre la langue comme code et son utilisation. En affirmant qu' « avant l'énonciation, la langue n'est que la possibilité de la langue » (Benveniste 1974 : 81), il montre que le système de la langue, virtuel, ne se construit que par les usages langagiers des sujets parlants, et ouvre la voie aux études sur l'actualisation de la langue en discours.

E. Benveniste (ibid.) définit ainsi l'énonciation comme « la mise en fonctionnement de la langue par un acte individuel d'utilisation ». Pour faire référence par la suite, cette définition n'en signale pas moins les limites des approches énonciatives sur la question de l'articulation entre le code et ses usages : si l'idée est claire chez Benveniste que le code, virtuel, se constitue par son actualisation dans les usages, l'acte d'énonciation, envisagé en tant qu'acte physique et mental, ne permet pas en lui-même de rendre compte de la constitution du code. Cela étant, l'intérêt reste de chercher à théoriser cet objet jugé non analysable par les structuralistes, i.e. la parole ou les usages effectifs, abordés en des termes énonciatifs en tant qu'actes individuels d'utilisation de la langue.

L'étude de cet acte individuel se heurte à un premier problème : « l'énonciation ne se reproduit jamais deux fois identique à elle-même » (Anscombe et Ducrot, 1976). Aussi les approches énonciatives doivent-elles fonder leur objet sur le postulat selon lequel dans l'énonciation, tout ne relève pas de l'individuel et du chaotique mais qu'une part notable peut être décrite en terme de système (Maingueneau, 1994). A l'encontre de la distinction entre le système de la langue et des usages individuels et uniques, l'idée est ici qu'en dépit du caractère singulier et événementiel de chaque acte d'énonciation, il existe des caractéristiques communes, ou un schéma général de l'énonciation, invariant à travers la singularité et la multiplicité des actes.

C'est dans cette perspective que vont être introduites, en tant qu'invariants de l'énonciation, deux dimensions mises à l'écart d'une linguistique structurale : la référence d'une part, la situation d'énonciation d'autre part. La référence est conçue comme ne pouvant être envisagée indépendamment de la relation entre interlocuteurs et du contexte effectif de l'énonciation. Ce contexte est quant à lui abordé en termes de situation d'énonciation, définie en tant que système de coordonnées abstraites associées à toute production verbale (Charaudeau et Maingueneau, 2002). La situation d'énonciation se trouve donc caractérisée au moyen de 3 paramètres : un temps, un lieu, des actants, diversement appelés locuteur et allocutaire, locuteur et interlocuteur, ou destinataire et destinataire.

Cela étant, avec le problème de l'unicité de l'acte d'énonciation s'en pose un second pour qui entend théoriser la mise en fonctionnement de la langue : cette mise en fonctionnement est en elle-même insaisissable. Comme le souligne C. Kerbrat-Orecchioni (1980 : 29), citant T. Todorov : « nous ne connaissons jamais que des énonciations énoncées ». Autrement dit, l'énonciation n'est observable qu'à travers les énoncés, que E. Benveniste (1966, 1974) prend soin de distinguer de l'acte lui-même en les définissant comme le produit de cet acte, ou en d'autres termes comme l'objet linguistique résultant de l'acte d'énonciation.

La distinction des deux objets et les définitions qui en sont proposées nous semblent d'une importance capitale. C'est le reproche que nous formulerons à l'égard du terme de « discours » : parce qu'il peut rendre compte à la fois du produit et de sa production, il nous semble présenter le risque de tendre parfois à une confusion entre les deux niveaux. Les approches qui distinguent texte ou matériau verbal d'une part et contexte, ou extérieurs au texte, ou encore extralangage d'autre part ne sont pas davantage satisfaisantes : l'idée d'intérieur et d'extérieur du texte, qui reproduit la délimitation du champ de la linguistique en termes d'intérieur et d'extérieur ignore la question du lien entre les deux dimensions opposées. En outre, il semble éminemment problématique de supposer que le « contexte », pour n'être pas du texte, est extralangagier. La notion de discours n'en dit pas davantage quant à l'ordre dont relève le contexte ou les conditions de production et d'interprétation d'un texte ; cependant, avec une définition du discours en tant qu' « inclusion du texte dans son

contexte c'est-à-dire dans ses conditions de production ou d'interprétation »¹, l'intérêt se manifeste d'aborder le rapport entre le texte et sa production ou son interprétation en termes de conditions.

La substitution d'une approche en termes d'intérieur/extérieur au texte ou en termes de texte/contexte par une approche fondée sur le couple production/produit nous paraît plus féconde. L'idée d'acte d'énonciation permet en effet d'envisager la production de l'énoncé en des termes proprement langagiers, même si cet acte est également qualifié par E. Benveniste d'acte physique et mental, et ce sont alors dans les relations entre l'énoncé et cet acte dont il résulte que se situe la problématique énonciative. Si l'acte n'est observable que dans son produit, en quoi ce passage de l'un à l'autre des deux niveaux est-il rendu possible ? Comment est-il expliqué ? Autrement dit, dans quelle mesure l'analyse de l'énoncé peut-elle permettre de dire quelque chose de l'acte d'énonciation ?

2.2..2. Les relations entre l'énoncé et l'énonciation : la réflexivité énonciative

Pour comprendre la problématisation de la relation entre l'énoncé et l'acte d'énonciation dont il résulte, il faut introduire une dimension qu'il est d'usage de nommer la propriété réflexive de l'énonciation (Récanati, 1979). Il s'agit de rendre compte du fait que dans l'énoncé se réfléchit le fait qu'il a été énoncé. C'est d'abord aux unités déictiques que la dimension réflexive de l'énoncé doit sa mise en évidence : ces unités du type *je*, *ici*, *maintenant* font référence à leur propre énonciation.

Une telle approche de l'énoncé met en péril une conception du langage fondée sur une idée de transparence entre les mots et ce qu'ils expriment : elle montre que l'énoncé n'est pas simplement quelque chose qui est dit sur autre chose (un extérieur, un monde, etc.), mais qu'il dit également quelque chose de sa propre production. Cette propriété réflexive par laquelle l'énonciation est « traversée par son auto-représentation » pour reprendre les termes de J. Authier-Revuz (1995) introduit de l'opacité dans la signification, et montre l'importance de considérer l'ordre langagier en tant que tel.

C'est sur la base de cette propriété réflexive de l'énonciation que s'établissent les approches indicielles en linguistique de l'énonciation. Dans un premier temps, il s'agit de rendre compte des unités déictiques en montrant la nécessité de tenir compte de l'acte d'énonciation pour comprendre leur sens. En langue, *je* ou *ici* renvoient virtuellement à un locuteur et au lieu de l'acte d'énonciation, et actualisés dans des énoncés, ils se trouvent instanciés par des locuteurs et des lieux effectifs, fictifs ou réels. Partant de ce constat, ce type d'unité va alors être envisagé comme trace, indice ou encore comme marque énonciatives : il s'agirait des traces, laissées dans l'énoncé, de l'acte d'énonciation dont il résulte.

Pour préciser ce que réfléchissent les traces ou marques énonciatives, on peut se référer D. Maingueneau (1994 : 13), qui explique en ces termes les fondements des approches indicielles en linguistique de l'énonciation : le langage est posé comme « une activité entre deux protagonistes, [appelés] énonciateur et allocutaire, activité à travers laquelle l'énonciateur se situe par rapport à cet allocutaire, à son énonciation elle-même, à son énoncé, au monde, aux énoncés antérieurs ou à venir. Cette activité laisse des traces dans l'énoncé, traces que le linguiste cherche à analyser ».

La démarche en linguistique de l'énonciation se caractérise donc comme une étude des indices ou des traces de l'énonciation dans l'énoncé. Une telle étude donne la priorité aux formes linguistiques : elle part des propriétés linguistiques de l'énoncé pour en rendre compte au moyen des éléments de l'énonciation. Ce faisant, un autre objectif, complémentaire, peut intervenir : s'il s'agit d'affirmer l'importance de tenir compte de l'acte d'énonciation pour

¹ Cf. le chapitre 1 pour une discussion terminologique sur « texte » et « discours » et cette définition de « discours ».

analyser pleinement ces marques, l'idée est aussi que ces marques doivent permettre de mieux cerner l'acte d'énonciation lui-même. C'est cet objectif qui prime dans l'approche énonciative développée dans les travaux d'A. Culioli (1990, 1999) : cette approche substitue à l'analyse indicielle une analyse en termes d'opérations : plutôt que de se centrer sur le produit qu'est l'énoncé, il s'agit de rendre compte des opérations énonciatives par lesquelles on aboutit à un énoncé. Ces deux objectifs

Essayons de cerner plus précisément encore cette question des relations entre l'acte d'énonciation et son produit, l'énoncé. On le voit dans la manière dont est abordée la situation d'énonciation, c'est bien sur l'énoncé lui-même que repose l'analyse de l'acte dont il résulte : la situation d'énonciation, nous l'avons dit, est définie à partir des éléments déictiques, ou éléments qui dans l'énoncé renvoient à ces extérieurs constitutifs que sont le temps, le lieu et les actants. Ces trois coordonnées de la situation d'énonciation sont retenues en tant qu'elles paraissent nécessaires pour rendre compte de ces unités déictiques observées dans les énoncés. Ces trois coordonnées, considérées comme inter-reliées mais également importantes dans un premier temps, font l'objet d'une nouvelle théorisation dans le cadre de l'approche praxématique, qui propose de voir dans le lieu le paramètre princeps duquel dépendent les deux autres (Détrie et al., 2001).

La sélection des caractéristiques pertinentes de l'acte d'énonciation, ici d'une des dimensions de cet acte, à savoir la situation d'énonciation, repose sur les unités observables dans les énoncés. On retrouve là le parti pris de J.P. Bronckart (1985) pour qui les conditions de production d'un texte ne peuvent être envisagées en tant que telles que dans la mesure où elles exercent une influence observable sur le texte produit. C'est ce qui permet de dire qu'elles conditionnent la production, et c'est aussi la seule manière qu'il y a pour le linguiste de traiter ce « contexte », si souvent convoqué comme « roue de secours » souligne G. Kleiber (1994), sans toutefois être véritablement questionné.

Si les observables linguistiques sont l'unique moyen de caractériser l'acte d'énonciation lui-même, il faut alors se prémunir d'une tentation consistant à prédéfinir l'acte avant d'en chercher les traces dans l'énoncé. Or, parler de traces de l'acte d'énonciation laissées dans l'énoncé est de ce point de vue relativement ambigu : en quoi le linguiste est-il autorisé à poser un acte d'énonciation pré-existant à l'énoncé, si ce n'est depuis l'énoncé lui-même ? Autrement dit, il convient de bien mesurer les implications de ce que rappelle C. Kerbrat-Orecchioni (1980 : 28), « ce n'est jamais qu'au travers des traces inscrites dans l'énoncé que l'on peut tenter de se représenter l'acte lui-même ».

C'est ce qui nous paraît problématique dans la définition reprise *supra* que donne D. Maingueneau (1994) des traces énonciatives : en disant que le langage est posé comme une activité qui laisse des traces dans l'énoncé, il prend comme point de départ l'activité langagière, en termes d'acte d'énonciation, alors que le point de départ du linguiste demeure l'énoncé.

Dans le cadre d'une analyse de type conversationnelle par exemple, il est intuitivement fondé d'envisager un lieu, un temps et des actants, et de montrer que dans les énoncés produits on trouve trace de cette situation. Cependant, si l'on considère une conversation mise en scène dans un roman par exemple, on notera que dans ce cas, la conversation en tant que situation mettant en jeu deux protagonistes en un temps et un lieu est reconstruite à partir de la conversation produite (qui n'est autre qu'un texte), laquelle implique qu'une pratique de conversation ait eu lieu. Ainsi le personnage d'un roman à la première personne n'existe-t-il que par ses propos. Si ce personnage est une instance physique et mentale, sujet de l'acte d'énonciation, ce n'est qu'en tant que représentation : l'énoncé à la première personne porte en lui la représentation d'une instance locutoire à son origine, qui n'existe cependant en tant que telle que parce qu'elle est impliquée par l'énoncé.

En définitive, les approches énonciatives n'échappent pas aux difficultés d'analyse des relations entre énoncé et énonciation que leur fondement même vise à éviter. Le problème tient à ce que parler de trace, c'est parler de trace de quelque chose, alors que ce dont on cherche la trace n'existe pas indépendamment de cette trace : l'univers fictionnel montre que le *je* ou le *ici* figurent l'acte d'énonciation plutôt que d'en être la conséquence.

Il s'agit donc de convenir du parti pris des approches énonciatives en posant premièrement que l'énonciation n'est pas l'énoncé et inversement, ce qui revient à ne pas confondre l'acte de production et son produit, et en affirmant leur interdépendance : s'il n'y a pas d'énoncé sans énonciation, il n'y a pas non plus d'énonciation sans énoncé. Ensuite, il importe de s'interroger sur cette idée que l'énoncé implique l'acte d'énonciation. Ce que l'on considère comme indice dans l'énoncé indique quelque chose de l'acte dont résulte l'énoncé, mais on ne saurait se représenter l'acte pour ensuite en chercher la trace : l'indice est un indice qui permet de se représenter cet acte. Il convient donc de distinguer ce que nous conviendrons d'appeler à l'instar de F. Rastier (2001) « l'énonciation effective » (l'acte d'énonciation) et « l'énonciation représentée » (telle qu'elle est représentée, ou réfléchie au niveau de l'énoncé).

La notion de représentation ici introduite en référence à F. Rastier nous paraît capitale : elle suggère en effet que le passage de l'indice à ce dont il est indicial ne doit pas être envisagé de manière mécanique, comme un simple reflet de ce qui est. Ce serait renouer avec une conception du langage comme reflet de la réalité, ici la réalité de l'acte d'énonciation. Si les marques de l'énonciation dans l'énoncé représentent cet acte d'énonciation, elles ne le reflètent pas de manière transparente. Il faut donc se garder en tant que linguistes de poser un acte d'énonciation physique et mental, pour s'en tenir à l'analyse de la manière dont est représenté cet acte dans l'énoncé ; rien ne permet d'affirmer que l'énonciation représentée est l'énonciation effective car la représentation que donne l'énoncé de son énonciation reste une représentation, non un reflet de l'acte effectif de production de l'énoncé. C'est en cela qu'il faut rester prudent lorsque l'on utilise l'analyse des textes pour dire quelque chose des pratiques où ils sont produits et interprétés.

Au terme de ce parcours sur la distinction entre énoncé et énonciation, et sur la manière d'appréhender leurs relations, on peut donc résumer les propositions faites de la manière suivante : l'analyse porte sur les énoncés. Elle postule un acte langagier à l'origine de l'énoncé et impliqué par l'énoncé lui-même, et entend rendre compte de la manière dont cet acte d'énonciation est représenté dans l'énoncé. Plutôt que de dire que le langage est posé comme une activité entre deux protagonistes qui laisse des traces dans l'énoncé, il faudrait donc reformuler la proposition de D. Maingueneau (1994) en disant que l'énoncé représente l'acte d'énonciation comme une activité entre des interlocuteurs ; il semble possible de poursuivre cette idée en disant que l'énoncé représente (ou implique) une interaction, c'est-à-dire à la fois une interlocution et une interdiscursivité.

Finalement, la question des déictiques qui sert de point de départ aux approches énonciatives se trouve donc compliquée par leur objet même : si *je* demande à être analysé en référence à la situation d'énonciation, il ne renvoie pas mécaniquement et simplement au locuteur en tant que sujet physique de la production verbale. L'idée est contre intuitive dans le cas de la conversation : je parle en utilisant le *je* qui renvoie bien à moi-même en tant qu'individu à l'origine d'un acte d'énonciation, mais le *je* d'un roman ne renvoie pas à son auteur, ce dont on convient en arguant que l'œuvre crée un monde, où *je* renvoie à une instance locutoire fictive (un narrateur, un personnage). Cette instance locutoire doit son statut de sujet en vertu d'une reconstruction par laquelle les propos énoncés se trouvent envisagés comme le fait d'une instance physique et mentale à l'origine de ces propos : le *je* implique

dans l'énoncé une instance qui énonce ce *je*, mais qui ne peut être assimilée au producteur effectif de l'énoncé.

Il reste alors à traiter des marques étudiées par les approches énonciatives, et à voir comment on est passé d'une approche attentive aux déictiques, en tant que marques de la situation d'énonciation *hic* et *nunc*, à l'analyse de la manifestation d'une subjectivité dans les textes qui dépasse largement cette trace du *je*, pour en venir enfin à une approche polyphonique des textes, attentive à l'hétérogénéité énonciative observable dans l'énoncé. Alors que les unités déictiques peuvent, en première approche du moins, conforter l'idée que l'acte d'énonciation est un acte physique, situé dans le temps et l'espace et produit par un sujet, la polyphonie contredit quant à elle l'idée que l'énonciation représentée est assimilable à l'énonciation effective, *a fortiori* telle qu'elle est conçue avec la situation d'énonciation *hic* et *nunc* et les trois paramètres des acteurs de l'énonciation, du temps et du lieu.

2.2.3. Les marques étudiées par les approches énonciatives

Historiquement, les approches indicielles en linguistique de l'énonciation trouvent leur point de départ comme nous l'avons dit avec l'analyse des unités dites déictiques. Leur étude fait apparaître que ces unités ne sauraient être comprises comme des catégories grammaticales : c'est plutôt en termes d'opérations que l'on peut envisager cette question de l'embrayage, par lequel un énoncé se présente comme s'ancrant dans une situation d'énonciation ou s'en autonomisant. C'est donc la représentation dans l'énoncé de la situation d'énonciation qui est étudiée, en tant qu'elle signale un rapport spécifique de l'énoncé à cette situation.

Ce rapport de l'énoncé à la situation d'énonciation est à la base des typologies énonciatives de textes proposées par E. Benveniste (1966), J. Simonin-Grumbach (1975) et J.P. Bronckart (1985). L'opposition entre le « discours » (énoncé ancré) et le « récit » (énoncé coupé de la situation d'énonciation) de E. Benveniste cède la place chez J.P. Bronckart à l'idée d'un ancrage ou d'une autonomisation tendanciels, que manifeste le texte par rapport à la situation d'énonciation. La typologie de J.P. Bronckart (1985), étayée par une description interne des textes, postule une influence des conditions d'énonciation sur l'organisation discursive, déjà pointée par E. Benveniste. Nous conviendrons avec J.P. Bronckart que cette influence des conditions d'énonciation ne peut être envisagée en tant que telle que dans la mesure où l'influence est observable sur les textes produits. Cependant, dans notre approche, nous parlons de genre plutôt que de conditions d'énonciation : le genre conditionne la production et l'interprétation du texte, et c'est lui qui détermine les caractéristiques du texte et le rapport spécifique que celui-ci manifeste quant à la situation d'énonciation.

Il reste que, et nous aurons l'occasion de l'évoquer à propos de l'article de recherche au chapitre 7, la grille proposée par J.P. Bronckart nous semble heuristique pour décrire énonciativement les genres, à la fois pour la manière dont s'y trouve formulée la question du rapport à la situation d'énonciation et parce qu'elle associe à ce premier paramètre la question de la mimesis spécifique d'un texte. C'est avec ce second paramètre que cette typologie des « architypes de discours » permet de comprendre pourquoi dans certains textes on convient que le *je* n'est pas celui du scripteur effectif tandis que dans d'autres on postule une identité entre les deux.

La question des déictiques conduit rapidement les approches énonciatives à celle du sujet et de son inscription dans l'énoncé : les déictiques ne sont pas les seuls indices d'une instance locutoire à l'origine de l'énoncé, et d'autres marques disent quelque chose de cette instance : les marques lexicales de la subjectivité. On rejoint alors l'idée énoncée par C. Bally (1909) que tout énoncé contient un « dictum », le dit, et un « modus », la manière de le dire,

l'attitude du locuteur par rapport à ce qu'il dit. A la dimension référentielle de tout énoncé, en vertu de laquelle l'énoncé parle de quelque chose, s'ajoute sa dimension modale, qui tient à ce que l'énoncé dit quelque chose de celui qui en est le responsable. A travers les questions de la modalisation et de la subjectivité langagière, c'est donc la représentation dans l'énoncé du sujet responsable de l'énonciation qui est étudiée. En ce qui concerne plus spécifiquement la modalisation, cette représentation est étudiée en tant qu'elle signale un rapport spécifique du sujet à ce qu'il dit.

On voit ainsi comment les approches énonciatives peuvent fournir des outils pour l'analyse d'une image du scripteur, d'un ethos ou d'une figure de l'auteur dans les textes ; c'est ce que souligne R. Amossy (1999) à propos de l'ethos, en référence aux travaux de C. Kerbrat-Orecchioni (1980), dont l'objet est d'étudier la subjectivité langagière en fournissant un large aperçu des marques lexicales de subjectivité et en proposant d'en identifier différents types (les termes affectifs, les évaluatifs axiologiques ou non, les épistémiques, etc.). Par ailleurs, on retrouve l'idée que cet ethos, pour relever d'une dimension rhétorique des textes, ne représente pas cependant une composante exclusivement ornementale ; la modalisation doit offrir une prise sur le degré de certitude d'une assertion et sur la question de l'évaluation, et pourrait intéresser dans le cas des écrits de recherche la question de la construction linguistique du savoir.

Cela étant, les approches énonciatives vont également mettre en avant le phénomène de la polyphonie. C'est cet objet qui domine aujourd'hui le champ des approches énonciatives. La théorie polyphonique est développée par O. Ducrot (1984) sur la base des écrits de M. Bakhtine (1970) relatifs à la polyphonie dans le roman. La polyphonie scandinave s'inspire de ces travaux, tout en proposant un cadre original, largement confronté d'ailleurs aux propositions de O. Ducrot lui-même. Il est aussi question de dialogisme, notamment dans le cadre de l'approche praxématique (Bres, 1999, Bres et Vérine, 2002, Détrie et al., 2001).

Avant d'entrer dans le détail de ces analyses qui intéressent de près notre étude, retenons ici que le point de départ est le constat de la multiplicité des « voix » ou instances énonciatives à l'origine des points de vue dans l'énoncé. La mise en évidence de cette hétérogénéité énonciative (J. Authier-Revuz, 1984) au niveau de l'énoncé va de pair avec une conception interactive ou intersubjective de la production verbale. Si le phénomène est décrit linguistiquement, il peut être compris aussi comme une représentation que donne l'énoncé de l'hétérogénéité constitutive de la production verbale. Autrement dit, l'énoncé comporte une représentation de sa production telle qu'il ne se présente pas comme le fait d'un seul sujet, mais comme le produit de deux dimensions de l'énonciation que sont l'interlocution et l'interdiscours.

Nous retiendrons donc qu'au-delà des seuls déictiques de personne, et du *je* que E. Benveniste (1966, 1974) désigne comme une marque de cette « véritable personne de l'énonciation » qu'est le locuteur, la question de la subjectivité langagière doit pouvoir offrir une prise pour l'analyse d'une image du scripteur dans son texte, d'un ethos ou d'une figure de l'auteur, en pointant un certain nombre d'indices qui permettent de mieux cerner à quelle subjectivité on a affaire.

Le constat selon lequel la subjectivité langagière est omniprésente (Kerbrat-Orecchioni, 1980) corrobore l'idée d'une « persona » y compris dans le discours scientifique (P.N. Campbell, 1975) ; cependant, ce dernier se caractérise comme le souligne J.P. Bronckart par une autonomisation à l'égard de sa situation d'énonciation, et en d'autres termes, par un « effacement énonciatif » (Philippe, 2002, Rabatel, 2004a). La subjectivité manifeste dans le texte scientifique repose donc pour partie sur l'effacement de la subjectivité de la personne

empirique à l'origine du texte. Les marques de subjectivité analysées comme telles en linguistique de l'énonciation doivent permettre de décrire à la fois cet effacement énonciatif et cette subjectivité spécifique du texte scientifique.

La question de la polyphonie linguistique doit permettre quant à elle de mieux cerner la manifestation de l'« autre en discours » (Bres et al., 1999) mais complique aussi l'analyse d'une subjectivité dans le texte, en pointant l'hétérogénéité des voix et des points de vue observable au seul niveau de l'énoncé.

Pour déterminer en quoi une approche polyphonique des textes peut servir l'analyse d'une figure de l'auteur en tant que figure constitutive du genre, il nous faut tenter de définir les fondements d'une telle approche. Les principaux enjeux de l'élaboration d'un cadre pour une analyse polyphonique de la figure de l'auteur sont d'abord de concilier la polyphonie linguistique avec l'idée, au niveau du texte, d'une image d'un « soi » suffisamment homogène pour être conçu comme tel. D'autre part, si les approches énonciatives reposent sur l'énoncé et l'actualisation de la langue dans l'énoncé, elles envisagent généralement ce dernier dans des unités phrastiques ; comme dans notre optique, nous nous intéressons au texte, il convient de résoudre le problème du niveau d'analyse, c'est-à-dire le problème du passage de l'analyse polyphonique au niveau de l'énoncé, restreint au cadre phrastique, à celui du texte.

2.3. De la polyphonie à la définition polyphonique d'une figure de l'auteur

C'est sur la base de la distinction entre l'énonciation et l'énoncé, ou entre *Le dire et le dit* que O. Ducrot (1984) développe une théorie de la polyphonie linguistique. Sur cette première distinction vient s'en greffer une autre, celle entre le sujet parlant et le locuteur. Le premier est défini comme le sujet empirique à l'origine de l'acte d'énonciation : producteur effectif de l'énoncé, il s'agit d'une entité psycho-sociologique. Le second, le locuteur, est l'« être présenté dans le sens même de l'énoncé, comme étant le responsable de l'énonciation ». Nous conviendrons de l'appeler LOC dans ce chapitre, pour le distinguer d'une acception commune du locuteur comme producteur effectif de l'acte d'énonciation (i.e. le sujet parlant dans la théorie ducrotienne), et nous verrons qu'il s'agit d'une instance communément mobilisée dans les principales approches qui intéressent notre propos.

Cette instance à l'origine de l'énoncé en tant qu'elle est représentée comme telle dans l'énoncé suppose comme le précise O. Ducrot (2001) que « le sens de l'énoncé comporte entre autres choses une description de son énonciation ». Autrement dit, il s'agit là encore de ne pas confondre l'énonciation et l'énoncé, et, partant, l'énonciation effective et l'énonciation représentée dans l'énoncé. Cette distinction, on l'a suggéré *supra*, est contre-intuitive : elle met en péril l'illusion subjective de la parole, selon laquelle le responsable de l'acte d'énonciation, producteur effectif de l'énoncé, serait bien le responsable de ce qui est dit dans l'énoncé. Les termes de cette distinction sont extrêmement variés, de même que les arguments en sa faveur, mais nous retiendrons pour notre part cette nécessité de ne pas assimiler l'observable, i.e. le texte dans notre perspective, et l'acte effectif de sa production.

Si l'ensemble des approches énonciatives actuelles intègrent d'une manière ou d'une autre cette distinction, différents problèmes surgissent dès lors que l'on cherche à les mettre en perspective : des problèmes d'ordre terminologique, qui s'expliquent en partie par des problèmes d'objets ou de niveaux d'analyse (la phrase, l'énoncé, le texte voire le discours) et enfin des problèmes de démultiplication des instances et de la question de leurs éventuels liens. Le problème se complique avec la question de la polyphonie, traitée parfois aussi en termes de dialogisme, ou encore avec celle de l'hétérogénéité énonciative, sans compter les travaux sur le discours rapporté, le discours représenté, les discours autres ou la circulation

des discours, qui peuvent ou non situer ces objets-là dans un cadre d'analyse polyphonique ou dialogique.

Nous n'entendons pas initier une théorie de plus, ni même comparer terme à terme l'ensemble de ces approches et les mettre en perspective dans un cadre unificateur. Notre objectif est de bien cerner les fondements et les enjeux d'une analyse polyphonique, afin de définir des pistes pour une analyse polyphonique des textes, qui permette de caractériser énonciativement la figure de l'auteur en tant que figure du genre.

Pour ce faire, nous prenons comme point de départ le principe de base de la polyphonie, à l'aide d'exemples souvent mobilisés dans la littérature. Nous essayons de voir comment le mettre en œuvre pour une analyse des textes, à partir de deux notions centrales, le point de vue et la question de sa prise en charge énonciative. Nous nous intéressons de plus près d'une part à la théorie scandinave de la polyphonie linguistique (intitulée ScaPoLine par ses auteurs), et avec elle à la théorie développée par O. Ducrot dont elle s'inspire, d'autre part aux travaux de A. Rabatel. Nous tentons de voir comment chaque approche résout les questions du point de vue, de sa prise en charge, et de l'analyse de la polyphonie au niveau du texte.

Nous définissons ainsi les fondements énonciatifs et polyphoniques de la notion de figure de l'auteur, que nous introduisons à la fin du chapitre. Les éléments de réflexion sur les approches polyphoniques dépassent nos objectifs d'analyse, mais montrent cependant à titre programmatique l'intérêt de poursuivre un travail de confrontation de ces approches pour tenter de mieux cerner encore la polyphonie linguistique et son intérêt pour l'analyse des textes et des genres.

2.3.1. Pour introduire la question de la polyphonie

La polyphonie peut être définie en première instance comme la multiplicité des voix et des points de vue observables dans un énoncé. Un exemple fréquemment convoqué pour en rendre compte, notamment chez O. Ducrot et dans la ScaPoLine est celui de la négation, dans :

(1) Ce mur n'est pas blanc

Cet énoncé est décrit sous l'angle du dédoublement des points de vue impliqué par la négation : l'énoncé comprend un point de vue dit pdv1, dont le contenu sémantique pourrait être glosé de la manière suivante : « ce mur est blanc », et un second point de vue, selon lequel « pdv 1 n'est pas valide ».

La question qui se pose alors est celle de la prise en charge énonciative de ces points de vue. Les deux points de vue impliqués dans l'énoncé (1) se différencient à ce niveau-là : le pdv1 (« ce mur est blanc ») n'est associé à aucune instance de prise en charge énonciative, tandis que le pdv 2 (« pdv1 n'est pas valide ») se présente comme pouvant être attribué à l'instance qui dans l'énoncé se présente comme le prenant en charge, le locuteur au sens ducrotien (i.e. LOC).

Les analyses polyphoniques, celle de O. Ducrot d'une part, la théorie scandinave de la polyphonie d'autre part, ou encore celle proposée par A. Rabatel, et les propositions de J. Bres et B. Vérine (2002) en termes de dialogisme ont toutes en commun de poser cette question, et de s'attacher à déterminer quels points de vue peuvent être attribués au LOC. L'analyse proposée *supra* de cet exemple bien connu est consensuelle : c'est le pdv2 qui revient au LOC, tandis que le pdv1 se caractérise par une absence de marquage d'une quelconque instance énonciative.

Si les approches qui prennent comme point de départ la notion de point de vue introduisent la question de ces instances, certaines approches se centrent prioritairement sur ces instances en prenant comme point de départ la notion de dédoublement énonciatif ou d'hétérogénéité énonciative.

Pour les illustrer, on peut prendre pour exemple un autre phénomène considéré comme un cas typique de polyphonie ou de dialogisme, le conditionnel. L'exemple est fourni par A. Culioli (1990 : 150) mais on en trouve de très proches dans les analyses du dialogisme développées notamment par J. Bres (1999), ou J. Bres et B. Vérine (2002) :

- (2) Paul viendra demain
(2') Paul viendrait demain

Dans (2), un seul point de vue ou une seule voix interviennent ; si l'on parle de point de vue, on conviendra qu'il peut être attribué au LOC ; si l'on parle de voix, on dira qu'il s'agit de celle de l'énonciateur de cet énoncé – qui correspond en ce cas au LOC. L'énoncé (2') représente un cas de polyphonie ou de dialogisme. En première approche, on pourrait dire que le conditionnel suggère une distance énonciative : le LOC ne prend pas complètement en charge l'assertion. C'est ce qui explique qu'on parle de conditionnel épistémique¹ pour rendre compte de ces formes de conditionnel : elles touchent à la question du degré d'assertion, c'est-à-dire indissociablement, la question du degré de prise en charge de l'assertion par le locuteur, et celle du degré de certitude de l'assertion².

Dire que le LOC émet un doute ou une réserve serait problématique : ce serait outrepasser les limites que doit se donner l'analyse linguistique, car cela reviendrait à conférer à l'instance énonciative qu'est le LOC des attributs propres au locuteur non plus au sens de représentation dans l'énoncé de son responsable mais au sens de responsable de l'acte d'énonciation. Autrement dit, si l'on s'en tient au matériau linguistique lui-même, la description ne saurait se satisfaire d'une analyse en termes de doutes ou de réserves, et si on peut parler de distance, encore faut-il la caractériser énonciativement.

C'est en ce sens qu'est introduite la notion de dédoublement énonciatif, pour rendre compte linguistiquement de la différence entre (2) et (2'). Si l'énoncé (2) présente un seul point de vue et une seule instance énonciative, l'énoncé (2') se caractérise par un dédoublement, par lequel le LOC se décharge de la responsabilité de l'assertion « Paul viendra demain » : par le conditionnel, cette assertion se présente comme imputable à une autre instance énonciative que le LOC. Impliquée par le conditionnel, cette instance pourrait se trouver marquée par d'autres précisions, du type « selon Marie, Paul viendrait demain », « Paul viendrait demain m'a-t-on dit », etc.

Le dédoublement énonciatif observable en (2') consiste donc à postuler deux énonciateurs, au sens d'instances de prise en charge énonciative : l'énonciateur responsable de l'assertion « Paul viendra demain », et l'énonciateur responsable de l'énoncé (2'), le LOC, qui par le conditionnel se présente comme ne prenant pas en charge l'assertion imputée au premier. Ce dédoublement énonciatif est conçu comme étant à la base du phénomène

¹ On le trouve également souvent désigné sous le terme de conditionnel journalistique, car il se trouve fréquemment employé dans ces textes journalistiques. On peut voir là un argument à l'analyse des formes de polyphonie comme caractéristiques propres des genres et des discours.

² Repérons dans cet exemple le lien qui existe entre les problématiques de la modalisation et de la polyphonie, la structure polyphonique déterminant une modalisation spécifique du pdv1 et du pdv2.

polyphonique selon J. Bres et B. Vérine notamment (2002), qui préfèrent quant à eux parler de dialogisme¹.

Pour mieux mettre en perspective les différentes approches de la polyphonie, la question qui se pose est de déterminer en quoi il est possible d'analyser le dédoublement énonciatif observé en (2') en termes de points de vue seulement, plutôt qu'en termes de dédoublement de voix.

A la manière de la description faite de l'énoncé (1), on pourrait donc suggérer qu'en (2) on a affaire à un seul point de vue, alors que deux points de vue se manifestent en (2') : un pdv1 qui correspond à l'assertion « Paul viendra demain », et un pdv2 qui se présente comme imputable au LOC.

Que dit ce second point de vue ? Sa glose ne paraît pas aussi aisée que dans le cas de (1) où le second point de vue consiste en une négation de la validité du premier. Ici, le dédoublement introduit par le conditionnel est tel que le pdv1 « Paul viendra demain » fait l'objet d'un commentaire par le pdv2, mais il semble que ce commentaire apporté par le pdv2 se limite à dire « pdv1 est une assertion possible ». On ne saurait en effet préjuger – et *a fortiori* à partir d'un énoncé certes attestable mais complètement décontextualisé ici comme l'est (2') – que pdv2 dise du pdv1 qu'il n'est pas justifié, comme c'est le cas dans l'énoncé (1). Pdv2 dit plutôt que « pdv1 est peut-être justifié mais peut-être pas », autrement dit, le pdv2 introduit en quelque sorte une interprétation en suspens de pdv1, l'exhibant comme une assertion virtuelle et envisageable tout en disant de ce pdv qu'on ne peut rien en dire d'autre que de retarder sa prise en charge ou sa négation.

D'autres descriptions polyphoniques de cet énoncé sont vraisemblablement envisageables. Cependant on voit qu'une analyse en termes de points de vue reste relativement limitée, car le conditionnel épistémique suggère non seulement un dédoublement des points de vue, mais peut-être plus fondamentalement un dédoublement des voix : dans (2'), le pdv1 « Paul viendra demain » est présenté comme un point de vue, mais comme un point de vue virtuellement pris en charge : ce pdv1 est présenté comme pouvant potentiellement être ou avoir été attesté ou attestable, et c'est ce qui caractérise la distance énonciative impliquée par le conditionnel à l'égard de ce point de vue.

Aussi l'analyse en termes de dédoublement énonciatif, i.e. de dédoublement des instances de prise en charge énonciative dans un énoncé nous paraît-elle ici intéressante. Il convient alors de souligner que les théories énonciatives ont en commun d'introduire cette question des instances de prise en charge des points de vue, autrement appelées les sources des points de vue. Autrement dit, si en première approche on peut distinguer des cadres d'analyse intéressés par les points de vue et ceux intéressés par les « voix », cette question de la source des points de vue est centrale dans les premières comme dans les secondes.

A quoi peut tenir la pertinence même de cette notion, ou quels en sont les enjeux pour l'analyse énonciative des textes ?

D'une part, il s'agit, comme on a pu le voir à travers les énoncés (1), (2) et (2') de déterminer quels points de vue peuvent être rattachés au LOC. En d'autres termes, la question est de savoir dans quelle mesure un point de vue est présenté comme imputable à l'instance qui dans l'énoncé se présente comme responsable de son énonciation. Ainsi, dans (2), le point de vue (dit « simple » dans la ScaPoLine), « Paul viendra demain » est imputable à l'instance qui dans l'énoncé s'y présente comme étant responsable de l'acte d'énonciation dont il

¹ Ils entendent par là mettre l'accent sur la hiérarchisation des énonciateurs. Nous reviendrons sur cette question, qui est également jugée centrale dans la théorie scandinave de la polyphonie linguistique en particulier ; l'importance de cette hiérarchisation distingue selon K. Fløttum (2005) la ScaPoLine de l'approche de la polyphonie dans le roman développée par M. Bakhtine (1970).

résulte, le LOC. Dans (1) et (2'), ce sont les pdv2 (dits « hiérarchiques » dans la ScaPoLine car ils impliquent un autre point de vue à l'égard duquel ils disent quelque chose) qui se présentent comme pouvant être imputés à cette instance.

D'autre part, cette première question s'intègre à l'objectif plus large de décrire la multiplicité des points de vue en termes de multiplicité des voix au niveau du texte. Nous avons souligné à partir de l'exemple (2') qu'elle permet de rendre compte du fait qu'un point de vue peut se présenter dans l'énoncé comme pris en charge, au moins virtuellement, même si ce n'est pas toujours le cas. Ainsi dans (1), rien dans l'énoncé n'est dit quant à la source du pdv1 (« ce mur est blanc ») (Nølke et Olsen, 2000). En (2') cependant, le conditionnel pourrait impliquer non plus seulement un autre point de vue, mais bien une autre voix : pdv2 consiste à montrer que pdv1 a été virtuellement assumé ou pourrait l'être. Imaginons pour poursuivre avec des énoncés phrastiques et fabriqués une extension à (2') de type « selon Marie », « on m'a dit que », etc. En ce cas, l'énoncé montrerait alors le pdv1 « Paul viendra demain » comme un point de vue actualisé, c'est-à-dire attesté ou en d'autres termes effectivement pris en charge par une instance désignée « Marie », « on », etc.

Retenons au final que non seulement donc cette dimension de la source d'un point de vue est utile à l'analyse de certains tours polyphoniques, mais elle semble pouvoir permettre de différencier ces tours en fonction de la manière dont est représentée ou non cette source dans l'énoncé. Enfin, cette question de la source du point de vue doit servir de point de départ dans la manière dont se construit le jeu des points de vue au niveau textuel.

2.3.2. Problèmes et enjeux de l'analyse polyphonique

La définition de l'instance à la source des points de vue

Les problèmes et les enjeux de l'analyse polyphonique trouvent leurs fondements dans ce double constat de la multiplicité des points de vue et des voix observables dans la plupart des énoncés.

La question de la multiplicité des voix semble davantage investie que celle des points de vue eux-mêmes. Elle est diversement traitée, et d'abord diversement désignée par les différentes approches. Sans faire l'unanimité, la désignation d'« énonciateur » semble cependant assez généralement utilisée pour rendre compte de la source du point de vue, dans la continuité du cadre d'analyse proposé par O. Ducrot (1984). C'est le cas en particulier dans deux approches qui nous semblent particulièrement intéressantes pour notre étude, l'approche de la multiplicité et de la hiérarchisation énonciatives développée par A. Rabatel¹ et celle du dialogisme, développée par J. Bres et B. Vérine².

Cela étant, la notion d'énonciateur comporte nombre d'ambiguïtés, qui tiennent en premier lieu à la diversité des approches et la manière spécifique dont chacune rend compte – en partie ou totalement – de la distinction entre le sujet parlant à l'origine de l'acte d'énonciation, l'instance qui dans l'énoncé se présente comme responsable de l'acte d'énonciation, et les instances à la source des points de vue impliqués dans un énoncé. Ainsi le terme d'énonciateur peut-il tout aussi bien référer dans la littérature à l'instance responsable de l'acte d'énonciation, le sujet parlant dans la perspective de O. Ducrot (1984) ou encore au locuteur au sens ducrotien (LOC). Cette polysémie justifie pour partie dans le cadre de la

¹ Cf. notamment 2004.

² Cf. 1999, 2002.

SacPoLine un refus – initial du moins – de recourir à cette notion d'énonciateur, sans pour autant qu'elle renonce à traiter la question de la source des points de vue.

La seconde difficulté que comporte ce terme d'« énonciateur » n'est pas étrangère à ce premier problème de distinction des instances. Le reproche que l'on peut faire à ce terme est d'encourager l'attribution de caractéristiques subjectives à l'instance ainsi désignée. Les théories énonciatives rendent tangible cette difficulté et la thématisent. Ainsi O. Ducrot (2001) désigne-t-il explicitement les « abus de langage » que ne manque pas de rencontrer son texte. Dans la théorie de la ScaPoLine, ce sont les guillemets qui viennent signaler ce type d'abus de langage quand est abordé le fait que les points de vue sont associés à des « sujets » ayant ces points de vue. Parler d'« être » ou de « sujet » paraît presque inévitable, mais ne va pas sans introduire de confusions. Les précisions qu'elles appellent induisent à leur tour de nouvelles complications, qui ne rendent pas d'accès facile ces théories.

Le problème est le même avec la notion de LOC. Les propriétés agentives ou subjectives que suggèrent les termes de locuteur ou d'énonciateur rendent confuse la distinction évoquée *supra* entre sujet parlant, locuteur et énonciateur. Elles impliquent en effet une forme de mimétisme entre ces différentes instances, tel que leur distinction n'aurait plus lieu d'être. Pourtant, il faudra y revenir, parler de sujet ou d'être n'est pas anodin, ni même attribuer à ces instances des propriétés non textuelles de sujet, d'agent, ou d'être, car c'est bien la représentation d'un sujet que livre parfois l'énoncé à travers les points de vue qu'il comporte. S'il revient à l'analyse polyphonique de spécifier en quoi ces instances peuvent ou non revêtir les attributs de sujets dans la manière dont elles se présentent dans l'énoncé, sans doute convient-il aussi dans un premier temps de se prémunir de ce type de formulation fortuite.

On pourrait dans cette perspective préférer au terme d'« énonciateur » celui d'« instance » ou de « source énonciative ». Ce choix pourrait cependant se voir objecter que la formulation ne lève aucunement l'ambiguïté : l'adjectif « énonciative » ne permet pas de savoir s'il est question d'énonciation ou d'énoncé. En définitive, comme il s'agit d'une instance linguistique, il importe de définir linguistiquement cette instance.

On peut alors se référer au cadre d'analyse linguistique de la polyphonie proposé par O. Ducrot (1984). Celui-ci définit les énonciateurs comme

des êtres qui sont censés s'exprimer à travers l'énonciation, sans que pour autant on leur attribue des mots précis ; s'ils 'parlent', c'est seulement en ce sens que l'énonciation est vue comme exprimant leur point de vue, leur position, leur attitude, mais non pas, au sens matériel du terme, leurs paroles (op.cité : 204).

Conformément au projet ducrotien d'une « extension à la linguistique des recherches de Bakhtine sur la littérature » (1984 : 173) on retrouve là des éléments mis en évidence par M. Bakhtine dans l'approche polyphonique du roman (1970) et dans la théorie dialogique qu'il développe pour une « esthétique de la production verbale » (1984).

Nous retiendrons de cela cette idée que les énonciateurs « parlent » au sens où en tant qu'instances à la source d'un point de vue, leur point de vue s'exprime dans l'énoncé. Cela rejoint ce que nous avons dit *supra* des voix impliquées dans un énoncé, en particulier dans le cas du conditionnel : on ne saurait dire dans le cas de « Paul viendrait demain » que l'énonciateur à la source du pdv1 « Paul viendra demain » parle. Dire que l'on entend sa voix, ou, mieux que l'énoncé fait entendre sa voix serait plus juste. Pour autant, la question d'une définition proprement linguistique de la notion n'est pas résolue. Dans cette optique, nous conviendrons donc de prendre comme point de départ le point de vue, pour aborder secondairement la question de sa prise en charge énonciative.

La définition du point de vue

Il importe d'approfondir la notion de point de vue pour mieux définir cette instance énonciative considérée comme étant à sa source. La notion de point de vue trouve son origine dans la critique littéraire. C'est dans ce cadre que J. Pouillon l'introduit en 1946, et c'est avec O. Ducrot en 1984 que la notion s'intègre à l'analyse de la polyphonie linguistique, qui n'a cependant pas l'exclusivité car la notion semble diversement mobilisée dans une perspective littéraire, linguistique, sémiotique ou herméneutique.

Retraçant « une histoire du point de vue », A. Rabatel (1997) souligne que l'analyse littéraire du point de vue tend à privilégier la question de sa source, ou du foyer de vision. Dans le champ linguistique, elle reste relativement peu théorisée chez O. Ducrot (1984), où elle est utilisée, comme le suggère H. Nølke et M. Olsen (2000), en concurrence avec la notion d'énonciateur.

A cette difficulté commune aux notions de point de vue et d'énonciateur s'ajoute, avec la notion de point de vue, un autre problème. On parle en effet communément de point de vue pour rendre compte de positions intellectuelles assumées par des sujets, acteurs du champ social, dans un sens finalement proche de ce qu'on entend par « avis » ou « opinion ». Dans la transposition didactique de ces catégories, ou dans leur mobilisation en contexte didactique du moins, on soulignera l'ambiguïté des attentes (s'agit-il de défendre une opinion, de marquer linguistiquement son point de vue, ou les deux), et les difficultés éprouvées par des étudiants pour lesquels l'exigence de marquer son point de vue n'évoque pas nécessairement les moyens linguistiques qui lui sont associés.

La rupture avec les conceptions doxiques, pour être nécessaire, ne doit pas intervenir seulement comme argument d'autorité. Plus même, il peut être utile une fois cette rupture entreprise d'envisager alors des éléments de continuité entre l'analyse linguistique du point de vue et ce que l'on nomme communément point de vue, au sens d'opinion, et qui paraît bien avoir partie liée avec la production linguistique du sens. Il conviendrait de s'interroger sur ce lien entre les points de vue observables dans l'énoncé et les points de vue en tant que positions intellectuelles assumées et défendues par les acteurs d'un champ social.

L'analyse de la polyphonie au niveau textuel s'inscrit dans cette perspective : elle doit permettre de cerner comment un point de vue, en tant qu'opinion assumée par un acteur, se construit linguistiquement, dans la multiplicité et la hiérarchisation des points de vue qui caractérisent la polyphonie du texte. L'idée est que, au moins dans certains genres, et en particulier dans le genre qui nous concerne, l'article de recherche, le texte affiche un point de vue cohérent, qui peut être mis au compte de son scripteur ou de son producteur effectif, au titre d'opinion, ou de thèse qu'il défend. Ce point de vue d'auteur se construit dans le texte, et résulte de la dimension polyphonique de ce dernier : le jeu des points de vue au sens linguistique doit être tel qu'il rende possible cette construction d'un point de vue cohérent, et indexable à l'auteur du texte.

Pour définir le point de vue et partant, la question de sa source en des termes linguistiques, on peut se référer à la théorie scandinave de la polyphonie linguistique (ScaPoLine) d'une part et aux travaux d'A. Rabatel d'autre part. Tous deux entendent répondre à cette exigence, dans un cadre d'analyse résolument ancré dans une perspective d'analyse des textes : même si la théorie scandinave se présente comme une théorie fondée sur l'énoncé, ses plus récentes évolutions mettent en avant l'idée que l'analyse linguistique menée au niveau de l'énoncé et de ses structures linguistiques sous-jacentes doit servir l'analyse des textes.

Avec l'analyse linguistique de la notion de point de vue, c'est une analyse des textes littéraires fondée sur leur matérialité linguistique qui est visée. Cette assise linguistique se veut utile plus largement à l'analyse de l'ensemble des textes et des pratiques langagières, en vertu notamment du lien entre polyphonie et argumentation, analysé dès le départ par O. Ducrot (1984) à travers la question des présupposés. L'intérêt de ces approches dans le cadre de notre étude est très net, s'agissant de définir les fondements d'une analyse des textes dans leurs dimensions énonciative et polyphonique, en tant que caractéristiques d'un genre.

Le problème du niveau d'analyse

En première approche, l'analyse polyphonique a de quoi dérouter ; non seulement les théories de la polyphonie sont multiples et complexes, mais le seul fait de s'attacher à cette description de la multiplicité des voix est contraire à une approche naïve du langage. Pourtant, l'analyse permet de décrire les phénomènes linguistiques observables dans les énoncés, et à partir de là, elle doit également servir à rendre compte de l'interprétation du texte, du moins dans la théorie scandinave de la polyphonie linguistique, mais aussi chez A. Rabatel comme nous le verrons.

La potentielle mécompréhension des théories polyphoniques est liée à une conception non problématisée de l'énoncé et à deux de ses difficultés : d'une part, celle de postuler que *je* renvoie de manière mécanique au responsable de l'énonciation et que l'énoncé est donc un simple reflet « interne » de ses « extérieurs », et d'autre part, celle de supposer que celui qui dit *je* est entièrement responsable de l'énoncé et de son sens, selon une illusion subjective de la parole. Dans cette conception spontanée pourrait-on dire de l'énoncé, la démultiplication des points de vue qu'y voit le linguiste est pour le moins déconcertante ; bien entendu certains phénomènes sont vraisemblablement moins concernés que d'autres : ainsi les publications sur la polyphonie insistent-elles souvent, en forme d'avertissement à un lecteur qui serait néophyte en la matière, sur le fait que la polyphonie ne se réduit pas à la citation ou aux formes marquées d'intrusion d'une voix explicitement signalée comme telle.

En ce sens, la négation descriptive ainsi analysée par O. Ducrot pourrait faire l'objet de résistances, par qui serait tenté d'y voir un jeu conscient instauré par le sujet parlant pour asseoir sa propre position. Cette difficulté d'analyse est d'ailleurs à l'origine même de la catégorie de la négation descriptive ; on lit dans O. Ducrot (2001) :

Pour vous signaler qu'il fait aujourd'hui un temps parfaitement beau, je peux aussi bien recourir à une phrase négative (« Il n'y a aucun nuage au ciel ») qu'à une phrase positive (« le ciel est totalement pur ») : il faudrait quelque tendance paranoïaque pour déceler dans la première une intention de m'opposer à quelqu'un, *et admettre en même temps que cette intention serait absente de la seconde* (si j'ai ajouté « et admettre en même temps... », c'est que, bien évidemment, tout énoncé, qu'il soit positif ou négatif, peut être interprété comme faisant allusion à un énoncé contradictoire, mais il s'agit là d'un fait pragmatique général, qui n'a pas de rapport spécifique avec la négation). C'est pour qualifier ces énoncés négatifs où la négation, en tant que telle, ne sert pas à l'opposition, que j'ai parlé de « négation descriptive ».

Opposée à la négation conflictuelle, qui regroupe la négation polémique et la négation métalinguistique, la négation descriptive est en elle-même, réduite à l'énoncé pris pour exemple, peu pertinente pour l'analyse polyphonique. Cela étant, si l'énoncé est envisagé dans un co(n)texte qui s'y prête, la structure polyphonique qu'il contient – virtuellement pourrait-on préciser – se manifeste clairement.

O. Ducrot (1984 : 216-217) souligne ainsi, à propos des deux possibilités pour la phrase dont il parle dans l'extrait cité d'être positive ou négative, la dissymétrie qui existe entre les deux par rapport à cette question de la polyphonie. Il montre qu'une phrase telle que

« Pierre n'est pas gentil » permet d'envisager un enchaînement comme « au contraire il est détestable », alors que l'affirmative « Pierre est gentil » ne saurait donner lieu à un énoncé comme « au contraire, il est adorable ». En toute logique, le « au contraire » ne s'oppose pas à l'assertion « Pierre n'est pas gentil » mais à l'affirmation (le p_{dv1} sous-jacent et nié) qu'implique la négation, y compris descriptive.

C'est ce qui permet à O. Ducrot d'argumenter en faveur de l'idée que « l'affirmation est présente dans la négation d'une façon plus fondamentale que la négation dans l'affirmation », et que la polyphonie est inscrite linguistiquement, en l'occurrence dans la négation. Nous ajouterons qu'elle est inscrite, de même, dans le « au contraire il est détestable » en réponse à un « il n'est pas gentil », ou encore dans un « c'est peu de le dire, il est adorable » suivant un « Pierre est gentil ». Mais à la polyphonie linguistique s'ajoute alors la « polyphonie »¹ de l'enchaînement des énoncés dans le dialogue ; M. Bakhtine appelait à distinguer ces deux dimensions, pour se centrer sur la polyphonie linguistique, manifeste dans le « au contraire », ou dans « c'est peu de le dire ». Développant les propositions de M. Bakhtine, c'est bien celle-ci que O. Ducrot (1984) entend décrire, non dans des échanges mais dans des énoncés apparemment monophoniques comme le suggérait son prédécesseur.

Si l'on revient à la structure polyphonique de la négation, il s'agit alors pour O. Ducrot d'envisager différents types de négation en fonction de leur co(n)texte : « il n'y a pas un nuage au ciel » est une négation descriptive si l'énoncé répond à une question sur le temps qu'il fait, et polémique dans le cas où l'affirmation vient contrer une affirmation inverse si l'énoncé répond à une assertion comme « il y a quelques nuages ». Fondée sur des énoncés, cette analyse de la négation appelle donc à envisager des suites d'énoncés pour voir comment, sous l'angle pragmatique-sémantique proposé par O. Ducrot, la négation sert une opposition entre énonciateurs et parfois entre locuteurs.

On pourrait répéter la démonstration pour le bien connu « ce mur n'est pas blanc », qui peut susciter des réserves similaires quant à sa dimension polyphonique. L'analyse polyphonique a un objectif de description des phénomènes linguistiques observables dans l'énoncé, mais manifeste aussi son intérêt si on associe à cet énoncé d'autres énoncés, et si on vise l'analyse des textes, qu'il s'agisse d'analyse des interactions, d'études sur l'argumentation, ou encore du texte littéraire. Bien entendu, telle n'est pas la perspective d'O. Ducrot, mais il s'agirait dans cette optique de voir comment le jeu des points de vue sert le point de vue du LOC, autrement dit de voir comment le texte construit un point de vue cohérent et sa dimension argumentative, au sens où l'entend R. Amossy (2000 : 24) de projet inscrit dans le texte d'amener le destinataire à modifier telle ou telle de ses représentations, autrement dit de faire apparaître ce point de vue comme légitime et convaincant pour un éventuel lecteur.

Cependant, comme les analyses polyphoniques ont d'abord montré leur pertinence à partir d'énoncés fabriqués, proches d'énoncés attestables, sinon authentiques, mais tout à fait décontextualisés et souvent restreints à un cadre phrastique (typiquement, « ce mur n'est pas blanc »), se pose la question délicate du passage entre l'analyse linguistique du point de vue

¹ Nous mettons ici le terme entre guillemets, car la tendance actuelle est de ne désigner par « polyphonie » que la polyphonie linguistique, et de préférer les termes de « dialogisme » et de « dialogue » au niveau de la production des énoncés, alors que les deux acceptions du terme polyphonie sont attestées chez M. Bakhtine. Rappelons ici aussi, mais nous avons eu l'occasion de le dire, que le terme « dialogisme » est cependant employé pour désigner la polyphonie linguistique chez J. Bres et B. Vérine (2002).

au niveau local et l'analyse de la polyphonie au niveau du texte¹. La multiplication des voix observable au niveau local se complexifie grandement en effet au niveau du texte.

Ne serait-ce qu'en s'en tenant à un énoncé comme « ce mur n'est pas blanc », la question du nombre de points de vue que l'on pourrait y voir nous semble difficile. On a décrit *supra* les deux points de vue impliqués par la négation, mais le pdv1 « ce mur est blanc » est-il un point de vue dit « simple » dans la théorie scandinave, c'est-à-dire un point de vue monophonique ? Ou peut-il être lui-même analysé comme un point de vue complexe, décomposable en structures du type « ceci est un mur » et « ce mur est blanc » ?² Nous ne prétendons pas ici répondre de manière tranchée à une telle question, ni assimiler l'analyse du point de vue initiée par l'approche polyphonique de O. Ducrot à ce type de questionnement ; cependant, à titre programmatique, soulignons qu'une notion comme le dialogisme de la nomination développée notamment dans le cadre de l'approche praxématique (cf. Détrie et al., 2001, Cassanas et al. 2003, Siblot, 2001), pourrait vraisemblablement offrir une prise pour l'analyse de la dimension polyphonique à la source de la présupposition existentielle qui caractérise le langage³.

Retenons simplement ici que non seulement l'analyse linguistique du point de vue mérite d'être approfondie, mais également son intérêt pour l'analyse des textes. Par ailleurs, si le passage de l'analyse locale ou pourrait-on dire, quasi grammaticale de la polyphonie, telle qu'elle est menée dans des énoncés restreints au cadre phrastique, à la polyphonie au niveau textuel est complexe, en raison de la prolifération des voix qui se manifeste, nous verrons que l'objet texte lui-même doit permettre de contenir en partie cette prolifération infinie des points de vue qui y est observable.

Comme c'est précisément ce double objectif de conférer une assise linguistique à la notion de point de vue et de fournir un cadre à l'analyse de la polyphonie au niveau du texte qui caractérise la ScaPoLine d'une part et les travaux de A. Rabatel d'autre part comme nous l'avons dit, nous évoquerons successivement ces deux approches.

Dans la perspective d'un cadre pour l'analyse des textes sur la base de leur dimension polyphonique, trois pistes guident notre lecture de ces deux approches : il s'agit de mieux cerner comment est défini le point de vue, comment est traitée la question de sa prise en charge, et de voir comment les auteurs se proposent de rendre compte de l'analyse de la polyphonie au niveau global du texte.

2.3.3. L'analyse du point de vue et de la polyphonie dans la théorie scandinave de la polyphonie linguistique

Le cadre général d'analyse de la ScaPoLine

En son point de départ, la ScaPoLine se donne comme objet la structure polyphonique en tant que fait de langue qui relève du niveau de la phrase ; secondairement elle s'intéresse également à la configuration polyphonique qui se présente quant à elle au niveau de l'énoncé (Olsen, 2002, Nølke, Fløttum et Noren, 2004). Celui-ci est entendu en adéquation avec les

¹ Cf. la « Note sur la problématique des niveaux de l'analyse polyphonique - de la phrase au texte » de K. Fløttum, 2000.

² La question rejoint celle que pose P. Dendale, « y a-t-il oui ou non des énoncés monophoniques ? », dans la rubrique questions du site « PolyphoniE » (<http://webhost.ua.ac.be/polyphon/>).

³ Cf. ce que dit P. Siblot (2001) : « toute nomination exprime une vision de la chose nommée, vue 'sous un certain angle', à partir du 'point de vue' auquel se place le locuteur ».

fondements des approches énonciatives *décrits supra* comme le résultat de l'énonciation de cette phrase, résultat qui fournit une image de son énonciation.

Les auteurs signalent qu'à l'instar de O. Ducrot (1984), ils parlent d'énoncé en tant que segment textuel doté d'un sens autonome et de phrase en tant qu'unité de la langue qui sous-tend l'énoncé. La description sémantique de l'énoncé et la description sémantique de la phrase sont désignées respectivement par les termes de sens et de signification. Cette dernière, objet d'étude dit « immédiat » de la théorie scandinave de la polyphonie linguistique, est envisagée comme se composant d'instructions à construire le sens de l'énoncé, et supposées en cela rendre compte des interprétations effectives d'un énoncé.

L'analyse prend donc comme point de départ l'énoncé et la polyphonie qui y est observable, pour rendre compte des instructions de la structure polyphonique, en tant que contraintes sur l'interprétation de la configuration. Si la structure polyphonique est affichée dans les premiers textes comme un objectif prioritaire, on note dès le départ cette volonté de mieux comprendre la construction du sens à partir des instructions, c'est-à-dire des éléments marqués dans la signification. Dans la perspective énonciative de la ScaPoLine, ces marques sont envisagées comme marques des éléments de la configuration polyphonique.

Dans les révisions et développements les plus récents de la théorie (Nølke et al., 2004), relier forme et sens est explicitement donné comme objectif d'analyse, puisqu'il s'agit de « préciser les instructions qu'apporte la forme linguistique pour l'interprétation du texte ». Le passage entre la structure polyphonique et la configuration polyphonique devient donc l'enjeu central et se fonde sur les notions de variables et de saturation. Nous aurons l'occasion de le détailler et de l'exemplifier par la suite, mais le modèle général qui est proposé est le suivant : dans les limites permises par les instructions posées par la signification (en langue), les stratégies interprétatives consistent en une saturation des variables et de leurs relations. Les variables (i.e. ce que nous avons appelé *supra* les énonciateurs) font partie des instructions, qui donnent également des indications relatives à leur saturation (i.e. leur indexation à des êtres-discursifs) et c'est le co(n)texte qui permet cette saturation relevant de l'interprétation.

On notera également que les évolutions de la ScaPoLine attestent d'un intérêt plus résolument ancré dans une perspective d'analyse des textes. Au départ, la distinction langue/énoncé ou phrase/énoncé est centrale ; elle rejoint ce que nous avons dit des approches énonciatives, intéressées par la question de l'actualisation de la langue dans les énoncés. Cela étant, elle nous paraît quelque peu problématique dans la mesure où les énoncés sont des énoncés fabriqués, restreints au cadre phrastiques ; l'intérêt est bien entendu de décrire des phénomènes linguistiques dans l'énoncé, mais c'est alors la question de la matérialité linguistique des énoncés qui doit à notre sens être posée. Aussi, reformulée en des termes de type forme/sens, la distinction s'inscrit nous semble-t-il dans une visée particulièrement intéressante, celle de chercher à rendre compte de ce qui, dans le texte, contraint son interprétation.

Le cadre d'analyse reste ainsi en adéquation avec celui de la polyphonie linguistique proposé par O. Ducrot : il s'agit bien de fonder l'analyse sur du matériau linguistique, et ce malgré les complications qu'introduit à notre sens l'opposition initiale entre langue et énoncé. Dans notre perspective, le texte est du matériau linguistique, de même que l'énoncé. Cela étant, il reste que l'originalité de la ScaPoLine est de tenter de véritablement cerner le problème du passage entre l'analyse d'énoncés souvent restreints et décontextualisés et l'analyse de la polyphonie au niveau du texte. C'est cet objectif renouvelé par rapport à l'approche initiale que précise H. Nølke en 2001 :

La ScaPoLine est une théorie discursive. Si son domaine (immédiat) est l'énoncé, elle n'oublie jamais que l'énoncé ne trouve sa raison d'être que dans le fait qu'il fasse partie d'un discours. Ce fait se reflète entre

autres dans les outils heuristiques dont elle se sert. C'est ainsi que les tests d'enchaînement différents servent comme son outil heuristique primordial. Il semble donc naturel d'essayer d'élaborer la ScaPoLine de manière à la faire fonctionner au delà de l'énoncé. Nous avons déjà fait quelques propositions concernant le développement d'une application proprement textuelle de la ScaPoLine et cet aspect va occuper une place centrale dans nos investigations à venir. Cette élaboration est en effet strictement nécessaire dans le cadre du projet des polyphonistes scandinaves, car ce n'est qu'au niveau textuel que peuvent se rencontrer linguistes et littéraires.

Les points de vue et l'instanciation de leur source

La question se pose alors de savoir comment aborder cette relation entre les points de vue dans l'énoncé et le niveau du texte. Soulignons à ce stade que l'enchaînement des énoncés évoqué ci-dessus par H. Nølke¹ n'est un outil au service de l'analyse de la polyphonie textuelle que dans la mesure où les points de vue ne sont pas seuls concernés ; comme le souligne H. Nølke (ibid.) tout le problème de l'analyse polyphonique au niveau du texte tient à la question de « l'instanciation des variables que sont les énonciateurs ». Pour rendre compte de la manière dont les variables de la structure polyphonique (analysée dans un énoncé restreint à une phrase ou dans un enchaînement d'énoncé) peuvent être saturées au niveau de l'interprétation du texte, la ScaPoLine entend déterminer les liens entre les points de vue et les être-discursifs (Nølke et al., 2004).

Les points de vue font l'objet d'une définition sémantique dans la théorie scandinave : ils sont envisagés comme une unité sémantique qui peut être déclinée en deux composantes, un contenu et un jugement.

Cette analyse du point de vue semble faire écho à celle que C. Bally (1909) introduit à propos de l'énoncé et non du point de vue, à savoir que l'énoncé comprend un dictum, le dit, et un modus, qui renvoie à l'attitude de celui qui dit sur le dit. C. Noren (2000) montre justement que la définition du point de vue donnée par H. Nølke dans la ScaPoLine se rapproche clairement de la conception de H. Kronning, qui propose de voir dans le point de vue un modus et un dictum (1996 : 44).

On connaît les limites de cette distinction au niveau de l'énoncé : pour être pertinente, elle reste peu opératoire, car il est difficile de distinguer dans la matérialité linguistique les deux composantes (le modus n'est pas marqué ou il se confond avec le dictum).

A propos de l'unité « point de vue », c'est donc le même problème que pose un point de vue dit simple dans la ScaPoLine, dans le cas d'un énoncé monophonique où un seul point de vue intervient (par exemple, dans « il fait beau »). En ce cas, dire que le point de vue se compose d'un contenu et d'un jugement ou d'une attitude sur ce point de vue reste difficile à décrire.

La difficulté peut être levée dans la distinction de C. Bally si l'on considère que le modus renvoie à la question de la prise en charge énonciative. Dans ce cas, on peut concevoir, que, même non marqué dans un énoncé comme « il fait beau », le modus rend compte de la présence d'un énonciateur (en l'occurrence d'un LOC, i.e. l'instance qui dans l'énoncé se présente comme l'ayant produit).

Au niveau du point de vue tel qu'il est analysé dans la ScaPoLine, ce ne peut cependant semble-t-il être la question de la prise en charge modale du contenu qui est visée par l'autre composante qu'est le jugement : cette question est résolue différemment, au moyen d'une variable, dite énonciateurs, et des liens énonciatifs qui permettent la saturation par un être discursif de cette variable, comme nous y reviendrons plus loin.

¹ Notons qu'en premier lieu, l'enchaînement des énoncés est envisagé dans la ScaPoLine comme un outil pour déterminer si on a affaire ou non à un point de vue, comme le souligne C. Noren (2000) qui montre toutefois les limites de ce critère.

Si l'on reprend l'exemple de « il fait beau », on a affaire à un point de vue monophonique, dont la variable « énonciateurs » est saturée par l'être discursif qui correspond au locuteur (LOC), ce qui revient à dire que le locuteur est attaché à ce pdv « il fait beau » par un lien énonciatif : celui-ci est envisagé comme une instruction, correspondant à une règle selon laquelle en l'absence de mention contraire, la valeur par défaut veut que le locuteur¹ soit à la source de ce point de vue, c'est-à-dire dans la perspective de la ScaPoLine, responsable de ce point de vue. Tout en approfondissant la question du modus et en la posant à partir du point de vue plutôt que de l'énoncé, la ScaPoLine rend donc bien compte de ce qu'est le modus dans ce type d'énoncés monophoniques.

Si la composante modus de l'énoncé monophonique est réglée par la question de la saturation de la variable « énonciateur » attachée au point de vue, comment alors interpréter ce couple contenu-jugement qui définit le point de vue – et partant l'énoncé monophonique, où un point de vue simple se manifeste ? A notre sens, la composante jugement du point de vue peut en quelque sorte être envisagée comme le nœud qui rend possible le lien entre le point de vue et la saturation par des êtres discursifs dans la perspective scandinave, ou le lien entre le point de vue et l'affectation d'une source ou d'une instance de responsabilité de ce point de vue.

Si l'on poursuit cette idée, dans un énoncé monophonique comme « il fait beau », on pourra donc poser :

- le point de vue ou association d'un jugement et d'un contenu,
- la variable énonciateur, le lien qui permet la saturation de cette variable par un être-discursif (en l'occurrence le locuteur, instance se présentant comme responsable de l'énoncé).

Plus rapidement, on dira que le locuteur de l'énoncé est responsable de ce point de vue. Après ce détour digne des poupées russes, on rejoint donc dans le cas des énoncés monophoniques une description finalement naïve du point de vue : « il fait beau » énonce un point de vue qui est à la fois un contenu (quelque chose de dit) et quelque chose qui est dit ou pensé.

Mais dans le cadre d'une approche polyphonique, ce cas est à envisager comme une exception : A. Culioli (1990 : 150) le notait également, le cas de coalescence entre le locuteur et l'énonciateur (ou même, dans la perspective polyphonique, le responsable du point de vue) est à envisager comme une exception. La déconstruction d'une approche naïve et la multiplication des instances qui en découle sont donc loin d'être fortuites, puisqu'il s'agit pour la ScaPoLine de rendre compte des instructions par lesquelles un point de vue peut-être rattaché au niveau du texte à des êtres-discursifs qui viennent saturer la variable énonciateur en fonction des liens énonciatifs.

N'y a-t-il pas cependant une forme de redondance dans le fait d'analyser la structure interne du point de vue en termes de contenu propositionnel et de jugement porté sur ce contenu, alors que la prise en charge du point de vue, i.e. sa composante modale, est résolue par ailleurs en termes de liens énonciatifs au départ, puis, dans les évolutions de la ScaPoLine, avec la variable énonciateur et sa saturation au moyen des liens énonciatifs qui relient points de vue et êtres discursifs ? Ces liens énonciatifs ne caractérisent-ils pas déjà la dimension de « jugement » attribué à tout point de vue ?

Cette redondance justifie dans un premier temps le refus par la ScaPoLine de reprendre la notion d' « énonciateur » telle qu'on la trouve chez O. Ducrot (1984) : comme le souligne C. Noren (2000), la notion est en effet jugée « superflue puisque le pdv implique obligatoirement que quelqu'un voit ». Traitée en termes de liens énonciatifs et d'êtres-discursifs, la question de la prise en charge des points de vue n'échappe pas complètement à

¹ Ou plus précisément, dans la perspective de la ScaPoLine, l0, locuteur de l'énoncé, nous y reviendrons.

ce problème de redondance, sans compter que la variable « énonciateurs » se trouve finalement réintroduite pour mieux mettre en évidence la proximité et de la compatibilité du cadre d'analyse proposé par la ScaPoLine avec celui de O. Ducrot : H. Nølke (2001) affiche cet objectif de mieux faire ressortir l'analogie entre la ScaPoLine et l'approche de Ducrot, suite à certaines mécompréhensions que pointe également O. Ducrot (2001), tout en soulignant que cette réintroduction de la notion d'énonciateurs est une « complication inutile dans l'économie du système ».

Si l'on propose donc de voir dans la composante dite « jugement » du point de vue un nœud qui rend possible le lien énonciatif par lequel un point de vue se trouve rattaché à un être discursif qui vient saturer la variable « énonciateur », on conviendra que l'appareil conceptuel ainsi constitué paraît quelque peu redondant ; il a cependant le mérite d'insister sur le fait que le point de vue comporte une dimension modale inter-reliée à sa dimension référentielle déjà mise en évidence par H. Kronning (1996 : 44), *a contrario* de O. Ducrot pour qui la question de la référence est exclue d'une analyse de la langue en vertu de la position structuraliste en linguistique.

On voit à nouveau que l'opposition entre langue et énoncé tend à opacifier les théories polyphoniques. Elles gagnent nous semble-t-il à se donner pour objet le texte, dans sa matérialité linguistique, en rendant compte des relations entre ce qui est observable au niveau de l'énoncé restreint au cadre phrastique et ce qui est observable dans leur enchaînement au niveau textuel.

L'idée que le point de vue, en tant qu'unité sémantique, associe quelque chose de dit (un contenu propositionnel) et une composante modale rejoint alors une idée assez largement admise en linguistique, quoique à des degrés divers, selon laquelle le langage ne saurait être envisagé exclusivement en termes de propositions sur une réalité empirique qui existerait indépendamment de lui : comme l'affirme J. Habermas (1973) pour ne citer que lui, ces contenus langagiers sont aussi des façons de produire le rapport au monde.

Aussi la conception des points de vue et de leur prise en charge dans la ScaPoLine ne passe-t-elle pas sous silence la question de la vérité et de l'assertion, qui servent généralement de fondements à l'analyse de la prise en charge énonciative comme le rappellent P. Dendale et D. Coltier (2005). Si celui-ci semble reprocher à la ScaPoLine de ne pas traiter explicitement ces questions autrement que par l'intermédiaire des points de vue et de leur prise en charge, c'est bien du statut énonciatif de l'assertion et de la question logique de sa vérité qu'il est question dans le fait de voir dans le point de vue une composante modale, interne au point de vue, et des liens énonciatifs par lesquels le point de vue peut se voir attribuer à ces entités sémantiques que sont les êtres-discursifs.

Les êtres discursifs et la saturation des points de vue dans le texte

Que représentent au juste ces êtres-discursifs intégrés à la théorie scandinave de la polyphonie linguistique ? On a vu qu'il s'agissait d'entités sémantiques dont la propriété est d'être susceptibles de saturer la variable énonciateur, c'est-à-dire de se présenter comme la source d'un point de vue. Un être-discursif lié à un point de vue par un lien de responsabilité (ou lien énonciatif) est dit être à la source du point de vue. Être source ou responsable du point de vue signifie dans la ScaPoLine être responsable de la composante jugement du point de vue.

Initialement définis de manière alternative comme des personnes linguistiques présentes dans le texte ou comme entités de l'univers de discours, les êtres-discursifs tendent dans les évolutions de la ScaPoLine à désigner des « images » de différents personnages présents dans le discours ; notons que ce dernier est parfois présenté comme équivalent de

« texte ». Là encore la question de l'objet de l'analyse n'est pas toujours très claire, mais l'idée est là que les êtres discursifs, comme les liens énonciatifs, peuvent être marqués dans la signification mais ne le sont pas forcément, auquel cas des règles permettent cependant de construire le sens approprié.

Les êtres-discursifs sont considérés comme étant construits par le locuteur. Rappelons que le locuteur fait l'objet d'une définition reprise de celle proposée par O. Ducrot : il est un élément du sens qui dans l'énoncé se présente comme le responsable de l'énonciation. Élément constitutif de la configuration polyphonique, le LOC est envisagé comme constructeur de sens et metteur en scène de cette configuration. Ainsi le LOC n'est-il pas un être-discursif : il est d'une part metteur en scène des êtres-discursifs, et d'autre part, il peut construire des êtres-discursifs comme images de lui-même.

Parmi l'ensemble des êtres-discursifs repérables, certains sont donc des images de LOC, mais il est bien précisé que ces images de LOC ne sont pas les seuls personnages qui se manifestent ; en particulier, on lit dans la ScaPoLine III (Nølke, 2001) que LOC construit aussi des images de l'autre instance énonciative qu'est l'ALLOC (élément de sens qui dans l'énoncé se présente comme celui à qui s'adresse l'énonciation).

Plus largement, les êtres-discursifs font l'objet d'une distinction entre diverses catégories, en fonction des attributs énonciatifs de ces êtres discursifs. La ScaPoLine propose ainsi de voir dans les êtres-discursifs soit des locuteurs virtuels soit des non-locuteurs : les locuteurs virtuels sont des êtres discursifs se présentant comme des images de locuteurs, c'est-à-dire des personnages susceptibles de prendre la parole, tandis que les non-locuteurs sont des êtres-discursifs plus abstraits, qui ne sont pas présentés comme ayant cette propriété de pouvoir produire eux-mêmes une énonciation.

Dans cette perspective, les images du LOC et de l'ALLOC correspondent à des locuteurs virtuels, avec les « troisièmes », définis comme des locuteurs virtuels indiqués sur un mode délocuté, par des pronoms de la troisième personne ou des syntagmes nominaux. Les non-locuteurs trouvent une marque privilégiée dans un *on* polyphonique, qui renverrait à la loi, les idées reçues, la vérité éternelle.

Pour rendre compte de ces différentes formes sous lesquelles se présente la source d'un point de vue, les locuteurs virtuels font l'objet dans la ScaPoLine d'une nouvelle distinction avec les vrais locuteurs (qui sont eux aussi des instances relevant du sens de l'énoncé) : si le locuteur virtuel a les propriétés d'un vrai locuteur mais ne s'en sert pas, le vrai locuteur actualise cette propriété virtuelle en se présentant comme prenant la parole. La distinction vrai locuteur/ locuteur virtuel peut donc également être comprise comme une distinction entre un locuteur virtuel dont les propriétés de locuteur sont actualisées dans le texte, et un locuteur virtuel dont les propriétés de locuteur ne sont pas actualisées.

C'est ainsi que le conditionnel, en accord avec l'analyse que l'on en a fait au début de cette partie, est envisagé comme impliquant un locuteur virtuel : le point de vue sous-jacent a comme source une instance qui se présente comme pouvant virtuellement énoncer ou avoir énoncé ce point de vue. Partant de là, si l'on associe à « Paul viendrait demain » une spécification du type « selon Marie », le locuteur virtuel impliqué par le conditionnel se trouverait actualisé par la désignation « Marie ». On aurait ainsi dans cet énoncé deux êtres-discursifs ou deux images de « vrais locuteurs » : le locuteur de l'énoncé et le locuteur « Marie ».

On voit donc là encore la proximité de l'analyse des êtres-discursifs avec celle des énonciateurs dans le cadre de O. Ducrot (1984) et dans d'autres analyses menées à sa suite, par exemple chez J. Bres et B. Vérine (2002) : il s'agit de distinguer des sources de points de vue qui se présentent comme disant un point de vue, ou pouvant le dire, et celles qui se

présentent à la source d'un point de vue sans pouvoir être assimilées à des sources locutoires capables de produire un énoncé.

Cette distinction n'est pas thématifiée chez O. Ducrot qui suggère cependant, on l'a vu, que les énonciateurs ne s'expriment pas tous au même titre. Par contre, elle fonde chez J. Bres et B. Vérine la mise en place d'un cadre d'analyse du dialogisme qui intègre les formes explicites comme le discours rapporté, où c'est un locuteur qui est mis en scène dans le texte, et les tours où l'énonciateur ne se présente pas comme locuteur, comme par exemple en l'absence de mention du type « selon Marie » dans un énoncé comme « Paul viendrait demain ».

De manière similaire dans la ScaPoLine qui rend compte du marquage, ou, en l'absence de marquage, des règles qui peuvent s'appliquer, la question de l'instanciation de la source des points de vue présente comme on l'a suggéré *supra* un intérêt pour différencier les tours polyphoniques d'une part, et pour analyser la manière dont les points de vue peuvent être rattachés au LOC au niveau du texte, à travers ses différentes images.

Sur ce point, la ScaPoLine poursuit l'analyse de O. Ducrot en proposant de distinguer deux images du LOC, images qui se retrouvent également pour l'ALLOC (Noren, 2000) et pour les troisièmes : un locuteur de l'énoncé et un locuteur du texte. Cette scission a été initiée par O. Ducrot concernant le seul locuteur dans son analyse des présupposés en 1982, et formulée au moyen des deux catégories du « locuteur être du monde » et du « locuteur en tant que tel ».

Malgré la différence terminologique justifiée dans la ScaPoLine par l'ambiguïté qu'introduit l'« être du monde » pour désigner un être discursif, il s'agit dans les deux cas de rendre compte du fait que les images données dans l'énoncé du LOC responsable de l'énonciation peuvent être diverses ; H. Nølke (2001) en donne comme exemple « je me demande si... », où *je* et *me* n'ont pas exactement la même référence et propose l'analyse suivante : *je* renvoie à un être défini uniquement par le fait d'être responsable de l'énonciation (et en l'occurrence, de l'acte de langage pourrait-on ajouter), et *me* renvoie à une image du locuteur qui est indépendante de l'énoncé particulier.

Cette distinction nous paraît intéressante dans la mesure où elle permet de rendre compte de certains phénomènes sémantiques et pragmatiques (par ex., la présupposition). Elle suggère aussi des pistes d'analyse quant à la manière dont se construit la subjectivité dans les textes : si l'on admet que c'est le fait de dire *je* qui constitue le sujet en tant que tel, on voit ici comment, dans les textes, se constitue l'image d'un sujet permanent par delà la multiplicité des énoncés qu'il produit, qui se trouve posé comme sujet non seulement par le fait de dire *je* mais par le fait d'être présenté selon une image générale de lui-même indépendante de cet acte particulier. La notion de locuteur du texte tend rejoint ainsi celle de figure de l'auteur, bien qu'il ne s'agisse pas dans la ScaPoLine d'envisager le lien entre le locuteur du texte et le sujet parlant à l'origine du texte, au contraire de ce que peut suggérer la notion de figure de l'auteur, et qui est plus central encore avec une notion comme celle d'image du scripteur.

Cependant, le fait de distinguer un locuteur de l'énoncé, au sens d'image de locuteur au moment de la parole et un locuteur textuel au sens d'image générale du locuteur ou d'image du locuteur à un autre moment de son histoire pose plusieurs difficultés. La première est terminologique : la distinction entre énoncé et texte concerne les niveaux d'analyse comme nous avons déjà eu l'occasion de le dire. Aussi, elle est intéressante pour identifier le locuteur de l'énoncé ; par contre, le fait de voir dans le locuteur du texte « une image du locuteur à un autre moment de son histoire » semble indiquer que dans le texte se construirait l'image d'un sujet extérieur au texte même, et à ce titre, la désignation de locuteur être du monde chez O. Ducrot serait plus pertinente, quoiqu'elle demande de bien distinguer cette instance du sujet parlant.

Finalement, c'est donc la question du lien entre le LOC et le sujet parlant, ou producteur effectif de l'énoncé qui affleure dans les théories polyphoniques, alors même qu'elle est écartée de l'objet d'une telle théorie, pour les raisons que nous avons évoquées en traitant des notions d'énoncé, d'énonciation, et d'énonciation représentée.

La seconde difficulté avec les notions de locuteur de l'énoncé et de locuteur du texte est de déterminer si cette bi-partition est toujours possible, autrement dit si elle est marquée textuellement ; si elle nous semble heuristique, elle peut aussi limiter la réflexion sur la construction du sujet dans le texte en supposant deux catégories qui pourraient aussi être envisagées plutôt sous l'angle de leurs relations.

Ces deux catégories ont ceci de problématique qu'elles tendent, nous semble-t-il, à projeter au niveau de l'énoncé des caractéristiques supposées de l'acte d'énonciation. On a en effet coutume d'envisager cet acte en distinguant une situation *hic* et *nunc*, décrite au moyen des trois paramètres suggérés par les déictiques que sont les acteurs, le temps et le lieu de la situation, et de convenir que cet acte se déroule dans un contexte plus général (par exemple, l'histoire conversationnelle entre les acteurs, leurs connaissances respectives et partagées). Les deux facettes des images du LOC dans le texte correspondraient ainsi à ces deux dimensions jugées nécessaires pour rendre compte du problème de la référence. En tout état de cause, il reste à évaluer dans quelle mesure la distinction est véritablement opératoire pour l'analyse des textes, et en quoi elle peut être utile à cette analyse. Il nous reviendra donc de la traiter dans le genre spécifique de l'article de recherche (cf. au chapitre 7).

Cette question s'intègre plus largement à celle de la prise en charge énonciative pour ce qui concerne le LOC, quoiqu'une originalité de la ScaPoLine est de tenter de l'appliquer également à l'ALLOC et aux dits « troisièmes »¹, pour voir en quoi, par exemple dans un énoncé comme « selon Marie, Paul viendrait demain », on peut distinguer en « Marie » un locuteur virtuel délocuté qui se présenterait comme locuteur *hic* et *nunc* d'un énoncé, ou comme locuteur « être du monde » comme le désigne O. Ducrot. On retiendra que pour le seul LOC, il importe de garder à l'esprit la complexité déjà pointée par M. Foucault (1969b) de ces images données du LOC dans l'énoncé : le seul *je* peut être analysé en multiples facettes qui ne se limitent pas à ces deux-là comme il le montre pour un énoncé scientifique où l'on peut voir un locuteur du texte, un locuteur de la démarche, un locuteur générique, mathématicien, etc.²

C'est le même problème que posent d'autres distinctions dans la ScaPoLine, comme celle entre les êtres discursifs indivisibles et les êtres discursifs divisibles, les premiers se présentant dans l'énoncé comme ne pouvant être conçus comme un collectif. On notera là encore une difficulté terminologique, un collectif (par ex. la Loi, l'Homme) n'étant pas davantage divisible, quoiqu'il doive pouvoir intégrer des êtres discursifs indivisibles (par ex. une image du LOC). Mais plus largement, les tentatives de distinctions des instances de prise en charge énonciative selon la manière dont elles se présentent dans le texte nous paraissent être délicates car elles se heurtent au fait que dans les textes, c'est vraisemblablement leur continuité et leur perméabilité qui est observable.

Pour finir, il nous faut enfin mentionner une dernière difficulté sur cette question du LOC, des êtres discursifs et de la prise en charge énonciative en termes de source ou de responsabilité des points de vue : l'idée récurrente dans la ScaPoLine que LOC met en scène et construit toutes ces images et tous ces faits textuels nous dérange quelque peu ; c'est la

¹ Les troisièmes sont des locuteurs virtuels, ou « êtres-discursifs présentés par le locuteur comme étant des images de (autres) locuteurs », et qui ne sont ni indexables au LOC ni à l'ALLOC, par exemple Marie dans l'exemple donné.

² M. Ali Bouacha (1995) aboutit à un constat similaire à propos du sujet tel qu'il se manifeste dans le discours philosophique.

même réserve que nous formulerions à l'égard de la désignation de la composante modale du point de vue en termes de jugement¹.

Si le jeu polyphonique est inscrit linguistiquement, et/ou inférable à partir de règles, il n'y a pas de raison pour dire qu'un LOC construit ou gère ces différentes instructions du jeu polyphonique. Le LOC est l'instance qui dans l'énoncé se présente comme responsable de son énonciation, et nous défendrons l'idée que le LOC ne construit ce jeu polyphonique que dans la mesure où ce jeu polyphonique donne de lui l'image d'une instance construisant ce jeu.

En d'autres termes, ce sont les caractéristiques de l'énoncé ou du texte qui contribuent à donner l'image d'une instance LOC, et d'une instance LOC qui mène le jeu polyphonique. Le jeu polyphonique doit être tel que se manifeste dans l'énoncé une instance qui s'y présente comme son responsable, et comme responsable du jeu polyphonique. En rappelant que même en disant *je*, un texte ne révèle pas son auteur réel, mais seulement un auteur supposé, M. Amorim (1995 : 88) souligne que M. Bakhtine notait déjà cette contradiction que nous pointons ici à propos du LOC : traitant pour sa part de l'« image de l'auteur », il montre que si l'image est créée, elle ne peut pas être créatrice. A notre sens, si on peut lui attribuer la propriété d'être créatrice, c'est à la condition que l'image créée implique cette propriété créatrice.

Il reste que pour être importantes, ces précisions ne doivent pas masquer l'intérêt de la démarche entreprise par la ScaPoLine : elle fournit des descriptions fines pour déterminer dans quelle mesure et en vertu de quelles règles les points de vue peuvent être rattachés aux êtres-discursifs par des liens de responsabilité². Elle rend ainsi possible une typologie des phénomènes polyphoniques, et offre des entrées textuelles à l'analyse de la construction du jeu polyphonique dans le texte.

2.3.4. L'analyse du point de vue et des postures énonciatives chez A. Rabatel

L'analyse du point de vue et de sa prise en charge chez A. Rabatel

Le projet d'A. Rabatel est *a priori* tout à fait différent de celui de la ScaPoLine, mais la rejoint nous semble-t-il sur un certain nombre d'aspects. S'inscrivant dans une perspective narratologique, l'auteur entend donner une assise véritablement linguistique à la notion de point de vue. A travers ses deux ouvrages de 1997 et 1998, il analyse les approches historiques qui ont investi la notion, et discute en particulier les travaux de J. Pouillon (1946) ou de G. Genette (1972, 1983), qui sont étrangers à la théorie polyphonique développée par O. Ducrot. L'objectif d'A. Rabatel est de dépasser les limites de ces approches en proposant des outils d'analyse de la construction textuelle du point de vue.

L'approche est résolument centrée au départ sur le texte littéraire : il s'agit de rendre compte de la construction linguistique de ce que l'on appelle le personnage et le narrateur, en prenant le contre-pied des approches antérieures qui cherchent à se saisir d'emblée de ces deux « candidats focalisateurs » : l'auteur reproche à ces approches une substantialisation excessive du point de vue, et montre au contraire l'importance de s'en tenir en premier lieu au point de vue lui-même. Bien que la perspective soit très différente, on retrouve donc le même parti pris que dans la ScaPoLine de prendre comme point de départ le point de vue, plutôt

¹ C'est ce que nous avons dit aussi à propos de la notion d'énonciateur. Il s'agit de se garder de voir dans la question de la prise en charge énonciative d'un point de vue une subjectivité dont les attributs sont calqués sur ceux des sujets parlants.

² Notons que les règles mises en évidence dans la ScaPoLine permettent de fait aussi de préciser comment fonctionne la non-responsabilité.

qu'une quelconque instance où il s'origine, pour s'interroger alors sur l'instanciation de la source du point de vue au niveau du texte.

S'agissant pour nous d'analyser en quoi l'approche polyphonique peut servir l'analyse de la figure de l'auteur dans le genre de l'article de recherche, nous n'évoquerons pas davantage les fondements narratologiques de l'approche d'A. Rabatel. Il ne s'agit pas de gommer les différences entre le cadre qu'il propose et celui développé par O. Ducrot et dans la ScaPoLine. Au contraire, nous cherchons à saisir la spécificité de son approche, mais à partir de ses travaux les plus récents, qui montrent qu'une réflexion sur le point de vue initialement inscrite dans une perspective narratologique peut au final servir l'analyse de textes hors du champ littéraire. Ce faisant, nous retrouvons des éléments de réflexion sur le point de vue et sa prise en charge énonciative qui sont thématiques chez O. Ducrot ou dans la ScaPoLine, et qui suggèrent l'intérêt qu'il y aurait, à terme, de travailler à l'élaboration d'un cadre d'analyse sinon commun, du moins cohérent en dépit de la diversité des approches.

Plus explicite que dans les travaux de O. Ducrot, la centration sur le point de vue défendue par A. Rabatel reste nous semble-t-il en adéquation avec l'approche polyphonique de ce dernier : A. Rabatel insiste sur la distinction entre point de vue et discours en montrant « l'irréductibilité du point de vue, qui repose d'abord dans le fait qu'il n'a pas besoin de paroles de personnage pour s'exprimer » (1998 : 14). On retrouve donc bien l'idée que la polyphonie ne se réduit pas aux marques de la présence d'une autre « voix », car ces voix peuvent ne pas parler si l'on reprend les termes de O. Ducrot (1984). C'est aussi ce dont vise à rendre compte la ScaPoLine en montrant comme nous l'avons vu que les êtres-discursifs se présentent pour certains comme des images de locuteurs virtuels (actualisés ou non) ou comme des images de non-locuteurs.

La définition que construit A. Rabatel de la notion de point de vue dans son ouvrage de 1998 repose sur la notion de perception et sur le constat de la bivalence de ce phénomène : à l'image pourrait-on dire de ce que l'on nomme un point de vue en montagne, toute perception associe quelque chose de vu (un contenu, une vision) et le fait de voir, le « lieu » d'où l'on voit. Bien que A. Rabatel cherche à dépasser les métaphores qui embarrassent la définition du point de vue pour lui conférer une assise linguistique, la notion telle qu'il la problématise tend à renouer avec la motivation métaphorique du terme « point de vue », exclue dans l'approche de O. Ducrot¹ ; elle porte en germe en cela aussi la possibilité d'un rapprochement que nous avons évoquée *supra* entre le point de vue linguistique et le point de vue au sens commun d'opinion ou d'avis.

Ces deux aspects sont pour l'auteur « les deux facettes d'un même phénomène d'appréhension du réel », « car il n'existe pratiquement pas [...] de perception qui se limite à la saisie des objets sans rien nous dire du sujet percevant, au niveau de l'expression linguistique » (op.cité : 14).

Ces deux aspects ne nous paraissent pas complètement différents de la double composition du point de vue en termes de contenu et jugement analysée dans la ScaPoLine. Cependant, A. Rabatel ne parle pas lui-même de jugement, mais simplement de la subjectivité du point de vue (op.cité : 14 et 23), qui consiste en « l'expression conjointe de perceptions et de pensées représentées » (op.cité : 23). Autrement dit, le point de vue est envisagé comme une « saisie perceptive représentée, co-référent à une subjectivité déterminée ».

Plus spécifiquement, la définition du point de vue chez Rabatel associe trois critères :
- le pdv suppose non seulement l'existence d'un sujet percevant,

¹ Contrairement à l'approche de O. Ducrot (1984), elle semble en cela aussi porter en germe la possibilité d'une réflexion que nous avons évoquée *supra* sur le lien entre le point de vue linguistique et le point de vue au sens commun d'opinion ou d'avis.

- il suppose encore un procès de perception,
- et il suppose enfin et surtout une expression particulière de cette perception.

L'analyse du point de vue repose en effet chez Rabatel sur une distinction entre la prédication d'une perception et sa représentation, c'est-à-dire l'expression d'une perception représentée qui seule peut être assimilée à l'expression d'un réel point de vue (op. cité : 24) : la première étape consiste donc à montrer comment linguistiquement une perception devient perception représentée ou, en d'autres termes, à « s'attacher aux mécanismes de représentation des perceptions, puisqu'ils indiquent un PDV¹ » (op. cité : 55).

Les propositions d'A. Rabatel quant à cette première étape de l'analyse du point de vue apportent ainsi des éléments de réponse à la question de savoir où commence un point de vue, ou qu'est-ce qui peut être considéré comme point de vue, autrement dit à la délimitation de cette unité « minimale » pourrait-on dire de l'analyse polyphonique. L'auteur montre en effet la difficulté de distinguer prédication d'une perception et perception représentée, et conclut que

s'il est difficile de déterminer où commence le PDV (ce type de raisonnement présupposant des frontières nettes et intangibles, ce qui reste à prouver...), au moins peut-on tomber d'accord sur cette observation de portée limitée, mais d'un usage pratique et commode, selon laquelle plus l'aspectualisation du focalisé est développée (et [...] plus elle associe perception et pensées et comporte d'inférences cognitives à partir des perceptions), plus on est assuré d'être face à un point de vue. (op. cité : 29).

L'analyse de la référentialisation du focalisé, ou plus précisément de l'aspectualisation du focalisé est donc à la base du repérage du point de vue, en tant que première étape de l'analyse. Cela étant, A. Rabatel le souligne, « l'existence de perceptions représentées n'indique pas toujours explicitement qui en est le focalisateur », défini comme l'origine du point de vue ou plus exactement du procès de perception. C'est, en d'autres termes encore, la question du sujet percevant qui est alors en jeu : à partir de la référentialisation du focalisé, il s'agit de s'intéresser à l'attribution du point de vue, c'est-à-dire au repérage de ce que A. Rabatel appelle l'énonciateur, défini en tant que responsable des choix de référentialisation.

Pour mieux comprendre cette seconde étape de l'analyse, il faut spécifier encore la définition du point de vue fournie par A. Rabatel. Reprenant les propositions de J. Guillemin-Flescher (1984), il montre que la structure virtuelle du point de vue est faite de trois composantes, selon le schéma suivant : x (verbe de perception et/ou de procès mental) P, où ce qui est noté x renvoie au « terme repère qui coïncide avec l'origine de la perception », P au « terme repéré, qui, dans l'analyse de Guillemin-Felscher correspond « au contexte ou à l'élément perçu », et où le verbe de perception ou de procès mental sert de « relateur qui établit une localisation entre ces deux termes » (Rabatel, 1998 : 55). En d'autres termes, cette structure du point de vue « rapporte la prédication de perception à une source énonciative explicite, par le biais d'une expression explicite d'un procès de perception et/ou d'un procès mental » (ibid.).

Le qualificatif « explicite », usité à propos de la source énonciative et du verbe relateur appelle cependant un commentaire : explicites dans la structure virtuelle du point de vue, ces deux dimensions ne le sont pas toujours dans ses réalisations. On a vu ainsi avec l'exemple de la négation dans un énoncé comme « ce mur n'est pas blanc », que le pdv1 ne peut être décrit qu'avec l'élément noté P dans l'analyse de A. Rabatel, ou terme repéré : « ce mur est blanc », mais que le procès de perception ou procès mental, de même que l'élément X, terme repère, ou origine du point de vue, ne sont pas représentés dans l'énoncé.

¹ A. Rabatel utilise l'abréviation PDV en majuscule pour distinguer son approche de celle d'O. Ducrot (« pdv »). Pour éviter les ambiguïtés, nous ne reprenons cette notation que dans les citations que nous faisons de l'auteur, et parlons de « point de vue » en dehors de ces citations.

Poser cette structure doit donc permettre d'en envisager des réalisations distinctes, comme nous l'avons suggéré en discutant à partir de la ScaPoLine de la nécessité d'associer au point de vue la question de sa source. C'est ainsi que A. Rabatel justifie cette triple composition de la structure du pdv :

cette structure virtuelle initiale s'avère utile dans la mesure où elle rappelle à l'interprétant que les trois éléments de la structure peuvent entrer dans des réalisations très diverses, qu'il faut pourtant rapporter à cette structure théorique, quand bien même, au niveau de la manifestation, elle n'est pas complète. C'est donc à titre heuristique que cette structure est importante, car c'est à partir d'elle qu'on peut imaginer des inférences susceptibles de combler les absences (relativement fréquentes) de l'un des deux premiers éléments de la structure [...]. C'est encore à partir de cette structure abstraite que l'on peut rendre compte des diverses modalités d'expression du sujet focalisateur, ou du procès de perception. (1998 : 56)

L'analyse de ces modalités d'expression est cruciale dans la perspective d'A. Rabatel. Le repère et le relateur sont considérés comme essentiels pour l'attribution du PDV à une origine déterminée, le focalisateur, et il s'agit donc pour lui de décrire « les principales marques linguistiques embrayant sur un PDV du personnage ou du narrateur (les deux seuls candidats focalisateurs) », pour mieux cerner comment se construit, textuellement, la sphère du narrateur et celle du personnage.

Ce faisant, A. Rabatel fournit une précision intéressante sur l'expression du point de vue et de sa source : « considérer que le repérage du foyer se limite au repérage du « sujet (sémantique) des perceptions et des pensées » ne doit pas masquer la difficulté, tant les structures syntaxiques exprimant ce rapport sémantique sont variées » (1998 : 58).

Il convient donc comme il le souligne de « distinguer le sujet sémantique des perceptions (l'agent) et le sujet syntaxique desdites perceptions » : « le sujet sémantique des perceptions peut endosser de multiples habits syntaxiques et il y a de notables différences entre un sujet percevant actif [« Pierre scrutait attentivement la scène qui se déroulait à quelques pas de lui] et un sujet percevant plus passif en raison du sémantisme du verbe » (A. Rabatel en donne pour exemple « Pierre entendait vaguement des coups », où « le sujet est agi par sa propre perception », ou des structures syntaxiques telle que la passivation, « des coups étaient perçus par Pierre ») (1998 : 55-57).

Cette distinction entre ce que l'on pourrait appeler trop schématiquement la prise en charge syntaxique et sémantique de la perception représentée permet à A. Rabatel (1998 : 58) de spécifier encore sa définition linguistique du point de vue :

les bases linguistiques de l'expression du point de vue reposent dans l'expression des perceptions et/ou des pensées représentées. Ces perceptions et pensées représentées sont sous la dépendance syntaxique d'un sujet et d'un procès de perception mentionnés dans les premiers plans¹ et/ou sous la dépendance sémantique d'un agent ou d'un procès que le texte ne mentionne pas explicitement et que le lecteur reconstruit par inférence.

Une fois cet objectif mis en œuvre de « rechercher les traces linguistiques d'un point de vue dans le mode de donation du référent de l'objet perçu » (ibid.), A. Rabatel montre qu'il n'est pas toujours aisé de distinguer les points de vue du narrateur et du personnage : sa description met en évidence le fait que ces points de vue sont « moins homogènes qu'on pourrait le croire », et qu'ils se caractérisent linguistiquement par leur intrication (1998 : 139).

S'agissant d'analyser des régularités quant à ces relations entre les PDV du personnage et du narrateur, il introduit d'une part les notions de volume de savoir et de profondeur de perspective, d'autre part les notions de consonance et de dissonance entre les

¹ I.e. la suite chronologique dans la perspective d'étude des textes narratifs adoptée par A. Rabatel.

points de vue. Sans reprendre ici l'ensemble de la démonstration, on se contentera de souligner les principales conclusions auxquelles elle aboutit.

A propos du volume du savoir narratif et de profondeur de perspective, A. Rabatel (op. cité : 165) montre que l'éventail des possibles est un peu plus ouvert au narrateur qu'au personnage, quoique l'analyse permette également de spécifier que le point de vue du personnage est plus étendu qu'on le dit traditionnellement. A propos de la consonance et de la dissonance des points de vue, l'analyse suggère que le système du PDV « malgré ses régularités est éminemment souple » et que les traces polyphoniques compliquent les choses, surtout si l'on cherche avec minutie à attribuer tel ou tel énoncé au PDV [du personnage ou à la subjectivité du narrateur] : même quand un PDV est clairement le fait du personnage, la voix du narrateur peut être toujours présente et plus ou moins prégnante.

Ainsi, on retiendra qu'il est difficile (ou parfois arbitraire) d'attribuer avec certitude à telle ou telle portion textuelle l'expression d'un PDV déterminé « et pour cause d'ailleurs puisque le propre des écritures consonnantes est d'intriquer ce que l'analyse tente pataudement et comme vainement de séparer » souligne A. Rabatel (op. cité : 183). Au contraire, l'ambition est donc de rendre compte de cette complexité inscrite linguistiquement dans les marques polyphoniques, par laquelle ce sont les relations entre points de vue et leur possible intrication qui caractérisent un texte, selon une polyphonie irréductible. Nous trouvons ici un écho à la discussion menée à propos des êtres-discursifs que la ScaPoLine cherche à distinguer, car quoique ces distinctions soient utiles à l'analyse, c'est nécessairement aussi leur intrication qu'il s'agit de mieux cerner.

Relations entre points de vue et construction de L1/E1 dans la topique énonciative

L'analyse des relations entre points de vue semble être le fil conducteur des projets que poursuit l'auteur, et c'est à ce titre qu'ils intéressent notre approche. A cette première caractéristique s'en ajoute une seconde par rapport à ses deux premiers ouvrages sur le point de vue : si A. Rabatel s'intéresse toujours au texte littéraire, son cadre d'analyse des relations entre points de vue est également appliqué à d'autres textes (pour n'en citer que deux exemples, les productions d'apprenants, 2002, le dictionnaire philosophique de Comte-Sponville, 2004b).

Ces outils pour l'analyse des relations entre points de vue sont décrits au moyen d'une « topique énonciative » (Rabatel, 2004aa). Celle-ci prend comme point de départ la disjonction entre le locuteur et énonciateur ; la définition donnée de l'énonciateur est en adéquation avec celle proposée par O. Ducrot. Cependant, si celui-ci définit solidairement l'énonciateur et le point de vue, A. Rabatel (op.cité : 3) souligne à juste titre que

derrière l'implication réciproque des termes (pas de pdv¹ sans énonciateur, pas d'énonciateur sans pdv), les deux concepts ne fonctionnent pas au même niveau, la notion de pdv servant à définir l'énonciateur (disjoint du locuteur) comme la source d'un contenu propositionnel exprimant un pdv – ce dernier ne se manifestant pas que dans le *modus*, mais aussi dans le dictum (O. Ducrot 1993 : 128).

La disjonction avec l'instance du locuteur se justifie par « le fait que, si tout locuteur est aussi énonciateur (en vertu du principe de sincérité), en revanche, tout énonciateur n'est pas nécessairement locuteur » (2004a : 2), au moyen d'exemples proches de ceux que l'on a examinés pour introduire la question de la polyphonie, et d'autres (le discours indirect libre, les énoncés ironiques, doxiques ou des points de vue narratifs).

¹ A. Rabatel utilise ici la notation en minuscules (pdv), contrairement au choix fait dans ses écrits antérieurs (PDV).

Cependant la définition de locuteur semble maintenir une ambiguïté, sur la question de sa distinction ou de son assimilation au sujet parlant : A. Rabatel (2004a) semble bien faire du locuteur une instance de l'énoncé, mais les propriétés de ce dernier font qu'il revêt parfois les attributs du sujet parlant, comme le montre sa définition *infra*. Il pourrait cependant s'agir d'une simple difficulté à systématiquement préciser la distinction comme on l'a vu avec les abus de langage évoqués par O. Ducrot, car en d'autres lieux, A. Rabatel revendique la distinction et souligne en référence à J. Authier-Revuz la nécessité d'échapper à une conception idéaliste du locuteur conçu comme « sa majesté le sujet » (2004 : 7). Cela étant, on soulignera que contrairement à O. Ducrot et à la ScaPoLine, la réflexion sur le lien entre les points de vue assumés dans l'énoncé et leur prise en charge par un sujet parlant producteur effectif de l'énonciation n'est pas écartée chez A. Rabatel.

Ce point est particulièrement important dans le cadre de notre approche : en parlant de figure de l'auteur, nous introduisons, à titre programmatique, cette question du lien entre la représentation que donne le texte de son auteur et la question des propriétés de l'énonciation effective de ce texte. Cette question, plus nette encore avec une notion comme l' « image de soi du scripteur », demande comme nous l'avons dit une certaine prudence, pour ne pas assimiler les propriétés de l'énonciation représentée à celles de l'énonciation effective, mais elle intéresse, à terme, l'idée d'une analyse de la construction, dans les textes, des points de vue en tant qu'opinions assumées par des acteurs du champ social.

Pour revenir à l'approche de A. Rabatel, le locuteur est ainsi défini comme instance qui « profère un énoncé embrayé ou désembrayé, dans ses dimensions phonétiques et phatiques ou scripturales, » tandis que l'énonciateur désigne l'instance des actualisations opérées par le sujet modal, qui « assume l'énoncé en un sens nettement moins abstrait que la prise en charge découlant de l'ancrage déictique » (2004 : 3).

Si l'on considère que le locuteur est à comprendre comme une dimension du texte, on conviendra donc de la compatibilité du cadre d'analyse proposé par A. Rabatel avec celui de la ScaPoLine : le locuteur est une instance qui se présente comme responsable d'un acte d'énonciation, tandis que l'énonciateur pose la question de la responsabilité du point de vue. Cependant, alors que la différence tiendrait à des niveaux d'analyse dans la ScaPoLine, entre point de vue et énoncé, et au fait qu'il n'y a d'énonciateur que dans le sens où cette variable peut être saturée par des êtres discursifs, la différence entre locuteur et énonciateur chez A. Rabatel rejoint ce que A. Culioli (1990, 1999) traite en termes d'opérations déictiques et modales, dans la mesure où l'on aurait dans le texte un repère des opérations déictiques (le locuteur) et un repère des opérations modales (les énonciateurs).

Par ailleurs, comme il s'agit de même que dans la ScaPoLine d'analyser le jeu polyphonique au niveau du texte, A. Rabatel souligne que, à la différence des propositions de O. Ducrot, « plutôt que de considérer qu'il y a autant d'énonciateurs que de contenus propositionnels, il paraît cognitivement moins coûteux de regrouper ceux qui peuvent avoir des contenus thématiques divers, mais que rassemblent la même origine et la même visée énonciative, surtout si les énonciateurs sont incarnés¹ » (2004c, cité dans 2004a : 3).

Se manifeste ainsi la proximité de cette question du regroupement avec l'approche de la ScaPoLine : A. Rabatel souligne qu'elle implique de tenir compte à l'instar de la ScaPoLine de la saturation sémantique des énonciateurs par les êtres-discursifs, en fonction aussi de ce qu'il appelle leur degré d'incarnation. Celui-ci correspond vraisemblablement au fait que les êtres-discursifs peuvent dans la perspective scandinave se présenter comme des non-locuteurs, ou comme des locuteurs virtuels dont les aptitudes de locuteur sont ou non actualisées – dans la manière dont elles se présentent dans le texte.

¹ L'auteur renvoie dans une note, à propos de ce terme « incarné », à la question de la saturation de la variable « énonciateurs » par des êtres-discursifs dans la ScaPoLine.

Par ailleurs, cette question du regroupement des instances énonciatives est abordée sous l'angle double de la hiérarchisation des points de vue et des énonciateurs d'une part, et des relations entre énonciateurs et locuteurs d'autre part. L'attention est accordée prioritairement à l'instance L1/E1, ou plus justement, comme le point de départ reste le point de vue et son énonciateur, à « l'énonciateur en syncrétisme avec le locuteur en tant que tel et en tant qu'être du monde (voire avec le sujet parlant) » (2004a).

Cette notion de syncrétisme est une manière de rendre compte de la possibilité pour un point de vue de se voir attribuer comme source le locuteur (au sens de LOC), par l'intermédiaire de son énonciateur. C'est vrai pour L1/E1 qui renvoie au locuteur premier, mais également pour d'autres locuteurs se manifestant dans le texte, dans le cas où la source énonciative d'un point de vue est « incarnée », i.e. est un locuteur, comme dans le discours rapporté.

On voit donc que l'analyse reste proche du projet de la ScaPoLine, quoiqu'elle porte ici préférentiellement sur le locuteur *via* les différents énonciateurs, plutôt que sur les liens permettant la saturation des énonciateurs par les êtres discursifs (impliquant eux-mêmes des images du LOC mais aussi de l'allocutaire ou encore les troisièmes). L'objet est cependant quelque peu différent, puisqu'il s'agit de rendre compte de la manière dont se construit ce L1/E1 à partir du jeu énonciatif.

D'une part, l'objectif, dans la perspective d'A. Rabatel, est semble-t-il de mieux définir cette notion de syncrétisme, pour voir dans quelle mesure un énonciateur peut être conçu comme étant en syncrétisme avec le locuteur, et pour voir dans quelle mesure des regroupements sont possibles, ou non, en fonction des caractéristiques des textes. Ainsi le projet d'une étude de la figure de l'auteur dans les textes à visée didactique (Grossmann et Rabatel, eds., à paraître 2007) entend-il montrer comment la construction d'une figure de l'auteur y est mise en péril par une hétérogénéité extrême des instances énonciatives et l'absence de point de vue englobant.

D'autre part, c'est la hiérarchisation des instances énonciatives qui est analysée pour rendre compte de la manière dont se construit (possiblement) une instance telle qu'elle puisse être notée L1/E1 : la hiérarchisation des points de vue et partant, des instances à leur source est à la base de la possibilité d'envisager au niveau du texte une instance du type du locuteur au sens ducrotien (LOC) qui s'y présente comme son responsable, ou plus précisément dans la perspective de A. Rabatel, qui se présente à la fois comme repère des opérations déictiques (et responsable de l'acte d'énonciation *hic* et *nunc*, dans ses dimensions phonétiques et phatiques ou scripturales, le locuteur) et comme repère des opérations modales (et responsable des points de vue pris en charge dans l'énoncé *via* les énonciateurs).

Cette hiérarchisation est analysée par A. Rabatel sous l'angle des trois possibilités de relations entre points de vue et entre énonciateurs : la coénonciation, la surénonciation et la sousénonciation (2004a). Il s'agit de traiter des interactions et inégalités de points de vue en termes de postures énonciatives.

Reprenons rapidement les définitions qu'en donne A. Rabatel. La coénonciation, manifestement rare tant le jeu des points de vue fait état d'effets d'inégalités, correspond aux cas où deux voix se superposent, momentanément, autrement dit à la « co-construction par les énonciateurs d'un point de vue qui se donne comme point de vue commun et partagé ». La surénonciation caractérise un point de vue interactionnellement dominant, surplombant, d'un énonciateur sur l'autre. Enfin la sousénonciation caractérise les cas où un point de vue se présente dans le texte comme interactionnellement dominé.

L'originalité du cadre d'analyse – ou de la topique ainsi constituée – est de chercher à rendre compte du jeu des points de vue dans des textes extrêmement divers, y compris dans

une textualité orale du type de la conversation. C'est ce qui explique sans doute les ambiguïtés que peuvent soulever ces notions si l'on occulte la disjonction première entre énonciateur (instance modale attachée au point de vue) et locuteur (responsable de la production de l'énoncé). Ainsi A. Rabatel s'efforce-t-il de bien préciser que c'est le jeu interactionnel des points de vue qui est visé par l'analyse, et non le jeu interactionnel des acteurs empiriques de l'échange¹.

Encore une fois, ce sont donc des caractéristiques textuelles qui sont décrites, sous l'angle de l'interaction des points de vue que rend tangible tout texte, qu'il soit le fait d'un seul producteur effectif (contexte monologal dans les termes de A. Rabatel) ou de deux ou davantage d'interlocuteurs co-présents (contexte dialogal).

La seconde difficulté que comportent ces notions est la question du marquage de ces postures de co-, sur- et sousénonciation. Notons que le problème n'est pas spécifique à ces notions mais se pose à toute approche énonciative en linguistique. Les postures énonciatives proposées par A. Rabatel (2004a) sont en effet codées dans les textes. Cependant, les marques sont innombrables, (elles incluent dans le cas de l'oral la gestualité), et peuvent être polysémiques, de même que *je* ne renvoie pas seulement ou pas toujours au locuteur, et que ce locuteur ne se manifeste pas que dans ce *je*. Par ailleurs, la difficulté s'accroît dans l'analyse du texte, car ces postures peuvent se relayer les unes les autres, si bien qu'une sousénonciation peut par exemple n'être que provisoire et servir finalement la surénonciation.

Ces notions introduites par A. Rabatel permettent d'aborder la polyphonie non plus seulement en termes de multiplicité mais aussi de hiérarchisation des points de vue, comme c'est aussi le cas dans l'étude polyphonique proposée par la ScaPoLine, ou dans celle des différentes formes de dialogisme que l'on trouve chez J. Bres et B. Vérine (1999, 2002). Comme nous l'avons dit, c'est d'ailleurs en ce sens que ces derniers préfèrent parler de dialogisme plutôt que de polyphonie, pour mettre l'accent sur cette hiérarchisation énonciative.

Si les faits textuels que concerne la topique énonciative d'A. Rabatel sont déjà décrits dans ces approches et dans d'autres, leur intérêt tient à ce que, comme pour la ScaPoLine, il s'agit de rendre compte de la construction d'une posture L1/E1- du moins dans certains genres qui appellent la construction d'une telle posture devrait-on préciser, en ajoutant que l'objet permettra de préciser les difficultés de sa construction dans d'autres.

Par ailleurs, la topique ainsi décrite fournit une assise linguistique et des outils d'analyse de la construction de la « visée argumentative » au sens entendu par R. Amossy (2000 : 24) de projet de modifier les représentations d'un lecteur présumé, qu'il s'agisse comme le souligne A. Rabatel (2004a)

des effets argumentatifs résultant de l'argumentation dans la langue, au sens ducrotien, ou des effets argumentatifs indirects reposant sur des schématisations (Grize 1990), sur des topoï qui orientent l'interprétation indépendamment de la présence des formes argumentatives « logiques » (Amossy 2000)

Les travaux d'A. Rabatel doivent ainsi permettre de mieux rendre compte du lien entre énonciation et argumentation, en fournissant une assise énonciative à la construction du point de vue dans le texte et de sa légitimité. Ils permettent ainsi d'envisager la visée argumentative du texte en tant que telle, autrement dit dans la matérialité des échanges entre sujets parlants.

¹ Même si la question des relations entre les deux n'est pas exclue par A. Rabatel, cf. 2004d.

2.3.5. Quel cadre pour l'analyse de la polyphonie ?

Les approches de la polyphonie développées dans la ScaPoLine d'une part et chez A. Rabatel d'autre part ont en commun ce fait inédit de tenter de rendre intelligible l'analyse de la construction du point de vue au niveau textuel. Si elles partagent aussi le même intérêt pour l'analyse du texte littéraire, le cadre que chacune d'elles propose fournit des outils pour toute analyse énonciative de texte. Leur examen montre cependant la difficulté qu'il y a à envisager un cadre à proprement parler pour l'analyse de la polyphonie, mais permet également d'en pointer les principaux fondements.

Les fondements de l'analyse

De manière centrale, il faut retenir qu'en dépit des différences nettes entre les approches de la ScaPoLine et d'A. Rabatel, le parti pris pour une analyse polyphonique des textes consiste à prendre comme point de départ les traces linguistiques du point de vue pour rendre compte de l'interprétation de la polyphonie.

Celle-ci manifeste toute sa complexité : d'une part il s'agit, et c'est vrai des deux approches, de rendre compte de la multiplicité des voix et des points de vue, et de montrer comment cette multiplicité se résout, en partie, par des possibilités de regroupements, notamment par la construction polyphonique d'une instance de locuteur dans le texte. Cette instance cependant se présente comme irréductiblement hétérogène, ne serait-ce que parce que les modes syntaxiques de prise en charge énonciative sont variés comme le souligne A. Rabatel, et que, sémantiquement, les images du locuteur sont multiples.

D'autre part, et cette seconde question n'est pas étrangère à la première, il s'agit de rendre compte de la hiérarchisation des points de vue ; la notion de points de vue relationnels dans la ScaPoLine montre comment un point de vue se construit par rapport à d'autres, et c'est aussi l'objet de la topique énonciative d'A. Rabatel, qui vise à décrire cette hiérarchisation par laquelle un point de vue, mais aussi, à partir de lui, le locuteur au niveau du texte, acquièrent dans le jeu polyphonique une posture dominante, ou dominée, ou momentanément superposée à un autre point de vue ou une autre instance.

Les éléments de réflexion et les outils d'analyse proposés dans ces deux approches sur le point de vue et sa prise en charge nous permettent ainsi d'envisager des pistes pour une étude de la polyphonie au niveau textuel, en vue de caractériser énonciativement le genre de l'article de recherche avec la notion de figure de l'auteur.

Trois éléments sont centraux dans l'analyse, quelle que soit la théorie considérée :

- le point de vue,
- la question de sa prise en charge, intégrée à la structure même du point de vue,
- la question de la prise en charge de l'énoncé dans son ensemble (le texte, dans notre optique).

Pour mieux saisir les différences terminologiques et conceptuelles entre les approches, il faut ajouter à cela le sujet parlant, ou producteur effectif de l'acte d'énonciation, quoiqu'il soit bien entendu systématiquement exclu de l'analyse linguistique.

Récapitulons schématiquement la manière dont O. Ducrot d'une part, en tant que fondateur de ces analyses polyphoniques, la ScaPoLine d'autre part et A. Rabatel entendent rendre compte de ces différentes dimensions.

	Le sujet parlant, producteur effectif de l'énoncé (ou responsable effectif de l'acte d'énonciation)	L'instance qui dans l'énoncé se présente comme responsable de l'acte d'énonciation dont il résulte	La source ou le responsable du point de vue	Le point de vue
O. Ducrot	Sujet parlant	Locuteur	Énonciateur	Point de vue (parfois utilisé pour énonciateur et inversement)
« ScaPoLine »	Sujet parlant	Locuteur	Une variable « énonciateur » saturée par des êtres-discursifs en vertu de liens énonciatifs	Point de vue : une composante référentielle et une composante modale
A. Rabatel	Sujet parlant (<i>ou locuteur dans certains cas</i>)	Locuteur ou énonciateur « principal » (L1/E1) ou figure de l'auteur	L'énonciateur (déterminé par le repère et les relateurs contenus dans la structure du pdv)	Point de vue : repère (origine de la perception), terme repéré et relateur (verbe de perception).

Figure 2.1. L'analyse de la polyphonie linguistique chez O. Ducrot, dans la ScaPoLine et chez A. Rabatel.

En accord avec la théorie scandinave de la polyphonie linguistique et avec l'approche de A. Rabatel, nous retiendrons comme premier principe l'importance de prendre comme point de départ le point de vue. Cependant, on a envisagé aussi comment dans ces deux cadres d'analyse l'enjeu ne consiste pas seulement à faire état de la multiplicité des points de vue, mais à rendre compte de l'interprétation polyphonique au niveau du texte. Dans cette perspective, c'est d'abord la question de la prise en charge du point de vue qui importe.

Le point de vue et sa prise en charge ou la question de la modalisation

Cette question de la prise en charge du point de vue se pose d'abord au niveau du point de vue lui-même, comme nous l'avons vu en discutant de l'analyse du point de vue dans la ScaPoLine et chez A. Rabatel. Nous conviendrons à partir de ces deux approches de définir le point de vue dans sa double composante référentielle et modale : le point de vue est à la fois quelque chose de dit et l'expression d'un rapport à ce qui est dit. Il peut être montré en tant que point de vue, dans les cas où, comme le souligne A. Rabatel, le procès de perception et/ou la source énonciative du point de vue sont marqués dans le texte. Cela rejoint ce que nous avons dit pour introduire la problématique polyphonique, à propos de « Paul viendrait demain », et d'une analyse du pdv2 en ces termes : « Paul viendra demain (pdv1) est une

assertion envisageable » : le conditionnel introduit un dédoublement des points de vue par lequel le pdv 1 est montré en tant que tel¹.

Tout en maintenant donc cette double composante, il nous semble important d'éviter de parler de jugement pour rendre compte du « rapport au dit », i.e. de la seconde composante du point de vue, impliquée par la composante référentielle. Ce « rapport à » suggère, moins qu'une subjectivité, cette possibilité qu'à tout point de vue d'être signalé en tant que tel, comme le fait d'un procès de perception et/ou d'une source énonciative.

L'idée que tout point de vue comporte une dimension modale revient à dire que tout point de vue peut virtuellement être asserté, au sens où tout point de vue peut être montré comme énoncé ou pensé ou encore énonçable ou pensable. Les contenus propositionnels, ou la dimension référentielle du point de vue, ne sont pas à envisager en termes de reflet d'une réalité pré-existante. Comme ils contribuent à informer cette réalité, i.e. à lui donner forme, le seul fait de nommer, ou, au niveau du texte, la seule dénomination peut être considérée comme un point de vue ; c'est ce que montre aussi le fait que toute dénomination puisse faire l'objet d'un dédoublement où le dit se trouve commenté en tant que tel, comme l'a analysé J. Authier-Revuz (1995).

Le point de vue trouve sa matérialité dans le texte (cf. le point de vue « ce mur est blanc » impliqué par la négation dans « ce mur n'est pas blanc »), de même que cette possibilité qu'à le point de vue d'être montré en tant que tel : dans « Paul viendrait demain », le conditionnel implique le point de vue « Paul viendra demain » et montre ce point de vue en tant que tel.

Dans cette perspective, un point de vue monophonique dont la dimension modale serait linguistiquement non marquée (par exemple dans « il fait beau ») est à comprendre comme une forme parmi d'autres sous lesquelles se présente un point de vue. Dire que sa dimension modale n'est pas marquée ne signifie pas qu'elle est inexistante : c'est bien une forme de prise en charge modale que cette absence de marque modale. En d'autres termes, le seul contenu référentiel dit quelque chose de l'instance responsable de ce contenu, y compris quand cette instance n'est aucunement marquée par des marques dites modales, ce qui renforce l'idée d'éviter d'envisager la composante modale sous l'angle des seules marques de personne et/ ou en termes de subjectivité².

A partir de là, il convient de contourner le problème que poserait une distinction entre des liens énonciatifs externes au point de vue et des jugements de probabilité, de vérité, etc., internes au point de vue : si le lien énonciatif relie la composante jugement et les êtres discursifs, le point de vue et sa prise en charge ne peuvent être traités en termes d'internalité/d'externalité car les êtres discursifs doivent pouvoir se manifester à l'intérieur des points de vue, de même que les instructions quant aux liens énonciatifs.

La question appelle de nouveau à poser celle de la délimitation des unités « points de vue » dans les textes, et nous avons dit qu'elle était particulièrement délicate. Nous avons

¹ En ce sens, nous ne souscrivons pas complètement à l'analyse proposée par K. Fløttum (2002b) à propos des passages entre guillemets avec l'exemple « Jean s'est acheté un "château" » : l'auteur montre qu'on peut y voir deux pdv : « Jean s'est acheté un "château" » (pdv1) et « "château" n'est pas le bon terme » (pdv2). A notre sens, le pdv2 tel qu'il se présente linguistiquement n'implique pas nécessairement une évaluation axiologique quant au choix du terme, mais simplement le fait que pdv1 est une assertion envisageable ; en d'autres termes, pdv2 dit qu'il est possible de parler de château, sans comporter de précisions sur la validité ou l'invalidité d'un tel choix dénomiatif. L'usage des guillemets repose bien cependant sur un dédoublement énonciatif, et correspond ici à deux fonctions de dénomination et de modalisation que nous retrouverons en étudiant les passages entre guillemets dans l'article de recherche au chapitre 5.

² Sauf à s'interroger ce que l'on entend par subjectivité. On comprend l'intérêt mutuel de certaines approches énonciatives (cf. en particulier les travaux de J. Authier-Revuz) et des approches psychanalytiques sur la question du sujet.

également discuté à partir des deux composantes du point de vue dans la ScaPoLine, en regard des notions de dictum et de modus chez C. Bally (1909) la difficulté de distinguer dans la matérialité des textes les deux composantes. S'il est opératoire de bien distinguer le point de vue de sa prise en charge, il reste que la question de la prise en charge d'un point de vue détermine également celle de l'évaluation d'un point de vue, comme on le voit avec la négation (le LOC s'exclut de la prise en charge du point de vue invalidé par la négation).

Si l'on convient donc que le point de vue comporte virtuellement la possibilité de se présenter comme le fait d'une perception (ou d'une assertion) et d'un sujet de perception, il importe d'envisager sa composante modale dans toute sa complexité. D'une part, elle pose le problème de ce que l'on a usage d'appeler le mode de prise en charge énonciative : le point de vue est-il énoncé au moyen d'un tour personnel, ou d'un tour délocuté ? D'autre part, elle mobilise l'ensemble des marques modales, autrement dit l'ensemble des formes linguistiques qui peuvent co-référer à cette instance impliquée par le « rapport au dit » que contient tout point de vue. En l'absence de marques, soit ce « rapport au dit » reste virtuel (comme c'est le cas du point de vue invalidé dans « ce mur n'est pas blanc), soit il ressortit à une instance inférable (le locuteur dans « il fait beau », un énonciateur dit collectif, etc.), autrement dit à ce que la ScaPoLine traite en termes d'êtres-discursifs, mais qu'au seul niveau du point de vue nous pouvons traiter en termes d'énonciateur.

En raison des différents problèmes posés par cette notion d'énonciateur, nous préférons dans la mesure du possible parler de source du point de vue, en soulignant qu'il s'agit d'envisager à la fois la question de sa prise en charge énonciative (résolue par la notion d'énonciateur ou par le système des liens énonciatifs et des êtres-discursifs dans la ScaPoLine), et la question du degré d'assertion et de vérité du point de vue : les jugements de probabilité ou de vérité évoqués par P. Dendale (2005) peuvent être envisagés comme référant à cette source modale qu'est l'énonciateur. Ils participent de la prise en charge énonciative du point de vue, au même titre à notre sens que le fait qu'un point de vue soit énoncé sur un mode personnel ou non. De même, l'explicitation d'un procès de perception dont parle A. Rabatel (1998) informe la composante modale du point de vue en termes de source du point de vue et de degré d'assertion et de vérité du point de vue.

Plutôt que de voir dans cette composante modale un sujet, le fait de l'envisager en termes de source du point de vue et, indissociablement, de degré d'assertion et de vérité de ce point de vue doit permettre d'en envisager la complexité. Comme le souligne A. Rabatel, cette source se construit au moyen de marques éminemment variées. Aussi est-il peut-être ambitieux de prétendre caractériser sémantiquement cette source en termes, par exemple, d'énonciateur collectif, ou encore de locuteur textuel et de locuteur de l'énoncé comme on l'a discuté à propos de ces notions dans la ScaPoLine : l'intérêt pourrait être non seulement de voir la complexité de cette construction de la source, et éventuellement son irréductible polyphonie en montrant comment des sources différentes peuvent se trouver intriquées.

Nous retiendrons cependant que la composante modale à la source du point de vue peut se présenter comme une source locutoire : sur ce point la notion de locuteur virtuel introduite dans la polyphonie nous paraît-elle particulièrement intéressante, pour rendre compte du fait que dans certains tours polyphoniques, non seulement un autre point de vue est impliqué, mais un point de vue montré comme tel comme on l'a dit *supra*, et que certains de ces points de vue montrés comme tels peuvent être montrés plus spécifiquement comme des assertions (énoncées ou énonçables). Ainsi, pour reprendre rapidement l'exemple qui nous a servi à préciser cette notion, si l'on associe à « Paul viendrait demain » une spécification du type « selon Marie », le locuteur virtuel impliqué par le conditionnel se trouverait actualisé par la désignation « Marie ».

Il nous reviendra donc de voir à partir de l'analyse de l'article de recherche comment cette source du point de vue se constitue dans les textes, et en particulier de mieux cerner des degrés d'actualisation de cette source en tant que source locutoire. Sans doute existe-t-il en effet, sinon des catégories du type « non locuteur », « locuteur virtuel », « vrai locuteur », du moins des degrés par lesquels un point de vue peut être simplement impliqué par une marque polyphonique comme dans « ce mur n'est pas blanc » sans que rien ne soit dit quant à sa source, ou peut se présenter comme point de vue (Paul viendra demain dans Paul viendrait demain) ou encore comme point de vue attesté ou attestable, i.e. étant ou pouvant être effectivement pris en charge par une source locutoire ou perceptive (du type « Paul viendra demain », « on peut penser que Paul viendra demain », « selon Marie Paul viendra demain »).

Ce problème se complique à deux niveaux : d'abord, ces degrés d'actualisation dépendent de la possibilité qu'à tout point de vue de se voir associé à son procès de perception et/ou sa source, et nous avons vu sur ce point avec A. Rabatel la diversité des modes syntaxiques de prise en charge et des valeurs sémantiques à associer. Il faudrait donc sans doute introduire différents critères pour tenter de caractériser une source, plutôt que de vouloir la nommer en tant, par exemple, qu'énonciateur collectif, source locutoire, etc. : certaines sources se laissent sans doute difficilement appréhender en ces termes, de par leur hétérogénéité observable dans les textes. C'est par exemple le problème du *on* en français, face auquel on convient d'analyser les différentes valeurs prises à l'aide de différents critères (notamment, inclusif ou non, générique ou non, défini ou indéfini, cf. D. Maingueneau, 1994).

Par ailleurs, ce problème de la source reste complexe car comme le souligne la ScaPoLine, une part de la saturation de la source peut se jouer dans l'interprétation. Affirmer que « ce mur n'est pas blanc » n'est pas en soi une opposition mais peut le devenir, dès lors que cet énoncé répond à un autre, dans l'échange immédiat mais aussi dans l'histoire conversationnelle des locuteurs, et dans le texte ou dans sa production ou son interprétation.

La polyphonie comme dédoublement des points de vue

Nous retiendrons comme définition de la polyphonie celle de dédoublement des points de vue, tout en admettant que la prolifération des points de vue demande à être associée à la question de leur prise en charge et à être traitée dans certains cas en termes de prolifération des voix, lorsque la multiplication des points de vue est telle que leur dénombrement même est mis en péril.

Tout point de vue peut faire l'objet d'un commentaire – de manière programmatique, nous pourrions dire que c'est peut-être ce qui caractérise un point de vue, et qui explique que, dans la ScaPoLine, l'enchaînement des énoncés soit pris comme test pour discriminer les points de vue. On a vu que la seule dénomination pourrait être considérée comme un point de vue : dans « ce mur est blanc », on peut envisager que « ce mur » (« ceci est un mur ») est un point de vue, qui peut faire l'objet d'un dédoublement et/ou être montré en tant que tel. La démonstration mérite cependant de prendre pour objet le texte, dont l'enjeu ne se satisfait vraisemblablement pas d'un recensement de l'ensemble des points de vue.

Cependant, si l'on retient au moins provisoirement cette idée, on pourrait aborder l'analyse des points de vue et de la polyphonie à l'aide de deux questions : celle de savoir si le point de vue est montré en tant que tel d'une part (au moyen d'un procès de perception et/ ou d'une source énonciative) ou non, et si le point de vue fait l'objet d'un dédoublement polyphonique, par lequel il se trouve commenté.

Nous entendons « commentaire » dans le sens où J. Authier-Revuz (1995) l'utilise pour rendre compte des formes de dédoublements méta-énonciatifs : comme elle le souligne,

le dit peut faire l'objet d'une boucle réflexive par laquelle il se trouve commenté en tant que dit, comme dans « c'est ce qu'on appelle des marginaux » (op. cité : 39). Ce procédé central en langue introduit de l'opacité : comme le souligne J. Authier-Revuz, le signe tel qu'il est envisagé ordinairement, sur le mode de la transparence, s'efface devant la chose nommée, tandis que le signe autonyme, montré comme signe,

résulte de ou consiste en – selon que l'on parle du résultat ou du processus énonciatif – une référence au monde accomplie en interposant sur le trajet de la nomination la considération de l'objet signe par lequel on réfère (op. cité : 31)

J. Authier-Revuz propose alors une analyse des formes de dédoublement énonciatif de cette modalité autonymique opacifiante ; la diversité des formes que prennent les boucles réflexives fait apparaître que le commentaire n'est pas toujours à envisager sur le mode d'une explicitation maximale : il peut se limiter à introduire de l'opacité sans rien dire du signe commenté sinon qu'il s'agit de l'envisager comme signe. Ainsi, les guillemets bloquent l'interprétation littérale du segment qu'ils entourent, mais ils ne fournissent en eux-mêmes d'autre instruction que cette rupture.

Si les phénomènes analysés par J. Authier-Revuz peuvent être analysés comme des formes de polyphonie, nous soulignerons que son analyse en termes de dédoublement des points de vue semble pouvoir être appliquée à un ensemble plus vaste encore de phénomènes. C'est ce que proposent J. Bres et B. Vérine (2002), qui entendent décrire le phénomène du dialogisme dans son unité, en repérant ce qui unit les divers tours analysés comme polyphoniques dans les approches énonciatives.

Préférant le terme de dialogisme à celui de polyphonie pour mieux rendre compte de la nécessaire hiérarchisation des voix qui se joue dans leur multiplication, ces auteurs soulignent que le dialogisme repose sur un dédoublement énonciatif qui peut dans certains cas s'afficher comme dédoublement des locuteurs. Si nous entendons comme nous l'avons montré nous en tenir aux points de vue avant d'aborder la question de leur prise en charge, nous conviendrons cependant d'envisager de manière intégrative les différentes marques dialogiques (ou tours polyphoniques) comme des formes de dédoublement des points de vue, en soulignant que le commentaire apporté par le dédoublement peut être restreint, comme c'est le cas dans le conditionnel : dans « Paul viendrait demain », il y a bien un dédoublement par lequel le pdv1 « Paul viendra demain » est présenté comme point de vue, sans que rien cependant ne soit dit de sa validité : le conditionnel, envisagé localement, laisserait cette question en suspens comme on l'a suggéré.

Comme le soulignent J. Authier-Revuz et J. Bres et B. Vérine, une perspective intégrative d'analyse de tours polyphoniques divers doit ainsi permettre de rendre compte de leurs spécificités ; c'est à ce titre que nous retiendrons l'intérêt qu'il y a à voir en quoi le point de vue est montré comme tel par la mention d'un procès de perception et/ou d'une source énonciative d'une part, et d'autre part comment cette première question rencontre celle du dédoublement des points de vue. Une telle piste pour l'analyse de la polyphonie permet alors de poser la question du lien entre composante modale du point de vue et polyphonie comme on l'a suggéré *supra* à propos du point de vue et sa prise en charge.

Ainsi, dans un tour polyphonique du type « ce mur n'est pas blanc » ou « Paul viendrait demain », il s'agira dans une perspective double d'analyse polyphonique et modale de s'interroger sur la composante modale du pdv1 (« ce mur est blanc », « Paul viendra demain », comme on l'a évoqué en traitant de ces deux exemples au début de cette partie), et sur celle du pdv2 : si celui-ci consiste en un commentaire sur pdv1 (« pdv1 n'est pas valide », ou « pdv1 est un pdv »), la composante modale du pdv2 dans le cas des tours polyphoniques se construit vraisemblablement de manière complexe, dans ce commentaire introduit par

pdv2, dans la question de sa composante modale, et dans la manière dont se manifeste la composante modale du pdv1.

Cela étant, au risque de nous répéter, nous conviendrons qu'une analyse tenant compte des implications mutuelles de ces phénomènes de modalisation et de polyphonie reste difficilement envisageable à partir d'énoncés décontextualisés. En prenant comme objet les textes eux-mêmes, on doit pouvoir y étudier la construction polyphonique du point de vue, dont la composante modale contribue au jeu polyphonique tout en se caractérisant par ce jeu.

Un tel objectif n'est pas contraire semble-t-il à l'objet des approches énonciatives, i.e. l'énoncé. La définition que donne E. Benveniste de l'énoncé n'exclut pas la possibilité de définir un texte en tant qu'énoncé : le critère définitoire est celui du résultat matériel d'un acte d'énonciation, dont une des propriétés est de consister en une énonciation représentée, i.e. de fournir une représentation de l'acte d'énonciation dont il se présente comme le produit. Cela étant, comme l'analyse polyphonique privilégie généralement des énoncés restreints au cadre phrastique, se pose comme nous l'avons dit à plusieurs reprises le problème du passage entre le niveau phrastique et une analyse menée au niveau textuel. Cette question est complexe : le texte ne saurait être envisagé exclusivement comme une suite d'énoncés (ou de phrases). C'est ce dont nous traiterons pour finir.

2.4. Bilan. La polyphonie au niveau textuel : la notion de figure de l'auteur

Si l'on donne pour objet à l'approche polyphonique de rendre compte de la polyphonie au niveau du texte, la question n'est évidemment pas seulement de s'attacher à la description minutieuse des points de vue et de leur prolifération.

Dans cette perspective, nous nous référerons à F. Rastier (2001) pour souligner que l'interprétation repose sur un double principe de détermination du global par le local et du local par le global. Si le point de vue est bien l'unité sur laquelle doit se fonder l'analyse polyphonique du texte, la démultiplication à l'infini des points de vue et de leurs sources au niveau du texte doit, c'est du moins une hypothèse que l'on peut faire, être contenue par ce qui se joue au niveau global.

Ce n'est que dans ce sens que se justifie à notre sens le recours à une instance qui dans l'énoncé s'y présente comme son responsable, le LOC, ou locuteur au sens ducrotien. L'introduction de cette instance est en effet loin d'aller de soi dans le cadre d'une approche polyphonique, attentive à la multiplicité des points de vue et des voix.

Alors que l'objet même de l'approche polyphonique, tel qu'il est annoncé par O. Ducrot comme fondement de son esquisse théorique, « est de contester – et si possible de remplacer – un postulat [...], [celui de] l'unicité du sujet parlant » (1984 : 171), il est étonnant de constater que le fait de « mettre en doute le postulat selon lequel un énoncé fait entendre une seule voix » (ibid.) ne va pas sans réintroduire une unicité du sujet au niveau de cet énoncé en supposant que, en dépit de la prolifération des voix, un LOC s'y dessine.

Postuler un tel locuteur est conforme à une conception naïve du langage comme instrument de communication entre deux locuteurs. Mais dans une telle approche spontanée, poser l'existence dans le texte d'un locuteur responsable de l'énoncé (en ce qu'il se présente comme tel dans l'énoncé) se comprend par le fait que la relation entre énonciation et énoncé ne peut être envisagée que de manière mécanique : le *je* intérieur à l'énoncé renvoie dans cette optique au locuteur physique des échanges, celui qui profère *je*.

Or, si l'on prend comme point de départ le point de vue, ou même, une instance supposée être à sa source, l'énonciateur, force est de constater l'irréductible hétérogénéité de cette instance dite locuteur, comme le montrent par exemple les travaux de J. Authier-Revuz

(1995) sur les phénomènes de dédoublement énonciatif. C'est aussi ce à quoi conclut A. Rabatel (1998) à propos de la sphère du narrateur : alors même qu'il se donne pour objectif d'analyser cette instance, le fait de prendre comme point de départ les points de vue le conduit en définitive, rappelons-le, à déconstruire en quelque sorte la notion, pour montrer que s'il existe bien une sphère du narrateur, celle-ci reste foncièrement hétérogène, et pour partie superposée à celle du/des personnage(s).

A notre sens, on ne saurait donc aborder l'énoncé en y postulant comme allant de soi un locuteur, sauf à renouer avec le paradigme d'intériorité et d'extériorité, celui même dont on entend s'écarter en problématisant les relations entre énoncé et énonciation et en définissant comme objet de la linguistique l'énoncé ou le texte. La manière dont est envisagée la représentation que fournit l'énoncé de l'énonciation, semble être à l'origine du postulat d'un locuteur d'une part, mais aussi de toute forme de substantialisation que dénonce A. Rabatel (1997) à propos du point de vue : dans l'analyse du point de vue, c'est la quête d'un focalisateur plutôt que de marques linguistiques qui a primé historiquement ; dans l'analyse de l'énoncé, on pourrait peut-être voir une forme de cette substantialisation dans le fait d'envisager le locuteur comme metteur en scène, doté d'intentionnalité selon une projection mimétique des attributs supposés du sujet parlant sur les caractéristiques du locuteur.

La métaphore théâtrale qui veut que l'on envisage l'énoncé comme une scène est particulièrement intéressante en ce sens, car elle est extrêmement fréquente : elle fait l'objet de développements (à la scène s'ajoutent par exemple les acteurs et le metteur en scène, sans compter les rôles), tout en étant éventuellement thématisée dans ses limites et ses difficultés, et présentée comme inadéquate ou imprécise. Si O. Ducrot (2001) parle d'abus de langage comme on l'a vu, à propos de la personnification de l'énonciateur, c'est chez Authier-Revuz (1998 : 71) que l'on trouve dénoncée la représentation d'un sujet tout puissant que suggèrent ces métaphores théâtrales : elles nient le sujet toujours centré, clivé, divisé tel qu'il se constitue dans les textes, en reproduisant une illusoire position de maîtrise du sens déconstruite par l'analyse même de la polyphonie.

Au terme de ce parcours des approches polyphoniques, nous sommes loin de pouvoir envisager dans un cadre intégrateur l'ensemble des propositions que nous avons évoquées, qui offrent un vaste programme de recherche, et montrent toute la complexité de la polyphonie. Cependant, un certain nombre de pistes d'analyse se révèlent centrales, et nous permettent d'avancer dans la définition polyphonique d'une figure de l'auteur, en tant que figure constitutive du genre – au moins dans le genre qui nous intéresse, l'article de recherche.

Pour notre part, nous refuserons ces attributs d'intentionnalité aux instances énonciatives relevant du texte, car il s'agit de les concevoir en des termes textuels précisément. Or, les observables font état de l'irréductible polyphonie dans les textes, et il faut alors ré-interroger la notion de locuteur en tant que représentation que donne l'énoncé de son responsable. Si dans une approche polyphonique, il s'agit d'analyser la multiplicité et la hiérarchisation des points de vue, c'est ce jeu polyphonique comme nous l'avons dit qui construit le locuteur plutôt que l'inverse. Par ailleurs, si l'on peut dire à l'instar de la ScaPoLine que le locuteur construit la configuration polyphonique au niveau du texte, ce n'est que dans la mesure où la configuration polyphonique est telle au niveau du texte qu'elle fournit l'image d'une instance qui construit la configuration.

En particulier, la notion de surénonciation proposée par A. Rabatel (2004a) doit être utile à montrer comment un énonciateur domine le jeu énonciatif, et en quoi cette position de surplomb caractérise dans certains genres l'instance désignée comme « l'énonciateur principal » dans la perspective de A. Rabatel, i.e. », i.e. l'énonciateur à la source d'un point de vue indexable au locuteur au sens d'image donnée dans le texte de son responsable.

Cependant, l'analyse linguistique de la polyphonie a besoin comme nous l'avons dit de postuler une telle image du locuteur dans le texte. C'est précisément le sens de la notion d' « énonciateur principal » que nous venons d'évoquer. Nous considérerons donc qu'en vertu d'une détermination du global sur le local, on aborde les textes avec, au moins pour certains d'eux, une attente quant au locuteur : ce serait-là une des caractéristiques du genre, qui explique que, lisant un texte, on s'attend à y trouver, ou à voir s'y construire une telle instance définie comme une image que donne le texte de son responsable.

Il nous faut alors nous référer à M. Foucault (1969b), car cette idée de principe herméneutique par lequel on postule un sujet à la source du sens dans le texte définit ce qu'il propose d'appeler la fonction-auteur. Sur la base notamment des écrits de M. Foucault, M. Couturier parle de « figure de l'auteur » pour évoquer à propos du texte littéraire ce principal sujet énonciatif reconstruit dans l'acte même de lecture (1995). Nous parlerons à sa suite de figure de l'auteur pour rendre compte de cette instance qui dans le texte s'y présente comme son responsable ; elle est appelée par le genre, et rend possible la saturation de l'irréductible polyphonie observable dans le texte, selon un principe d'économie dans la prolifération des voix.

La figure de l'auteur représente donc à la fois une illusion nécessaire, appelée par un genre, et la résultante du jeu polyphonique dans le texte. Elle implique qu'existe, en dépit de la prolifération des voix, ce que K. Fløttum (2002a : 342) propose de nommer une « cohérence polyphonique », c'est-à-dire une cohérence entre les différents points de vue et leurs énonciateurs, cohérence dont on peut faire l'hypothèse avec cet auteur qu'elle est attendue par le récepteur.

Dans notre perspective, il s'agit de postuler la construction d'une figure de l'auteur suffisamment cohérente pour justifier qu'on la traite en ces termes : le genre qui nous intéresse, l'article de recherche, appelle une cohérence des points de vue et de leurs énonciateurs telle qu'un point de vue – ici au sens commun d'opinion ou de thèse – cohérent se construit dans le texte, et avec lui la figure du responsable de ce texte, qui est à la fois celui qui produit le texte et le garant des contenus qui y sont assertés, si l'on se réfère à la définition donnée par M. Foucault (1969b) de la fonction-auteur.

Comme le souligne M. Foucault, et comme le montrent les analyses polyphoniques que nous avons évoquées, cette fonction de saturation dans l'interprétation de ce que les faits linguistiques ne peuvent saturer demande à tenir compte cependant de son irréductible hétérogénéité. C'est l'objet même d'une analyse énonciative et polyphonique de la figure de l'auteur que de tenter de caractériser cette figure dans la matérialité des textes, en montrant en quoi se constitue une figure de l'auteur, déterminée par le genre, et en quoi cette figure reste foncièrement hétérogène.

En guise de bilan, soulignons d'abord que ce que nous avons dit de la polyphonie, avec les approches de O. Ducrot, de la ScaPoLine et de A. Rabatel dépasse nos objectifs d'analyse du genre, mais fournit des pistes de réflexion pour une description des marques de « soi » et de l' « autre » dans les textes, et suggèrent l'intérêt de travailler ces approches les unes par rapport aux autres, car elles semblent pouvoir s'éclairer mutuellement.

Nous résumerons ici quelques principes pour l'analyse énonciative et polyphonique de la figure de l'auteur dans le genre de l'article de recherche. Le point le plus important sans doute est que la figure de l'auteur ne saurait être restreinte à ce qui est généralement traité en termes de « présence du scripteur dans son texte » au moyen des seules marques de personne.

Elle repose sur une problématisation des notions d'énoncé, d'énonciation effective et d'énonciation représentée : l'énonciation telle qu'elle est représentée dans l'énoncé ne peut être vue comme un simple reflet, secondaire par rapport à l'énonciation effective. Les

marques personnelles intéressent cependant, avec d'autres faits linguistiques, la question de déterminer dans quelle mesure le texte tend à un effacement énonciatif (Philippe, 2002, Rabatel, 2004a). La subjectivité manifeste dans l'article de recherche doit donc reposer pour partie sur cet effacement énonciatif, et sur d'autres marques de subjectivité telles qu'analysées par C. Kerbrat-Orecchioni (1980).

Les approches polyphoniques suggèrent qu'une telle figure peut être postulée comme principe d'économie dans la prolifération des voix, mais qu'elle doit cependant demeurer foncièrement hétérogène : en passant d'une phrase à l'autre, rien ne permet d'affirmer qu'on aura affaire au même locuteur représenté *via* un *je* par exemple. Il s'agit donc d'analyser une subjectivité telle qu'elle est à la fois hétérogène et cohérente d'un point de vue énonciatif, et à la fois relativement stable et variable dans le genre.

Avec les marques de subjectivité, les approches de la polyphonie offrent des outils pour l'analyse des marques de « soi » et de l'« autre » caractéristiques du genre. Elles montrent que les marques de l'autre sont loin de se réduire aux formes les plus attendues de la présence de points de vue autres (citation, discours rapporté) : il est possible d'y rattacher les points de vue montrés comme tels par le dédoublement des points de vue qui caractérise la polyphonie (par exemple, dans la négation, le conditionnel, etc.).

La question de la source de ces points de vue mérite qu'on s'y attarde, car elle doit permettre notamment de rendre compte de différences entre les tours polyphoniques, et de la manière dont un genre détermine le jeu polyphonique : avec la citation, la source est désignée en tant que source locutoire (locuteur actualisé), mais comme d'autres cas sont attestés, il importe de prendre appui d'abord sur les points de vue pour traiter alors de cette question de leur source.

La figure de l'auteur se présente comme une figure englobante, à la source du texte, des contenus qui y sont assertés et du jeu énonciatif qui s'y manifeste. La hiérarchisation des points de vue doit permettre de décrire en quoi la figure de l'auteur repose sur une position de surplomb énonciatif, et en quoi elle peut être conçue en cela comme une figure de garant du texte. La notion doit ainsi offrir une prise sur l'analyse de la dimension argumentative de l'article, à partir de la construction d'un point de vue d'auteur et de la légitimation de ce point de vue selon des procédés énonciatifs et polyphoniques caractéristiques du genre.

Chapitre 3.

Problématique et méthodologie de la recherche.

Les chapitres 1 et 2 nous ont permis de préciser les fondements de la notion de genre et celle de figure de l'auteur et de progresser vers la définition d'un cadre pour l'analyse des pratiques d'écriture, qui détermine les orientations de la présente étude. Notre objectif est de caractériser un genre en particulier, celui de l'article de recherche, dans ses régularités et sa variabilité, et de proposer des pistes pour poursuivre cette étude par des analyses ultérieures sur le genre, et la description des textes et des unités linguistiques.

Nous nous intéressons au genre de l'article de recherche dans deux disciplines de sciences humaines, les sciences du langage et les lettres. Nous faisons l'hypothèse qu'une identité disciplinaire se manifeste dans l'écriture de l'article, au sens où la discipline doit exercer une influence observable dans les textes. La comparaison disciplinaire représente avec le genre un point de départ pour mettre en relief les caractéristiques de l'article dans chaque discipline, et pour l'analyse de la diversité des textes. Une seconde hypothèse intervient, pour étudier les pratiques d'écriture de l'article en termes d'acculturation au genre, celle d'un effet du statut du scripteur sur les textes produits.

Par ailleurs, nous faisons également l'hypothèse d'usages singuliers que chaque auteur fait du genre, i.e. les styles. L'objectif est de décrire des styles d'auteurs dans l'écriture de l'article, et en cela, de mieux cerner encore la diversité des textes ; en d'autres termes, il s'agit de rendre compte de la manière dont cette diversité se structure en fonction de critères linguistiques de différenciation, et en fonction des conditions de production de textes, envisagées comme telles à titre d'hypothèse, et dont la pertinence sera déterminée en termes d'effets sur les textes.

La caractérisation différentielle des textes constitue un fil directeur de l'analyse, et il s'agit d'aborder le genre dans son ensemble, en multipliant les aspects pris en compte : ainsi les aspects sémiographiques, structuraux, morpho-syntaxiques, lexicaux et énonciatifs font-ils l'objet d'une analyse exploratoire, à la fois quantitative et qualitative. Nous entendons cependant ne retenir que les aspects dont l'analyse aura suggéré l'intérêt pour la description des régularités et de la variabilité du genre, et pour une description du genre qui tienne compte de la fonction dans les textes des traits linguistiques étudiés.

Si certains des traits analysés peuvent être interprétés en termes de rapport au lecteur ou à l'objet, nous accordons une attention spécifique à la figure de l'auteur ; le caractère institutionnalisé de l'article de recherche autorise l'hypothèse d'une figure de l'auteur

constitutive du genre, qui permette d'aborder de manière intégrative les « marques de soi et de l'autre » dans les textes, en lien avec la construction des savoirs et la dimension argumentative de l'article. L'analyse énonciative et polyphonique de la figure de l'auteur dans l'article vise à asseoir dans la matérialité du texte l'hypothèse rhétorique que le genre implique un rôle relativement stable endossé par le scripteur, ou plus exactement une représentation dans le texte de l'instance qui en est à l'origine, en tant que sujet social et praxéologique, et qui sert de garant aux contenus assertés.

L'objectif étant de caractériser le genre de l'article de recherche, la linguistique de corpus représente une entrée méthodologique particulièrement adaptée. Celle-ci est loin bien entendu de constituer un champ homogène, ne serait-ce que parce que la définition de ce qu'est un corpus dépend des fins de l'analyse.

Dans notre optique, il s'agit de constituer un corpus de textes génériquement homogène, et de le structurer de manière à contrôler sa variabilité et à traiter de l'effet de la discipline et du statut de l'auteur sur les textes. Pour décrire le genre dans ses régularités et dans sa variabilité, la quantification des traits caractéristiques des textes s'avère nécessaire, de même que le traitement statistique des données. L'automatisation de l'analyse est mobilisée pour l'étiquetage morpho-syntaxique et pour les requêtes dans le corpus, mais est précédée et relayée par des observations manuelles : le but est d'une part de prendre en compte davantage de traits textuels, et d'autre part de mettre en relation les données quantitatives avec le fonctionnement des caractéristiques linguistiques dans les textes, et c'est le rôle notamment de l'exemplification. L'analyse du genre que nous proposons se caractérise donc par une automatisation partielle d'une part, et par des va-et-vient entre une approche quantitative (manuelle et/ou automatique) et une approche qualitative d'autre part.

Nous définissons ainsi notre approche en termes de linguistique de corpus quanti-qualitative. Elle porte sur un nombre important de traits textuels, et doit articuler une description globale du genre et une description fine des caractéristiques des textes. Ces dernières concernent le niveau macro-textuel (la structure de l'article), le niveau de la phrase et les composantes linguistiques « traditionnelles » (par exemple les temps verbaux, le lexique, mais aussi les ponctuations), et les aspects énonciatifs et polyphoniques du genre.

Avec la discipline et le statut de l'auteur, une dernière entrée doit comme nous l'avons dit permettre de mieux rendre compte de la variabilité de l'article : la variation stylistique, propre à l'usage singulier que chaque auteur fait du genre. L'étude de cette dernière requiert un protocole méthodologique spécifique, que nous abordons pour cette raison au chapitre dédié à cette étude (chapitre 6.). Les réflexions sur le corpus et sur la démarche d'analyse fournies ici valent cependant aussi pour cette analyse spécifique.

Nous abordons dans ce chapitre la construction d'une méthodologie adaptée à notre objet, en nous centrant successivement sur le corpus (partie 3.1.), l'analyse quanti-qualitative (partie 3.2.) et les traits textuels retenus (partie 3.3.) pour décrire le genre de l'article.

3.1. Le corpus

Quand elle ne se fonde pas sur des exemples fabriqués pour les fins de l'analyse¹, l'étude des phénomènes en linguistique repose sur un corpus, c'est-à-dire, en première approche, un recueil de pièces ou de documents attestés. La définition du corpus en tant que collection de données qui servent d'échantillon du langage (Sinclair, 1991, 1996) reste cependant sans doute trop limitée : le corpus ne renvoie pas à du matériau langagier tout venant, mais doit résulter d'un « regroupement raisonné » (Habert, 1998 : 35) ; cela revient à affirmer que les données ne sont jamais que celles que l'on se donne. Dans cette acception que B. Habert et al. (1997 : 144) qualifient de restreinte, le corpus est donc entendu comme une construction, dictée en premier lieu par les objectifs théoriques de la recherche : le corpus repose sur des critères explicites, établis en référence aux hypothèses qu'il doit permettre d'étayer empiriquement.

Ce premier trait définitoire signale un parti pris qui n'est pas partagé dans l'ensemble des approches en « corpus linguistics »², ou, en contexte francophone, en linguistique de corpus. L'essor des conférences et des publications sur le sujet, depuis les années 1990 en particulier atteste de la constitution d'un domaine³ ; loin d'être unifié, celui-ci résulte cependant de la volonté de confronter les différentes approches : il s'agit de discuter la diversité des objectifs et des méthodes d'exploitation des corpus.

Dans ce cadre, une ligne de démarcation se manifeste comme le souligne M. Péry-Woodley (1995)⁴ entre les tenants du « gigantisme » (« more data is better data ») et ceux, auxquels notre étude se rallie, du corpus en tant que construction raisonnée. Derrière cette question, c'est bien entendu celle de la représentativité des données qui se pose. Or, celle-ci est extrêmement délicate concernant les usages langagiers, et la difficulté de constituer un corpus « équilibré », représentatif de la langue « générale » sert d'argument au gigantisme, rendu possible par les développements informatiques.

La question de la représentativité n'est pas pour autant résolue, et aboutit à une impasse. Qu'est au juste ce « tout » de la langue que le linguiste entend cerner ? La multiplication des données ne dédouane pas d'un questionnement sur les objectifs de l'analyse, car il reste à positionner son approche par rapport à la langue que l'on cherche à décrire. Non seulement l'exhaustivité est irréaliste, mais le problème de la variation des productions langagières et des paramètres qui la structurent se pose de manière centrale⁵. La linguistique du genre, à ce titre, doit permettre comme nous l'avons vu (chapitre 1.) de dépasser les obstacles posés par la quête de cette « langue introuvable » qu'est la langue générale (Gadet et Pêcheux, 1981).

S'agissant de décrire un genre, la conception du corpus dans le cadre de cette étude se trouve encore spécifiée. Le corpus doit réunir des textes et c'est là encore un positionnement qui ne fait pas l'unanimité. Comme l'approche en corpus répond d'une exigence d'empirisme,

¹ On pourrait dire cependant qu'il s'agit là d'un corpus « introspectif », issu tout droit de la compétence linguistique du chercheur ; cette option méthodologique, utile en fonction des objectifs poursuivis, ne va pas sans débats quant à l'acceptabilité/ l'inacceptabilité des énoncés retenus, et pose aussi la question de leur authenticité et de leur représentativité en termes d'usages de la langue.

² Issue d'une linguistique descriptive de tradition anglo-saxonne (Habert, Nazarenko et Salem, 1997).

³ Dans le sens où certains se réclament de l'utilisation de corpus et se définissent même par cette utilisation comme le souligne M. Péry-Woodley (1995).

⁴ Cf. aussi M. Bilger, 2000, A. Condamines et al., 1999, B. Habert et al. 1997.

⁵ Cela revient à dire, comme le souligne F. Gadet, que « le constat de l'hétérogène est co-extensif à la notion de langue » (1989 : 7), et que face à « l'idée fallacieuse d'une homogénéité discursive comme règle générale du langage » (Vargas, 1985 : 53), la variation demande à passer du statut de « phénomène latéral » à celui de « phénomène central » (François, 1980).

l'idée de données langagières attestées fait consensus ; cependant la conception implicite ou explicite du corpus, et, à travers elle, de la langue que l'on cherche à décrire détermine de nouveau une opposition entre l'analyse *des textes* et les approches qui entendent recueillir *du texte*. Ces dernières posent prioritairement le problème de « soumettre à l'analyse un objet dont l'hétérogénéité est totalement opaque » (M. Péry-Woodley, 1995 :216).

Dans notre optique, la prise en compte de textes intégraux est requise et l'importance de la question de leur diversité demande de se doter de critères d'homogénéité et de contrastivité des textes (Mayaffre, 2001). A l'encontre de l'idée d'une collection de textes, la pertinence du corpus doit reposer sur sa structuration et sa documentation. Nous reprendrons donc la définition du corpus que donne F. Rastier (2002) :

un corpus est un regroupement structuré de textes intégraux, documentés, éventuellement enrichis par des étiquetages, et rassemblés : (i) de manière théorique réflexive en tenant compte des discours et des genres, et (ii) de manière pratique en vue d'une gamme d'applications.

La réflexivité des corpus, défendue notamment par D. Mayaffre (2002) au titre de ligne d'horizon vise à terme une objectivation des rapports intertextuels et un contrôle des parcours de lecture – ou d'analyse – des textes.

En tant qu' « objet heuristique »¹ et construit qui a pour but de décrire le genre de l'article, le corpus se doit de représenter la diversité des usages, tout en répondant à un impératif, celui de l'homogénéisation générique des textes retenus. Comme le genre se caractérise par un double principe de stabilité et de variation, il importe d'éviter que des genres différents se télescopent, annulant par là réciproquement leurs spécificités (Rastier, 2002). Les critères de sélection des textes retenus dans le corpus de travail opèrent ainsi des restrictions quant au choix des textes, et vont permettre de contrôler leur diversité par la structuration du corpus en fonction d'un certain nombre de paramètres. Certains de ces paramètres ont le statut de variables dont on cherche à évaluer la pertinence en termes d'effets sur les caractéristiques des textes.

3.1.1. Un corpus de textes du même genre : critères de sélection des textes

La sélection des textes repose sur l'exigence principale de constituer un corpus génériquement homogène. Si la question de la représentativité est délicate, il importe pour le moins de se doter d'un corpus d'une taille suffisamment importante et présentant une certaine diversité interne pour limiter les risques que comporte toute entreprise de généralisation (Habert, 2000), à savoir l'incertitude liée au caractère aléatoire de l'échantillonnage et les biais que peut introduire notamment la sur-représentation de tel ou tel domaine de spécialité, d'une revue, d'un auteur, etc.

Nous abordons donc les critères de sélection en termes d'identité du genre. Ils visent le contrôle de la diversité des textes en opérant des restrictions et en structurant les textes retenus ; l'étude requiert d'exclure certains axes d'investigation et de neutraliser ou à l'inverse de rendre observable l'effet potentiel de certains paramètres sur le genre.

Nous traitons ici des critères de sélection des textes en termes d'exclusions : ils ont trait notamment à la langue de l'article, à la nécessité d'étudier le genre en synchronie, et à la dimension éditoriale du genre. A cela s'ajoute, pour les besoins spécifiques de l'étude la question de l'identité énonciative des textes. Nous envisageons ensuite les critères de constitution du corpus en termes d'analyse de la variabilité du genre, avec deux variables

¹ C'est ainsi que D. Mayaffre (2002) propose de concevoir le corpus, pour pallier les difficultés posées par la notion de « données » en linguistique.

principales qui retiennent notre attention, la discipline et le statut de l'auteur. Comme le texte de l'article a ceci de particulier qu'il relève d'un autre texte, celui du numéro de revue, les critères affectent la sélection des articles, des numéros et des revues.

Des articles de recherche français

Le premier critère employé se situe dans l'identité linguistique et culturelle du genre. Nous avons fait le choix de ne retenir que des articles francophones, issus de revues elles-mêmes complètement ou majoritairement francophones. Par ailleurs, la diversité culturelle de la francophonie n'étant peut-être pas sans effet sur les pratiques d'écriture de l'article, il nous a semblé important de nous restreindre aux revues publiées en France. L'identité linguistique et culturelle concerne donc prioritairement les revues retenues.

Le genre en synchronie

Veiller à une homogénéité des textes du corpus suppose aussi de se doter de critères ayant trait à la dimension historique du genre. Dans cette optique, comme nous cherchons à étudier les pratiques contemporaines de l'écriture de l'article, nous avons privilégié les revues actuellement encore en activité (au moins au moment où nous recueillons les données) ; plus important encore, les articles étudiés doivent être publiés dans des numéros récents de la revue : la période délimitée va de 1998 à 2003, de manière à garantir une appréhension synchronique des normes du genre, cette synchronie représentant, comme les autres critères de sélection du corpus, un artefact¹ nécessaire à une caractérisation du genre qui rende compte de sa stabilité et de ses principes de variabilité.

La dimension éditoriale du genre

Le regroupement de textes du même genre prend son point de départ dans la catégorisation sociale des genres ; tout locuteur, dans un champ donné, est à même d'attribuer une identité générique à un texte, de sorte qu'une catégorisation générique *a priori* peut se donner comme fondement les étiquettes génériques communément et spontanément adoptées. Cela étant, dans le cas d'un genre institutionnalisé comme l'est l'article, force est de tenir compte de l'éditorialisation des textes : le dispositif éditorial joue sur la définition du genre, par l'assignation d'une appartenance générique² aux objets textuels et par les procédures de sélection et de correction des épreuves.

L'importance de l'éditorialisation dans la définition du genre impose donc de veiller à une homogénéité éditoriale, à l'origine de trois critères de sélection :

- ne retenir que des textes publiés
- ne retenir que des textes publiés en tant qu'articles de revue
- ne retenir que des textes publiés en tant qu'articles dans des revues « à comité »

¹ Cf. J. M Adam et U. Heidmann (2004) pour une réflexion sur la synchronie comme artefact dans l'analyse des genres.

² Les étiquettes génériques assignées au niveau éditorial peuvent entériner la classification spontanée des genres, comme elle peut, aussi, en jouer, dans le cas du discours littéraire en particulier.

Le premier critère consiste à ne retenir dans le corpus d'étude que des articles publiés, et de disposer des textes tels qu'ils ont été publiés, et non d'une version antérieure (ou postérieure). Ce critère a déterminé le choix de nous adresser directement aux responsables des revues pour réunir les articles. Il est en effet délicat de s'adresser directement aux auteurs pour leur demander leurs textes, sauf à s'assurer que ces versions sont parfaitement conformes à celles qui sont effectivement publiées, quitte à les modifier, le cas échéant, dans le sens de cette conformité.

Des textes publiés en tant qu'articles de revue

Au sein du discours scientifique, il existe nombre de genres qu'on peut, *a priori*, considérer comme différents, ne serait-ce qu'intuitivement, parce que tout scripteur et lecteur a bien conscience du fait qu'écrire un compte-rendu d'ouvrage n'est pas écrire une communication orale pour un colloque.

Pour paraître évidente, cette remarque est pourtant loin d'aller de soi, car elle pose la question de savoir comment se structurent les différents genres au sein d'un discours donné. Cette question conduit F. Rastier (2001) à introduire entre le niveau des genres et celui des discours un niveau intermédiaire, le champ générique, défini comme un groupe de genres qui contrastent ou rivalisent dans une pratique. Dans une approche contrastive, on sera ainsi amené à privilégier la caractérisation des genres au sein d'un même champ générique ; il ne s'agit pas de renouer avec une approche taxinomiste mais d'envisager comment un genre se structure dans des réseaux de genres.

Parmi les champs de contraste auxquels prêter attention pour répondre aux besoins d'homogénéisation du corpus, nous avons d'abord distingué l'article des autres genres que l'on trouve dans un numéro de revue : l'avant-propos, l'introduction ou la présentation du numéro qui apparaissent généralement dans les opus, au moins en partie thématiques, de même que les éventuels « conclusions » et « bilan » du numéro d'une part, et les comptes-rendus d'ouvrages, notices ou notes de lecture d'autre part. Nous considérons par hypothèse que ces textes ont chacun une identité générique spécifique et un statut particulier dans l'économie du numéro. En témoignent les indications portées dans le sommaire, ou sur l'entête des pages à venir : « présentation », « notes de lecture », etc. Produites au niveau éditorial, ces indications structurent le numéro en différents genres¹.

Un autre champ de contraste possible se situe dans l'opposition entre « article de revue » et les textes qui consistent en des contributions à des ouvrages collectifs², des « hommages à... », ou des actes de colloque, lesquels peuvent paraître sous la forme d'ouvrage ou de numéro de revue. En l'absence d'études sur la question, il convient de ne se centrer que sur les articles de revues, en excluant les numéros de revue consistant en des actes de colloque. Mais le problème est délicat, car il n'est pas possible, et tel n'est pas notre objet,

¹ Chaque revue définit donc à sa façon le champ générique qui structure ses livraisons. Il arrive que la partie correspondant aux « articles » soit ainsi titrée dans le sommaire, une précision sur le caractère thématique ou non de cette partie étant apportée quand le numéro comprend les deux. Cette structure varie ainsi d'une revue à une autre, voire d'un numéro à l'autre dans quelques rares revues, mais l'on retrouve globalement les mêmes éléments, qui contribuent à définir ce qu'est, textuellement, un numéro de revue.

² Contributions généralement alors désignées comme des chapitres.

de déterminer pour chaque texte sa lignée génétique. Nous ne pouvons nous fier qu'à la matérialité des textes eux-mêmes, en l'occurrence aux indications portées sur le numéro de la revue, une fois mis à l'écart les volumes autres que les revues. Il se peut cependant que certains numéros retenus dans notre corpus soient constitués de contributions liées à un colloque, sans que cela soit mentionné dans le numéro.

Des articles publiés dans des revues « à comité »

Comme nous l'avons souligné, l'évaluation des articles en amont de leur publication exerce par hypothèse une contrainte forte sur le genre. Le rôle joué par l'évaluation dans la sélection des textes et dans leur production (demandes de modifications, etc.) est à l'origine d'un critère supplémentaire : la présence d'une instance chargée de cette évaluation dans les revues représentées dans le corpus, à savoir un comité de lecture et/ou de rédaction.

Les revues identifiées comme étant « à comité » bénéficient d'une accréditation plus grande dans le champ. L'évaluation entre pairs, en tant que procédure de validation externe de la construction du savoir, est perçue comme un gage de scientificité : l'expertise des articles est censément garante de leur qualité. Cette valeur distinctive se manifeste jusque dans les curriculum vitae, où les publications dans les revues « à comité » font l'objet d'une rubrique spécifique. Mais ce n'est pas tant la valeur sociologique de cette spécificité qui nous intéresse : si nous excluons les articles autres que ceux publiés dans de telles revues, c'est parce que nous faisons l'hypothèse qu'il s'agit là d'un critère qui peut exercer une influence observable sur les textes publiés, ne serait-ce que parce que ne sont pas publiés les textes évalués comme non recevables.

Mise à part cette exigence de présence de comité(s), aucune restriction n'est faite quant à la notoriété de la revue ; on cherche au contraire à réunir des articles publiés dans des revues nationales et locales, éditées par de grandes chaînes d'édition ou au sein de laboratoires. Certaines revues représentées dans le corpus, d'envergure nationale, bénéficient ainsi d'une large reconnaissance de la part des chercheurs du champ (on peut citer notamment *Langages*, ou *Langue française*, pour les sciences du langage, la *Revue de Littérature Comparée*, *Romantisme*, *Littérature Classique* pour les études littéraires). À côté, d'autres revues, locales, restent davantage méconnues. Si la diffusion d'une revue peut être prise comme indice de la qualité scientifique des contributions, nous abordons pour notre part cette qualité en termes d'expertise, en raison du rôle supposé de cette dernière sur les normes du genre ; la question de la notoriété demanderait une réflexion spécifique, d'ordre sociologique, sur les indices à retenir, il nous semble préférable de la neutraliser en privilégiant une diversité des revues à ce niveau.

Une identité énonciative

Signalons que nous avons également dû respecter un critère de sélection répondant aux besoins spécifiques de l'analyse énonciative de l'article. Nous n'avons en effet retenu que des articles signés par un seul auteur. On introduit là un biais, car la présence d'articles co-signés n'est pas négligeable, en sciences du langage tout particulièrement si l'on en juge aux textes que nous ont adressés les responsables de revue. Mais dans la perspective d'étude des textes dans leur dimension énonciative, les articles co-signés compliquent l'analyse en créant un nouvel axe de partage dans le corpus. Comme rien ne permet de préjuger de l'absence de différence entre les articles co-signés et les articles signés par un seul auteur, nous avons préféré écarter le premier cas pour nous centrer sur le second, et envisager des pistes pour

l'analyse de la figure de l'auteur dans des textes signés par un seul auteur. La question reste ouverte de savoir s'ils sont produits par un seul scripteur.

Il semblait difficile d'avoir prise sur une telle question autrement qu'à partir de ce qui est observable, en l'occurrence le texte de l'article et du numéro. Il n'est pas impossible qu'un article ait été écrit par plusieurs auteurs pour être finalement signé par un seul. Si cette pratique ne semble pas être monnaie courante dans les champs disciplinaires concernés, il faut convenir que la signature ne représente pas une marque transparente du scripteur comme le montre D. Pontille (2001, 2004) à propos du fonctionnement complexe de la signature dans divers champs des sciences dures. Notons pour finir que, hormis ce critère relatif à la signature, aucune restriction n'est faite concernant l'auteur (sa nationalité, sa langue maternelle, son statut, ses fonctions...); nous cherchons cependant comme nous le verrons à intégrer au corpus des articles signés par des doctorants.

Les critères de sélection répondent aux besoins de l'étude en procédant par exclusions : nous nous centrons sur l'article de recherche publié en France entre 1998 et 2003 dans des revues nationales et locales avec comité, et signé par un seul auteur ; nous délaissions ainsi la dimension diachronique du genre, la question du contraste entre langues, celle du contraste entre genres. Sélectionner des textes revient aussi à neutraliser certaines variables, ou au contraire à rendre possible leur analyse en termes d'effets sur les textes. Autrement dit, la constitution du corpus requiert un contrôle de la variabilité du genre.

3.1.2. Le contrôle de la variabilité du genre : diversification des textes et définition de variables externes

En restreignant le choix des articles retenus dans le corpus en vue de respecter une homogénéité générique, les critères de sélection visent le contrôle de la diversité interne du genre, i.e. sa variabilité. Il s'agit d'une part, de neutraliser d'éventuels biais et d'autre part, de rendre possible une analyse contrastée en fonction de la discipline et du statut de l'auteur.

La diversification des revues, des numéros et des auteurs

S'agissant d'étudier le genre de l'article, il importe d'abord de se prémunir d'éventuels biais. Le premier est celui des styles personnels, i.e. de l'usage singulier que chaque auteur fait du genre. La variation stylistique représente un de nos axes d'investigation, mais pour l'étude du genre, il importe de la neutraliser en retenant des textes d'auteurs variés. La part d'aléatoire dans l'échantillonnage a pour conséquence que le corpus contient parfois deux textes d'un même auteur ; cela vaut pour cinq auteurs (quatre en études littéraires et un en sciences du langage). Quant à l'étude des styles personnels, elle requiert un corpus qui lui soit dédié, et que nous présenterons dans le chapitre consacré à ce facteur de variation (chapitre 6).

Nous avons également cherché à diversifier les revues et les numéros représentés dans le corpus, en tentant également d'éviter un déséquilibre quant au nombre d'articles issus d'un même numéro ou d'une même revue, et quant à la composition du corpus dans chaque discipline. Ce critère de diversification peut cependant entrer en conflit avec la disponibilité des textes, en particulier pour un corpus électronique constitué à partir de l'archivage que chaque revue fait de ses livraisons.

La diversité des revues et des numéros est cependant essentielle, parce qu'elle a partie liée avec la diversité thématique des articles. Si les revues se définissent d'abord par un ancrage disciplinaire, nombre d'entre elles tendent aussi à une spécification que l'on pourrait

dire sous-disciplinaire (par exemple, dans les études littéraires, un siècle ou, de manière plus spécifique encore, un auteur). Les numéros sont *a fortiori* concernés par la question des sous-domaines d'une discipline, car la plupart des revues font le choix de numéros thématiques, même si cela n'exclut pas la présence dans la livraison de certaines contributions hors thématique.

La question des thématiques nous amène alors à celle des disciplines et de leur structuration en une diversité de sous-domaines et d'écoles. Le problème du rattachement disciplinaire demande une attention spécifique dans la constitution de notre corpus, s'agissant de comparer les pratiques d'écriture de l'article dans le champ des sciences du langage et dans celui des études littéraires.

La variable « discipline » : sciences du langage *versus* lettres

Une des principales hypothèses de notre recherche est que le genre, parce qu'il est lié aux communautés de discours, varie avec la discipline. Pour étayer cette hypothèse, le rattachement disciplinaire des articles retenus comme des revues représentées dans le corpus est un critère de toute importance.

Mais cette question de la discipline pose deux problèmes :

- celui de la diversité interne d'une discipline ; toute discipline est structurée par un ensemble de domaines, de courants, d'écoles, et, finalement, d'objets. La question de la représentativité du corpus intervient de nouveau, car entendre étudier l'article de sciences du langage supposerait de disposer d'un corpus qui en reflète la diversité et les tendances principales.
- le second problème lié au premier est de savoir où commence et où s'arrête une discipline, autrement dit de s'interroger sur ses frontières, c'est-à-dire sur ses frontières avec d'autres disciplines relevant du même champ d'activité de la recherche, et ses frontières avec d'autres champs d'activité¹.

Le problème de frontière se pose de manière cruciale dans le cadre de notre approche : la plupart des études sur le discours scientifique qui comparent les pratiques disciplinaires tendent à préférer des champs éloignés, susceptibles d'offrir un panel large de ce qu'est le discours scientifique. Avec les sciences du langage et les études littéraires, nous avons au contraire affaire à deux champs que l'on peut considérer comme liés ; c'est vrai de leurs objets (le langage, avec une spécification dans les études consacrées au langage littéraire), ainsi qu'au niveau institutionnel et historique, comme en témoigne par exemple la création d'UFR² de Sciences du Langage au sein des facultés de Lettres.

Il semble possible d'avancer qu'aucun champ disciplinaire ne saurait faire l'objet d'une cartographie objective ; les dictionnaires disciplinaires entendent offrir une vue panoramique et synthétique des acquis et des questions d'un champ, mais peuvent toujours être soumis à la discussion, ne serait-ce qu'au niveau du choix des entrées retenues³. La structuration du champ disciplinaire en sous-composantes, de même que les objets que doit se donner une discipline, ne peuvent faire consensus que sur certains aspects, qui peuvent en cela être considérés comme des acquis stables du champ. La sociologie de P. Bourdieu (cf. notamment 1991, 1994) rend compte de ce problème de définition du « champ », en termes notamment de rapports de force, et montre qu'un champ ne peut pas être homogène, mais

¹ L'article de recherche en Lettres pourrait en ce sens être confronté au champ de la critique littéraire.

² Unité de formation et de recherche (composante institutionnelle des universités françaises).

³ Cette question est analysée en termes de posture énonciative à propos des dictionnaires de sciences du langage, dans F. Grossmann et F. Rinck, 2004.

qu'il ne se satisfait pas non plus d'une conception relativiste selon laquelle toute approche est possible et toute approche se vaut.

La différence disciplinaire aux fondements de notre approche correspond à un artefact. Elle est à comprendre comme une entrée pour analyser le genre de l'article dans ses régularités et sa diversité interne, et avec lui la diversité des pratiques de recherche. S'il s'agit de mettre au jour des identités disciplinaires du genre, il importe de penser en premier lieu la continuité des pratiques textuelles. Les sciences du langage et les études littéraires sont liées nous l'avons dit d'un point de vue épistémologique et institutionnel, certains linguistes peuvent avoir suivi une formation initiale en Lettres, un chercheur institutionnellement rattaché à un laboratoire de recherches en Lettres peut adopter un point de vue linguistique (cf. par exemple avec le structuralisme), un même objet peut être interrogé sous les deux angles disciplinaires (par exemple, la fiction), et certains domaines de recherche représenteraient une sous-discipline commune aux deux disciplines considérées¹.

Dans le cadre d'une approche attentive aux genres, on ne peut qu'envisager les productions textuelles en termes d'intertexte, et convenir que la stabilité des champs est relative, comme celle des genres. S'il existe des catégorisations institutionnelles, reconduites ou réinterprétées par les acteurs du champ, il existe aussi des références partagées et des pratiques communes. Méthodologiquement, il nous faut donc introduire une homogénéité là où la question des frontières mériterait un examen spécifique. Les choix effectués reviennent à procéder à une catégorisation *a priori* du genre, que l'analyse doit alors permettre d'étayer. La comparaison disciplinaire, comme l'analyse de la variation stylistique serviront ainsi de point d'entrée pour rendre compte de la diversité du genre et mettre ainsi au jour des logiques plurielles.

Pour revenir au choix des articles en fonction de ce critère disciplinaire, il faut souligner que c'est d'abord à partir de la revue que la sélection a été opérée. Nous avons ainsi pu constater que les productions interdisciplinaires sont rares : la ligne éditoriale des revues impose un rattachement disciplinaire assez strict, signalé également avec leur classification documentaire dans les bibliothèques et les bases de données. Nous avons alors sélectionné au sein de ces revues des articles et des numéros sans équivoque quant au rattachement à l'une ou l'autre des deux disciplines considérées.

En termes de diversité interne des deux champs disciplinaire, nous avons tenté d'annuler autant que faire ce peut la spécificité de certains sous-domaines. Pour les études littéraires, nous acceptons tout aussi bien des revues, numéros ou articles traitant de littérature comparée, de littérature d'un siècle ou d'un auteur en particulier et tentons de diversifier les époques ou les œuvres concernées ; en linguistique, de même, nous entendons regrouper différentes branches, telle que la phonétique, la syntaxe, la sémantique, la sociolinguistique, et l'analyse du discours. La nécessité de neutraliser le poids de certains sous-domaines ne va pas cependant sans opérer des restrictions : les revues, numéros ou articles concernant les branches « appliquées » de la linguistique et des études littéraires sont mis à l'écart, notamment tout ce qui touche à l'ingénierie linguistique et à la didactique des langues et de la littérature, afin de contrôler au mieux les facteurs de variabilité du genre de l'article. Du point de vue du lectorat en particulier, ces travaux de recherche nous semblent avoir un fonctionnement particulier. C'est dans cette même optique que certaines revues d'études littéraires, s'adressant explicitement aux étudiants préparant l'agrégation, ont été exclues.

Comme le montre cette délicate question de la discipline, c'est finalement au moyen d'un ensemble d'exclusions que l'on circonscrit l'objet à analyser. Cela étant, un critère de

¹ C'est par exemple l'idée que défend B. Combettes dans son article « Analyse linguistique des textes et stylistique » (2002).

disponibilité joue nécessairement aussi dans la constitution d'un corpus, et avec lui se pose la question des droits. Le choix de nous adresser directement aux revues s'est imposé pour cette raison, et pour une acquisition des données facilitée par l'accès aux archives (textes publiés et en version électronique). L'un et l'autre aspect ont cependant posé des difficultés : l'archivage électronique n'est manifestement pas le fait de toutes les revues au moment où nous procédons au recueil de données (2002-2003)¹. Se pose également des problèmes de formats de documents, dans la mesure où l'archivage peut être réalisé sous des logiciels d'édition spécifiques et ne peut être livré tel quel. Quant aux droits, nous avons pu relever une certaine réticence de la part de certains responsables éditoriaux à mettre à disposition les textes. La condition de gratuité n'est pas acceptée par tous, et d'autres motifs de refus semblent davantage tenir à la « fouille de textes » que nous entendons mener.

Une dernière difficulté à laquelle nous avons eu à faire face est qu'en procédant à un recueil de données par l'intermédiaire des revues et des responsables éditoriaux, il n'était pas toujours possible d'obtenir des données sur l'auteur, et en particulier sur son statut, qui intéresse un second objectif de notre étude.

La variable « statut de l'auteur » : doctorant *versus* non-doctorant

A titre d'horizon, notre recherche sur l'article se voudrait utile à un questionnement sur la didactisation du genre. Comme celle-ci pose prioritairement le problème de savoir quel modèle de référence convoquer, nous avons fait le choix d'étudier, avec les pratiques expertes du genre de l'article, la question de savoir si ses usages varient en fonction du statut plus ou moins expert et plus ou moins néophyte de son auteur.

Comme nous l'avons dit, nous ne retenons que des articles publiés en tant que tels après avoir fait l'objet d'une expertise. Nous pouvons en cela considérer leurs signataires sur un pied d'égalité : ils sont tous auteurs d'un article qui a été jugé recevable dans un champ de recherches. Aussi, par hypothèse, la spécificité statutaire peut exercer une influence observable sur les textes en termes d'acculturation au genre, mais de manière moindre toutefois que l'influence exercée par les cultures disciplinaires.

Parmi les indicateurs d'ordre sociologique qui permettraient d'objectiver la question du statut du scripteur, et d'observer en quoi ce statut induit des pratiques textuelles spécifiques, nous avons fait le choix de ne retenir qu'une seule question : celle de savoir si le scripteur est doctorant ou non. A notre connaissance, cette question n'a jamais été analysée dans le cadre d'une étude du genre de l'article, mais elle est bien entendu thématisée, en particulier dans le champ de l'English for Academic Purposes, ou encore dans l'ouvrage de H.S. Becker (2004), qui propose une réflexion sur l'écriture des sciences sociales à partir de son expérience d'encadrement de mémoires des étudiants entrant dans l'écrit de recherche.

L'opposition entre deux groupes « doctorant » *versus* « non-doctorant » repose donc sur le statut institutionnel de « doctorant », et se base sur le nom d'auteur qui sert de signature à l'article². On retrouve donc le problème de la signature, évoqué *supra* à propos de la sélection d'articles signés par un seul auteur. Rien ne permet de garantir que le nom signé est celui du scripteur effectif du texte, et c'est la raison pour laquelle nous parlons de variable « statut de l'auteur » ; celle-ci nous sert à aborder la question du statut du scripteur effectif

¹ Entre cette période et la fin de notre écrit, nous constatons même une nette évolution de la visibilité des revues sur le web. Au départ, peu d'entre elles disposent d'un espace dédié, *a fortiori* d'un espace bien documenté (sommaires des numéros, voire archives en ligne), et cela tend à devenir le cas à présent.

² Se pose donc de nouveau le problème évoqué *supra* à propos des textes signés par un seul auteur et des textes co-signés, dans la mesure où la signature ne garantit pas l'identité du scripteur effectif de l'article.

dans la mesure où, *a priori*, les articles signés par un seul nom de doctorant doivent avoir été produits par le scripteur signataire.

Par ailleurs, l'opposition entre les deux groupes de scripteurs masque la diversité interne du groupe qui n'est défini que négativement, i.e. les « non-doctorants », ainsi que la diversité interne du groupe de doctorants, si l'on convient que la durée du doctorat doit jouer sur leur acculturation au champ et au genre. Là encore, c'est donc en termes de continuité des pratiques qu'il importe d'envisager les textes et la question d'un degré de « nouveauté » ou d'« expertise » dans le champ.

La variable se veut donc utile à une focalisation sur le groupe de doctorants, pour étayer l'hypothèse d'une spécificité de leurs textes et avec elle la question du modèle de référence à convoquer pour didactiser le genre de l'article. Elle doit là encore être comprise comme une entrée qui permette avec d'autres de rendre compte de la diversité des textes, et d'interroger les facteurs linguistiques et extra-linguistiques qui structurent cette diversité.

Concrètement, notre objectif se devait d'être exposé aux responsables éditoriaux afin que, dans la mesure du possible, ils en tiennent compte en intégrant des textes de doctorants parmi les articles mis à disposition pour notre étude. Cependant les données sur le statut de l'auteur ne sont parfois pas ou plus en possession des revues ; nous avons donc dû compléter la documentation du corpus *a posteriori*, en contactant les auteurs eux-mêmes ou en nous appuyant sur une trace de leur parcours (par exemple, les pages personnelles sur le web).

Une démarche consistant à recueillir du texte tout venant et à sélectionner ensuite un certain nombre de textes de doctorants aurait sans doute rendu plus difficile encore la sélection des textes ; en contrepartie, elle permettrait cependant de sonder les usages et de voir la part prise par les articles de doctorants. On remarquera en effet que, pour avoir fait expressément cette demande aux revues, elle s'est avérée difficile à satisfaire¹ et nous avons dû revoir nos objectifs à la baisse.

Un rapide questionnaire adressé aux revues nous a montré que publier des articles de doctorants pouvait représenter un des enjeux de la ligne éditoriale (c'est le cas, selon son responsable, de la revue *Travaux de l'Institut de Phonétique de Strasbourg*), ou au contraire ne pas être avéré : nous apprenons ainsi que les doctorants ne sont pas eux-mêmes signataires d'articles, sauf en collaboration avec un enseignant-chercheur chevronné, dans les revues *Langages* et *Langue Française*, dont le responsable souligne toutefois que la question dépend du coordinateur du numéro, qui préside à la sélection des textes.

Le statut de l'auteur ne semble donc pas être en lui-même un critère exclusif, puisque la sélection reste textuelle, mais il fait possiblement l'objet de ce qu'il est commun aujourd'hui de désigner comme de la discrimination positive : parce que le « nouvel entrant dans la communauté scientifique »² qu'est le doctorant a besoin de publier pour entrer dans le champ, certaines revues – mais c'est vrai aussi de certains colloques – se proposent de les aider en leur offrant un espace réservé.

Enfin, c'est le contrôle des facteurs potentiels de variabilité du genre qui détermine les textes exclus et les textes retenus dans le corpus. La documentation du corpus doit alors permettre de rendre compte de sa structuration interne. Chaque texte est décrit au sein du corpus à partir des différents paramètres qui le définissent et le rapprochent ou le distinguent des autres. Une base de données est ainsi constituée, et définit pour chaque article la

¹ Notamment au regard des autres critères, et celui de date en particulier. Mais le fait que certains responsables éditoriaux aient prêté une attention toute particulière à satisfaire notre requête sur ce point est sans doute à mettre au compte de la diversité des attitudes sur la question de la publication des doctorants.

² P. Bourdieu (2001 :95).

discipline et le statut de l'auteur, la revue, le numéro, la date et l'auteur (le signataire du texte), et associe à ces informations les traits linguistiques analysés. Elle sert alors l'analyse statistique de l'effet des deux variables « discipline » et « statut de l'auteur », et le cas échéant, l'effet de la revue sur les caractéristiques des textes.

3.1.3. La composition du corpus

La composition exhaustive du corpus est présentée en annexes (Annexes 1) : chaque article y est identifié au moyen de ses références bibliographiques. Nous présentons ici les caractéristiques générales du corpus.

Le corpus réunit comme nous l'avons dit des articles publiés en France entre 1998 et 2003 dans des revues nationales et locales avec comité, et signés par un seul auteur. Il représente plus d'1,5 millions de mots¹ et comprend 220 articles, soit deux sous-corpus homogènes d'un point de vue disciplinaire, avec 110 articles d'études littéraires (LET) et 110 articles de sciences du langage (SCL). Concernant la variable « statut de l'auteur », le corpus rassemble 35 articles de doctorants et 185 articles d'autres chercheurs, répartis de la manière suivante dans les disciplines :

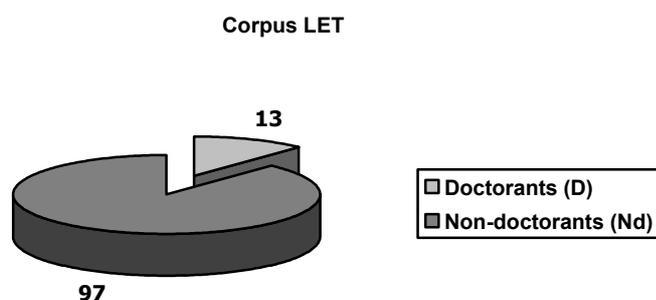


Figure 3.1. Articles de doctorants et de non-doctorants dans le corpus LET : poids en nombre d'articles.

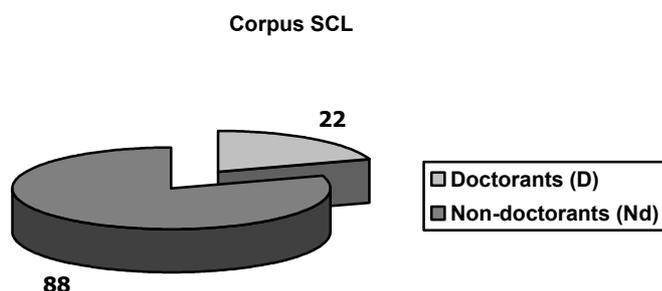


Figure 3.2. Articles de doctorants et de non-doctorants dans le corpus SCL : poids en nombre d'articles.

¹ 1535254 mots selon le comptage effectué par le logiciel Cordial Analyseur sur le corps du texte et les notes (soit l'article à l'exclusion des bibliographies et annexes).

Les articles produits par des scripteurs doctorants représentent donc un nombre relativement peu important d'articles, et moins encore dans le corpus LET que dans le corpus SCL. Ce taux relativement faible au regard nos objectifs initiaux résulte des difficultés de la constitution du corpus. La sélection opérée à notre demande par les responsables éditoriaux n'a pu être que partiellement contrôlée : cette variable a en particulier induit des pratiques très diversifiées d'une revue à l'autre, selon leur archivage ou non de ces données sur le statut de l'auteur, et la présence ou non d'articles de doctorants dans les livraisons récentes.

Les articles de doctorants ne se répartissent donc pas équitablement dans les différentes revues ; c'est là un fait dommageable pour la diversification interne du corpus, tout particulièrement en LET où les articles des doctorants ne proviennent que de 3 des 8 revues (et de 6 numéros), et où une revue, *La Licorne*, n'est représentée que par des articles de doctorants (6 articles). C'est le cas aussi en SCL, mais dans une moindre mesure : les articles de doctorants se répartissent sur 5 des 9 revues (et sur 12 numéros). Ces remarques appellent à envisager avec prudence les données observées sur le corpus d'articles de doctorants ; le recours aux statistiques doit cependant permettre de pallier la taille restreinte du groupe analysé, au regard du groupe des textes de non-doctorants.

Les articles de LET proviennent de 8 revues, et de 22 numéros différents¹. Les articles de SCL proviennent de 9 revues, et de 28 numéros différents. Le tableau ci-dessous indique les revues représentées dans chacun des deux sous-corpus disciplinaires (LET puis SCL) et leur poids en nombre de numéros et en nombre d'articles (au total dans l'ensemble du corpus et pour chaque catégorie des « doctorants » (D) et des « non-doctorants » (Nd).

Discipline	Code revue	Revues	Nombre de numéros	Nombre d'articles		
				Total	D	Nd
LET	BSA	Bulletin de la Société des Amis de Montaigne	5	22	5	17
	LCL	Littératures classiques	2	18	2	16
	CRM	Cahiers de recherches médiévales	2	17	0	17
	ROM	Romantismes	2	17	0	17
	RLC	Revue de Littérature comparée	4	17	0	17
	RMO	Revue de Montesquieu	2	7	0	7
	LIC	La Licorne	3	6	6	0
	RET	Recherches et Travaux	2	6	0	6
SCL	CDP	Cahiers de praxématique	5	23	3	20
	FDL	Faits de Langue	1	17	3	14
	LES	Langage et Société	4	13	5	8
	TIP	Travaux de l'Institut de Phonétique de Strasbourg	4	13	8	5
	RLV	Recherches linguistiques de Vincennes	2	10	0	10
	CFC	Cahiers du français contemporain	2	10	3	7
	LGG	Langages	5	9	0	9
	LIA	Liaisons-Airoe	1	8	0	8
	LFR	Langue Française	4	7	0	7

Figure 3.3. Représentation des revues dans le corpus de LET et de SCL en nombre de numéros et en nombre d'articles.

Le profil du corpus est tel qu'il ne respecte pas une distribution parfaitement équitable des articles en fonction des numéros et des revues. Celle-ci s'est avérée délicate en regard de la

¹ Ce comptage considère comme numéros différents les différentes livraisons d'une revue (soit en fonction du trimestre de publication), abstraction faite du mode éditorial de numérotation des livraisons ; ainsi a-t-on compté, en raison de la spécificité thématique de chaque livraison, 4 numéros de la *Revue de littérature comparée*, référencés 76-3, 76-4, 77-1 et 77-2 par la revue. Pour les livraisons désignés 27-28 ou 29-30 du *Bulletin de la Société de Montaigne*, nous avons affaire à des numéros doubles, qui font l'objet d'une même livraison et d'une unité thématique, et nous comptons donc un seul numéro.

disponibilité des textes, en raison de la taille variable des numéros d'une revue (en nombre d'articles notamment), et de la part non contrôlable de la pré-sélection opérée par les responsables éditoriaux.

Le corpus répond cependant à nos objectifs de réunir un nombre relativement important de textes, et d'éviter la surreprésentation de sous-domaines et de thématiques spécifiques, en privilégiant au contraire leur diversité. L'importance prise par le critère de disponibilité des textes montre quant à elle que le développement du document numérique ne doit pas faire illusion : si l'augmentation des possibilités de diffusion des textes et de stockage informatique représente un atout majeur, elle se heurte à un ensemble de difficultés dès lors qu'il ne s'agit plus de récolter du texte tout venant mais de constituer un corpus de textes en fonction d'objectifs déterminés, et en tenant compte des droits d'auteur.

3.1.4. Acquisition des données et pré-traitement du corpus

Pour disposer d'une version électronique du corpus, nécessaire à l'automatisation de son analyse, trois solutions d'acquisition des données sont offertes : numériser les textes par une saisie clavier ou en les scannant, ou récupérer les textes dans un format numérique. C'est cette dernière option que nous avons choisie, parce qu'il nous fallait tenir compte également des droits d'auteur et que s'adresser aux revues semblait donc recommandé. Les revues représentées dans le corpus le sont parce que leurs responsables ont bien voulu répondre à notre requête, faire confiance à nos engagements, et nous transmettre leurs archives. Nous tenons ici à les remercier pour la mise à disposition des textes et pour leur disponibilité à notre égard.

Une fois les textes recueillis, intervient le problème des formats de fichiers¹. Le format le mieux toléré par les analyseurs est le format texte brut, qui représente le format textuel le moins enrichi puisque le codage se limite aux caractères et aux tabulations. Cependant, l'absence d'autre enrichissement typographique et topographique (relatif à la mise en page) introduit un certain nombre de difficultés. Elle demande à confronter cette version en texte brut à une version originale du texte (version papier ou numérique) qui rende compte des italiques, des caractères étrangers, de symboles mathématiques, des tableaux et autres figures, de la hiérarchisation des titres lorsque celle-ci est signalée exclusivement à l'aide de la typographie, autrement dit de tout ce qui échappe au codage en texte brut mais qui peut être utile à l'analyse des textes.

La conversion des textes dans ce format complique également le nettoyage des textes, dans la mesure où les caractères non reconnus sont généralement traités à l'aide d'un autre caractère (par exemple, le point d'interrogation), ce qui là encore peut être préjudiciable à l'analyse. Plus largement, le nettoyage des textes demande à vérifier la conformité du texte avec la version effectivement publiée, et à supprimer tout élément susceptible de parasiter l'analyse. Ainsi, les en-têtes et pieds de page des articles publiés n'ont pas de sens spécifique dans l'étude mais peuvent corrompre les caractéristiques des textes et sont donc ôtés.

La question de savoir jusqu'où nettoyer les textes n'a rien d'évident ; par exemple, nous avons fait le choix de ne pas corriger d'éventuelles coquilles ou fautes d'orthographe. Finalement, il faut donc réunir les conditions requises par l'analyse automatique, tout en se référant à la version originale des textes lorsque son pré-traitement introduit des modifications importantes. Le problème se pose notamment au niveau du comptage des caractères et des

¹ La conversion des formats de fichiers a été rendue difficile lorsqu'il n'a pas été possible pour les responsables éditoriaux de nous fournir les textes dans un autre format que les formats propriétaires des logiciels d'édition. Le fait de s'adresser aux auteurs plutôt qu'aux responsables éditoriaux des revues est un moyen de contourner cette difficulté liée au mode d'archivage numérique choisi par les revues.

mots comme indice de la taille des textes, dès lors que le texte n'est pas fait que de mots (cf. chapitre 4).

La normalisation des textes consiste également à isoler certaines parties, qui pour faire partie intégrante du texte de l'article peuvent également corrompre les caractéristiques que l'on cherche à cerner. Les résumés, annexes et bibliographies ont ainsi été mis de côté en raison de la spécificité de leurs propriétés linguistiques et de leur statut péritextuel dans l'économie du texte. Nous conservons donc le corps de l'article (l'article à l'exclusion de ces trois éléments) et les notes. Les exemples et les citations (cf. chapitre 4) sont également isolés d'une version intitulée « Corps sans exemples ni citations » ; celle-ci correspond au corps de l'article une fois soustraits les exemples et les citations décrochés dans un paragraphe spécifique, et ceux qui sont intégrés au texte mais comptent plus de 20 mots. Nous avons également isolé les parties introductives et conclusives de l'article lorsque cela s'est avéré possible (cf. chapitre 4 pour la structure des textes et les difficultés posées par les introductions et les conclusions).

Au final, on obtient donc trois sous-corpus à utiliser pour l'analyse (automatique ou manuelle) :

- un sous-corpus « Corps », composé du corps de l'article (l'article à l'exclusion des résumés, annexes et bibliographies)
- un sous-corpus « Corps sans exemples ni citations » composé du corps sans les exemples ni les citations
- un corpus « Notes » qui ne réunit que les notes (de bas de page ou de fin d'article). La spécificité de cet espace du texte est prise en compte dans certains aspects de notre analyse, mais demanderait une étude à part entière.

Ces trois sous-corpus sont confrontés comme nous l'avons souligné à la version originale des textes afin de prendre en compte notamment les figures et tableaux, le lien entre le corps de l'article, les notes et la bibliographies, etc.

Le corpus est documenté mais ne fait pas l'objet d'un balisage logique. Sa structuration le rend échangeable, mais sa réutilisation dépend de la question des droits, initialement accordés au titre de la présente recherche. L'enrichissement du corpus à l'aide de balises logiques et d'une DTD (description du type de document) représente un prolongement requis de notre étude dans cette perspective d'une réutilisabilité du corpus.

Il demande à s'interroger en particulier sur les recommandations adoptées par la TEI¹. Les problèmes de format que nous avons évoqués ne peuvent que mettre en relief l'importance capitale de se doter de normes internationales pour l'édition et l'échange de documents numériques ; ces recommandations posent cependant un problème de lourdeur du protocole, et l'évaluation de leur pertinence met de nouveau au centre du débat sur les corpus la question de leurs objectifs. Nous verrons par exemple à propos de la structure des articles à quel point il est difficile d'envisager un modèle commun aux textes (chapitre 4).

Comme nous allons le voir, le corpus fait l'objet d'une analyse en partie automatisée et en partie manuelle. Plus coûteuse en temps, l'analyse manuelle répond cependant aux limites des assigneurs automatiques d'étiquettes linguistiques et à la complexité de certains phénomènes textuels. L'analyse manuelle est dans certains cas menée sur les 220 articles du corpus (c'est le cas par exemple pour la plupart des aspects relatifs à la structure des textes). Dans d'autres cas aussi, elle exige de restreindre le nombre de textes pris en compte pour pouvoir procéder à

¹ Text Encoding Initiative (<http://www.tei-c.org/>).

un examen plus fin des phénomènes. Nous utilisons alors des sous-corpus ; ceux-ci présentent des configurations variées en fonction des objectifs.

Nous utilisons ainsi, avec le corpus de 220 articles :

- un corpus de 44 articles de SCL pour l'analyse des références à l'existant dans l'article et la différence entre les textes de doctorants et de non-doctorants (partie 7.4.)
- un corpus de 35 articles de LET et de 35 articles de SCL pour l'analyse des exemples et des citations (partie 4.5.)
- un corpus de 20 articles de LET et de 20 articles de SCL pour la plupart des analyses énonciatives.

Ces différents corpus entretiennent tous un rapport d'inclusion depuis le corpus de 110 articles par discipline jusqu'au corpus de 20 articles par discipline.

Pour résumer, notre corpus d'articles est un corpus électronique constitué et structuré pour l'analyse du genre, comme nous l'avons indiqué en introduction de cette partie (3.1.). Une fois mises en évidence un certain nombre de caractéristiques en termes quantitatifs, il peut également être défini comme un réservoir d'exemples pour une observation plus qualitative des phénomènes. Ces deux pôles définitoires du corpus correspondent aussi à deux pôles dans l'analyse du genre, avec d'un côté l'automatisation et la quantification des phénomènes, de l'autre des études de cas sur des phénomènes plus fins, sans visée de généralisation ; nombre d'analyses se situent de manière intermédiaire entre ces deux pôles, avec notamment la quantification de phénomènes analysés manuellement, sur l'ensemble du corpus ou sur un corpus plus restreint.

Notre analyse du genre vise à articuler ces deux pôles, celui d'une analyse numérique quantitative et partiellement automatisée, et celui d'une analyse linguistique manuelle et qualitative, en les mobilisant de manière complémentaire. C'est vrai de l'ensemble de notre entreprise de caractérisation du genre, et c'est vrai aussi de l'analyse de la variation stylistique, qui a demandé la mise en place d'un protocole méthodologique spécifique, que nous traitons pour cette raison dans le chapitre dédié à cette étude (chapitre 6).

3.2. Une analyse quanti-qualitative partiellement automatisée

Compte tenu de la taille du corpus et de l'objectif de description des régularités et de la variabilité du genre, l'automatisation de l'analyse représente un atout majeur. Elle est produite à partir de l'assignation automatique d'étiquettes morpho-syntaxiques avec le logiciel Cordial comme nous le verrons ci-dessous. Cependant, un ensemble de phénomènes demanderait, pour être analysés automatiquement, à faire l'objet d'une annotation manuelle, en particulier les phénomènes énonciatifs qui intéressent notre étude. Une telle annotation repose sur une analyse manuelle mais représente une évolution possible de notre corpus : elle a ceci d'intéressant qu'elle peut permettre une fois le corpus enrichi de réutiliser les annotations (par exemple, pour des requêtes sur le corpus¹), à des fins de recherche ou d'exploitation didactique.

L'analyse manuelle peut ou non intégrer une composante quantitative, et nous le précisons pour chacun des traits finalement retenus dans l'étude du genre telle qu'elle est présentée dans les chapitres 4 à 7. L'analyse manuelle tente de proposer des typologies (par exemple, les guillemets au chapitre 5., les références à l'existant au chapitre 5), ou, à défaut, de proposer une description multi-critériée (par exemple, pour la structure des textes au chapitre 4) ; heuristique, la discrétisation introduite par l'entreprise typologique – à la base

¹ On en trouve un exemple en France avec la base catégorisée du corpus Frantext (base textuelle pour le français, <http://atilf.atilf.fr/frantext.htm>).

notamment des annotations manuelles envisageables à terme dans notre corpus – requiert cependant aussi de tenir compte aussi de la polysémie observable (par exemple, avec les guillemets, cf. chapitre 5. ou les pronoms *je* et *nous*, cf. chapitre 7.) ; dans ce dernier cas de figure, nous délaissions l’objectif de quantification.

Dans tous les cas, y compris suite à l’analyse automatique, nous entendons discuter les unités linguistiques caractéristiques du genre en regard de leur fonctionnement dans les textes, et associons à l’analyse des données quantitatives des analyses circonstanciées, par le recours à l’exemplification.

3.2.1. L’analyse automatique et quantitative

Le développement des étiqueteurs morpho-syntaxiques rend possible la mise en œuvre d’analyses automatisées, qui présentent un intérêt indéniable pour traiter de volumes importants de textes. Notre objectif est à la fois quantitatif et qualitatif, et impose une contrainte forte sur la phase automatisée de l’analyse : celle-ci doit viser l’économie ; l’objectif est de la compléter par une analyse manuelle des textes, qui permette de préciser les résultats obtenus par l’analyse quantitative, et de décrire quantitativement et/ou qualitativement des phénomènes que le logiciel ne peut prendre en charge.

Nous avons fait le choix d’utiliser le logiciel Cordial Professionnel (Cordial 9.0 Professionnel, 2003), développé par la société Synapse Développement (Toulouse). Celui-ci produit un étiquetage morpho-syntaxique, une analyse statistique de cet étiquetage, et une étude lexicale, stylistique et sémantique du texte ; le texte est analysé en regard du corpus de référence (corpus de plus de 500 millions de mots structuré en référence aux discours : juridique, littéraire, etc.). Cordial fournit d’une part des résultats que l’on pourrait qualifier de « clé en main », mais comporte également des fonctionnalités qui permettent à l’utilisateur de paramétrer l’étiquetage et de formuler des requêtes dans le texte.

L’objectif en recourant à ce logiciel est d’obtenir des informations morpho-syntaxiques et lexicales sur les textes. Il ne s’agit en aucun cas d’exploiter l’ensemble des fonctionnalités, car c’est principalement l’étiquetage automatique qui nous intéresse. Le choix de ce logiciel s’explique par son utilisation dans diverses analyses de textes et d’exploration de corpus (notamment Malrieu et Rastier, 2000, Beauvisage, 2001, Poudat, 2003, Poudat et Loiseau, 2005), qui poursuivent des objectifs en partie commun avec notre étude du genre de l’article. L’étiquetage est également utilisé en amont de traitements automatiques par différents outils (par exemple, Lexter). Il propose un grand nombre de descripteurs, qui dépassent les besoins de notre étude, et sa qualité d’annotation a été largement reconnue dans les campagnes d’évaluations internationales des étiqueteurs morpho-syntaxiques pour le français (par exemple l’action GRACE¹), qui font de Cordial professionnel un des meilleurs étiqueteurs automatiques du marché.

L’intérêt de ce logiciel au niveau de l’étiquetage tient à la performance de sa technologie, et au principe sur lequel elle repose. Cordial Professionnel intègre des fonctions d’analyse mais ses fonctions de base sont celles du correcteur orthographique et grammatical

¹ GRACE : Grammaires et Ressources pour les Analyseurs de Corpus et leur Evaluation. Pour des précisions sur la méthodologie d’évaluation et les critères de qualité de l’action GRACE, concernant l’assignation des parties du discours pour le français, voir l’article de G. Adda et al. (1999). Notons comme le souligne J. Véronis (2000) que c’est à la suite d’une interaction avec des équipes universitaires dans le cadre de l’action GRACE que la société *Synapse* a rendu disponible une version Universités dans laquelle le module grammatical est accessible et permet l’étiquetage de corpus, alors que ce module était au départ encapsulé dans un logiciel de correction orthographique et grammaticale qui ne permettait pas à l’utilisateur d’accéder aux paramètres de l’étiqueteur ni à ses résultats.

Cordial. A ce titre, le logiciel est robuste face aux distorsions du texte analysé par rapport aux normes de l'écrit¹. La technologie permet en particulier de contourner en partie les problèmes posés par le caractère multilingue très souvent à la source d'erreurs dans les analyseurs car les segments en langue étrangère ne sont pas reconnus comme tels. Il gère mieux également les mots inconnus de son dictionnaire, traités en tant que tels plutôt qu'analysés de manière erronée, mais aussi les éventuelles coquilles que comporterait un texte publié.

La technologie se fonde sur différents dictionnaires, sur des règles linguistiques explicites et sur des informations statistiques à partir desquelles le logiciel va calculer la probabilité pour qu'un terme, en fonction du contexte local, relève de telle ou telle catégorie. C'est donc du corpus de référence que dépendent les performances de l'analyse.

L'étiquetage automatique dans Cordial se caractérise par la reconnaissance formelle des unités grammaticales et lexicales, par l'indexation de l'unité à son lemme (par exemple, la forme masculin singulier pour les noms et adjectifs, l'infinitif pour les verbes, etc.), et par l'assignation d'étiquettes morpho-syntaxiques et sémantiques. Nous n'utilisons cependant pas les étiquettes sémantiques, en raison des difficultés qu'elle pose : si, comme le montre J. Véronis (2000), l'étiquetage grammatical est aujourd'hui une forme d'annotation automatique opérationnelle, c'est loin d'être le cas pour les annotations sémantiques (Véronis, 2004), qui posent par ailleurs la question de la pertinence des catégories retenues, résultant d'ontologies nécessairement discutables. L'annotation syntaxique se situe pour sa part sur un échelon intermédiaire : elle fonctionne bien pour certains éléments (comme l'identification de types de subordonnées, que nous exploitons dans notre étude), mais les arborescences sur la structure syntaxique des phrases restent périlleuses.

Pour donner un exemple de l'étiquetage morpho-syntaxique de Cordial, citons celui donné dans le manuel utilisateur, qui présente l'étiquetage de la phrase : « Le livre est sur la table » :

N°	Phrase	Lemme	Type grammatical	Fonction
===== DEBUT DE PHRASE =====				
1	Le	le	DETDMS	T
2	livre	livre	NCMS	T
3	est	être	VINDP3S	V
4	sur	sur	PREP	H
5	la	le	DETDMS	H
6	table	table	NCFS	H
7	.	.	PCTFORTE	
===== FIN DE PHRASE =====				

Figure 3.4. Exemple d'étiquetage d'une phrase dans Cordial Professionnel, cité dans le Manuel Utilisateur (Synapse Développement, 2002).

Selon les paramètres choisis ici, l'étiqueteur associe à chaque mot :

- son positionnement dans la phrase,
- sa forme originale dans le texte,
- sa forme normalisée ou lemmatisée,
- son type grammatical,
- sa fonction (T = « appartient au sujet », V = « verbe de base de la proposition », H = « circonstanciel »).

¹ C'est ce que soulignent A. Valli et J. Véronis (1999 :118) qui proposent de tester le logiciel pour l'étude de l'étiquetage grammatical des corpus de parole.

On soulignera d'abord la finesse du jeu d'étiquettes : comme le montre l'exemple donné, les étiquettes contiennent la catégorie grammaticale, mais accompagnée d'un certain nombre d'informations morpho-syntaxiques (genre, nombre, temps, personne, etc.), ce qui n'est pas le fait de tous les étiqueteurs¹. La fonction syntaxique, qui est sans doute le point sur lequel se posent le plus de problèmes, n'est mobilisée que dans quelques-unes des variables que nous avons adoptées dans notre étude, avec des étiquettes *a priori* peu problématiques (les pronoms personnels en fonction sujet).

Plus largement, les choix de variables que nous avons faits (nous les présentons plus loin dans ce chapitre, dans la partie 3.3.) visent à ne retenir que les fonctionnalités de l'étiqueteur qui ne posent pas de difficultés majeures : la fonction de segmentation du texte en phrases, l'identification du type de proposition, non signalée dans l'exemple présenté, la fonction de lemmatisation dans l'analyse du lexique et de la morpho-syntaxe, et l'assignation des types grammaticaux pour l'analyse des différentes variables morpho-syntaxiques, relatives notamment aux parties du discours, aux temps verbaux, aux pronoms personnels et à la ponctuation.

L'étiquetage permet alors au logiciel de fournir un comptage des différentes catégories identifiées (un comptage des formes, des lemmes, et des catégories). Les résultats de l'étiquetage (texte étiqueté, et comptages en valeurs absolues et/ou en pourcentage d'une catégorie par rapport à d'autres) sont accessibles dans différents fichiers en format texte brut. Cela permet d'évaluer l'étiquetage du texte, et également de réexploiter les comptages, *via* une base de données, dans différentes analyses, en particulier des analyses statistiques.

L'analyse automatique menée par Cordial comporte nécessairement des limites : tout étiquetage induit du bruit et du silence. Le bruit désigne les résultats non attendus par la requête (non pertinents), et le silence les résultats absents que la requête aurait du fournir ; cependant, au vu des étiquettes mobilisées par les variables que nous retenons (cf. partie 3.3.), les résultats ne doivent poser que quelques problèmes limités. Comme le souligne J. Véronis (2000), les étiqueteurs atteignent pour ce qui concerne les catégories grammaticales des résultats très satisfaisants, de l'ordre de 95% d'étiquettes correctes.

Si le bruit et le silence sont inévitables, ils sont donc largement minorés. L'étiquetage a à affronter le problème crucial de l'ambiguïté linguistique et si la technologie développée dans un analyseur a précisément pour but de viser la désambiguïsation, un certain nombre de faits linguistiques posent des difficultés telles que les analyseurs automatiques ne sont pas capables de les résoudre. C'est le cas par exemple du passif, avec une confusion fréquente entre un tour au passif et un tour au passé composé ou encore du subjonctif présent, confondu avec l'indicatif présent dans certains contextes.

Nous avons procédé à une évaluation de l'étiquetage, beaucoup plus modeste bien entendu dans ses objectifs et ses moyens que les évaluations qui entendent comparer les analyseurs disponibles sur le marché² ; il ne s'agissait pour nous que de vérifier la qualité de l'étiquetage dans le cas de l'article de recherche, et à partir des seules étiquettes retenues dans notre analyse (soit 97 variables, cf. partie 3.3.). Sur 50000 mots analysés (l'équivalent de 10 articles), on note que le bruit et le silence tiennent à ce type d'ambiguïtés problématiques pour toute analyse automatique. Les mots étrangers (fréquents dans les articles de Lettres où ils correspondent à de courts extraits du texte étudié) et les noms propres sont étiquetés comme mots inconnus, et il n'est donc pas possible d'envisager l'analyse automatique des noms

¹ J. Véronis (2000) y voit un critère déterminant pour la qualité d'un étiquetage grammatical, car le caractère opérationnel de la plupart des étiqueteurs de ce type masque d'importantes différences à ce niveau.

² Un des problèmes cruciaux posés par l'évaluation des étiqueteurs est en effet celui de la comparaison (Adda et al., 1999, Véronis, 2000).

propres, mais ce résultat prévisible nous avait incité à écarter cette variable de nos objectifs d'analyse avec Cordial.

Les problèmes d'ambiguïtés linguistiques et de limites des dictionnaires, propres à tout analyseur automatique peuvent être résolus par le recours à un étiqueteur entraîné : l'entraînement du logiciel sur des genres spécifiques rend possible la désambiguïsation, comme le montre C. Poudat (2006a), qui opte pour cette solution dans le cadre de son analyse du genre de l'article, en particulier à des fins d'évaluation des outils et d'analyse contrastive en français et en anglais. Cette solution particulièrement intéressante a été écartée de notre étude en raison de la place de l'analyse automatique en regard de nos objectifs, s'agissant d'analyser un certain nombre de traits structurels ou énonciatifs des textes, qui ne sauraient être traités automatiquement qu'une fois étiquetés manuellement dans les textes.

Une autre difficulté de Cordial tient aux catégories linguistiques utilisées. On relève notamment certaines fluctuations dans les choix dénominatifs, vraisemblablement à mettre au compte des versions successives du logiciel. Par exemple, pour la classe des déterminants, selon le menu du logiciel auquel on a affaire, les déterminants démonstratifs sont désignés « adjectifs » ou « déterminants ». Plus largement, ce problème est à mettre au compte d'un choix de catégories qui relèvent de la grammaire traditionnelle, de type scolaire¹. Ce choix s'explique par les applications de la technologie de l'analyseur, dans des logiciels commerciaux de correction orthographique et grammaticale : les étiquettes doivent être comprises par le grand public.

Cela étant, nous avons dit aussi que le texte analysé peut être comparé à un corpus de référence structuré en fonction de discours². Ce parti pris nous semble particulièrement intéressant, au moins à titre de perspective de développement de versions ultérieures : plutôt que de viser la langue générale, les analyseurs gagneraient à tenir compte de la spécificité des discours et des genres pour affiner le travail de désambiguïsation ; comme le souligne D. Biber (1993), les sorties varient en fonction des types de textes.

Le problème le plus délicat posé par le logiciel Cordial Professionnel est l'absence de transparence sur la technologie utilisée et ses fondements conceptuels en linguistique. Un tel constat n'a rien de surprenant, s'agissant d'un logiciel commercialisé, mais il est certain que cela peut à terme faire obstacle à l'exploitation du logiciel à des fins de recherche. Le problème se pose plus particulièrement pour l'ontologie à la base de catégorisations sémantiques, et c'est pour cette raison notamment que nous n'avons pas mobilisé les annotations de ce type.

Au final, il ne s'agit donc pas de nier les limites de l'outil ni son intérêt, mais de s'interroger sur ce qu'il est possible de demander à un tel outil de faire. L'analyse lexicale et morpho-syntaxique, pourvu qu'elle reste prudente face à certaines catégories qu'aucun outil n'est pour l'heure capable d'analyser de manière entièrement automatique, reste intéressante pour fournir des informations sur un nombre important de textes. Relayée par des procédures de validation statistique et par un examen des occurrences en corpus, elle n'a pour but que de faciliter une première approche d'un corpus volumineux, et ne retient que quelques variables pour lesquelles les résultats de l'étiquetage automatique sont bien éprouvés dans les travaux existants – y compris dans les évaluations faites de ce type d'outils – et s'avèrent satisfaisants.

Bien entendu, cette analyse automatique demande un travail important de préparation des textes, pour éviter que des données parasites ne corrompent l'analyse. Elle se prolonge par

¹ Ce problème de la pertinence linguistique des catégories est vrai de l'ensemble des analyseurs grammaticaux selon J. Véronis (2000).

² C'est vraisemblablement là que se manifeste l'influence de F. Rastier, que le Manuel Utilisateur (2002, Synapse Développement) cite comme principale référence dans la conceptualisation linguistique nécessaire aux développements du logiciel.

des analyses manuelles, qui pourraient à leur tour, lorsqu'une entreprise typologique est envisageable¹, permettre d'enrichir le corpus par de nouvelles annotations. Si l'assignation d'étiquettes grammaticales est donc automatique, on voit finalement comme le soulignent B. Habert et al. (1997) qu'il est difficile de maintenir une distinction entre, d'un côté, une procédure d'analyse automatique et de l'autre une procédure d'analyse manuelle.

L'automatisation, en tant que perspective nouvelle offerte en linguistique, est à envisager sur un continuum de possibilités entre l'annotation purement manuelle et l'automatisation complète (Véronis, 2000). Sur ce continuum, notre étude n'exploite qu'une partie des possibilités d'annotation automatique, celles qui à l'heure actuelle peuvent être reconnues comme opérationnelles ; les analyses manuelles pourraient quant à elles trouver des prolongements dans l'annotation des textes, et les possibilités permises par ces annotations, en termes de requêtes et d'extraction d'informations du corpus². Dans une perspective didactique en particulier, l'utilisation de corpus annotés pour l'observation, la manipulation et la génération automatique d'exercices représente une voie particulièrement prometteuse.

3.2.2. *Le recours aux statistiques*

La quantification des traits textuels repose en partie sur l'étiquetage automatique de Cordial. Dans d'autres cas cependant, elle a été opérée manuellement, selon des modalités diverses (soit il s'agit d'une reconnaissance de certaines unités formelles, par exemple les tableaux et figures que comprend un article, soit à cette première phase s'associe la construction d'une typologie, et un retour sur le corpus pour l'analyse de la représentation de chaque catégorie).

Qu'elle ait reposé sur l'analyse automatique ou manuelle, la quantification s'est presque systématiquement associée à un traitement statistique des données ; c'est vrai de toutes les quantifications, sauf celles concernant le *il* impersonnel : les objectifs de l'étude sont de pointer certaines récurrences dans les constructions du *il* impersonnel, mais la fréquence de chaque type de concordances rend caduque le recours aux statistiques.

Notre choix pour le traitement statistique s'est porté sur les outils de statistiques générales, plutôt que sur des outils dédiés aux statistiques textuelles comme c'est le cas du logiciel DTM (Data and Text Mining), développé par L. Lebart, et que nous avons utilisé pour les besoins spécifiques de l'étude sur les styles (chapitre 6), en particulier pour les méthodes d'analyse multidimensionnelle des textes et les méthodes de classification. Nous y avons toutefois adjoint des procédures de validation avec des traitements statistiques non intégrés à ce logiciel.

Le logiciel que nous avons utilisé est le logiciel Statview, version 5.0., développé par SAS Institute Inc. Ce choix s'explique par l'intérêt de l'outil, mobilisé dans nombre de domaines de recherches, mais aussi par sa disponibilité et le fait que nous ayons reçu une formation spécifique à ce logiciel. Ses fonctionnalités dépassent les besoins requis dans le cadre de la présente étude ; la performance de l'analyse et la diversité des méthodes et des paramétrages proposés à l'utilisateur représentent à notre sens les principaux avantages, mais les sorties de l'analyse montre que si Statview est un outil de travail très intéressant, il représente un outil de visualisation très rudimentaire. Nous avons donc importé les sorties de Statview dans le logiciel Excel (Microsoft, 2002) pour rendre compte des analyses au moyen de graphiques.

¹ Nous le verrons par exemple avec les différents types de passages entre guillemets au chapitre 5, ou avec la distinction entre *il* impersonnel et *il* anaphorique au chapitre 7, ou à propos des différents modes de références à l'existant dans ce même chapitre.

² Notre étude sur les guillemets (chapitre 5) s'inscrit dans un projet de ce type (cf. Rinck, Tutin, Hermann, 2004).

Principes de l'analyse statistique

La production de statistiques répond à un besoin d'analyse de données transversal aux différents domaines disciplinaires. L'utilisation de la statistique demande cependant à être un usager formé¹ et conscient : si les logiciels existants sur le marché sont généralement bien documentés et interfacés pour faciliter les procédures², il n'en reste pas moins que l'analyse statistique de données requiert un travail de conceptualisation spécifique. Le choix de méthodes statistiques et l'interprétation des résultats reposent par ailleurs sur des conditions strictes, dont on ne saurait s'affranchir. Si l'usage de la statistique est aujourd'hui très répandu en sciences humaines et sociales³, il n'est pas en lui-même garant de la pertinence des analyses et montre l'intérêt qu'il y aurait à systématiser l'évaluation de cet usage dans la littérature scientifique⁴.

L'analyse de données que requiert notre étude du genre de l'article mobilise deux aspects des statistiques, d'une part les statistiques descriptives, d'autre part les statistiques inférentielles. Ces deux visées de la statistique, la description et l'induction, s'associent à des procédures de visualisation (représentations graphiques) qui demandent là encore à faire l'objet d'un usage averti. Se pose aussi, de manière transversale, la question de l'interprétation des résultats.

Point de départ

Les statistiques consistent comme nous l'avons dit à traiter des données. Le point de départ de l'analyse statistique repose sur l'échantillon, qui ne renvoie non pas aux individus mais aux données sur ces individus, données supposées représentatives de celles de la population parente (i.e. l'ensemble des mesures possibles, compte tenu de l'ensemble des variables incontrôlables) ; dans notre cas, l'échantillon représente l'ensemble des données sur chaque article du corpus, et la population est le genre de l'article en sciences du langage et en lettres.

Chaque individu ou chaque article est décrit au sein de la base de données par des variables correspondant aux caractéristiques des textes (par exemple, le nombre de mots, la présence ou non d'une bibliographie, etc.), et par des variables « externes » ou indépendantes (la discipline, le statut de l'auteur, la revue).

Les données mettent en correspondance un nombre à l'objet observé, autrement dit fournissent une mesure. La théorie de la mesure développée par S. Stevens (cf. par exemple 1959) déconstruit l'idée que mesurer c'est compter, pour montrer que les données numériques obéissent à différents types de règles de constitution, en fonction de différentes échelles de mesure. Sans entrer dans le détail de ces échelles de mesures, soulignons cependant qu'elles

¹ Une évaluation des méthodes statistiques dans les thèses de médecine menée par J.L. Bosson et al. (1995) montre les problèmes posés par des lacunes au niveau de la formation des étudiants. Un des constats les plus frappants de l'étude est que près de 11% des thèses (16 sur les 140 analysées) concluent à des différences significatives entre groupes de patients sans utilisation de tests statistiques. On peut y voir sans doute aussi le problème posé par une utilisation médiatique des statistiques qui, sans être nécessairement biaisée, n'est du moins jamais explicitée ni questionnée.

² De sorte qu'il n'est parfois pas même nécessaire de disposer des connaissances mathématiques pointues qu'impliquent les statistiques. Ces connaissances aident cependant à conscientiser la pratique des statistiques et à l'envisager comme étant toujours perfectible.

³ Au point de devenir peut-être, dans un certain nombre de champs, une dimension attendue des thèses de doctorat.

⁴ L'article de J.L. Bosson et al. (1995) note que cette évaluation de l'usage des statistiques est de pratique courante dans la littérature médicale, et cite un article de 1991 qui recense plus de 150 articles sur ce sujet. A notre connaissance, une telle entreprise n'a pas été menée dans le domaine des sciences du langage.

déterminent la nature des variables, selon la manière de caractériser les objets observés par des nombres.

Les variables textuelles et indépendantes dans le cadre de notre analyse peuvent être de deux types, les variables nominales et les variables continues, reposant sur une échelle de rapports.

- Avec les variables nominales, la mesure consiste à assigner une étiquette catégorielle aux objets : par exemple, pour la variable « présence ou absence de bibliographie », on a deux modalités, « présence » et « absence » ; de même, pour la variable « discipline », on a deux modalités « LET » et « SCL ». Y compris quand le codage est numérique (par exemple, 0 et 1 pour absence et présence d'une bibliographie), le chiffre a une valeur nominale et n'introduit pas d'idée d'ordre entre les catégories.

- Avec l'échelle de rapports, les variables continues sont telles que l'ordre a un sens, que les intervalles entre deux échelons sont équivalents (on parle d'unité de mesure) et que le zéro n'est pas arbitraire, mais correspond à l'absence de l'élément mesuré. C'est le cas par exemple pour le nombre de mots du texte.

Statistiques descriptives

La base de données est donc constituée de telle sorte que la nature des variables soit systématiquement définie. Il est alors possible de leur appliquer les procédures statistiques. Comme le soulignent par exemple les ouvrages de statistique de D.C. Howell (1998) et de H. Rouanet et al. (1987), deux types de procédures peuvent être mises en œuvre, selon la visée de l'analyse statistique : les statistiques descriptives et les statistiques inférentielles.

La statistique descriptive a pour objet, comme son nom l'indique, de décrire des données. En d'autres termes, il s'agit de décrire la distribution d'une variable. Elle se prête donc bien à la visée descriptive qui est la nôtre, s'agissant de caractériser le genre de l'article : elle répond à l'objectif de décrire de manière exhaustive l'ensemble des individus du corpus à partir de l'ensemble des variables retenues.

Les descriptions fournies dépendent de la nature des variables examinées, ou des échelles de mesure considérées. Décrire la distribution d'une variable nominale consiste à calculer l'effectif de chacune des modalités (par exemple, le nombre d'article de LET et le nombre d'articles de SCL, le nombre d'articles contenant une bibliographie et le nombre d'articles n'en contenant pas, etc.). Concernant les variables continues, leur distribution est décrite en termes de tendance centrale et de dispersion des valeurs. La tendance centrale correspond à la valeur qui représente au mieux l'ensemble des valeurs, tandis que les indices de dispersion vont permettre de décrire comment les valeurs s'écartent de cette tendance centrale.

La tendance centrale peut être appréhendée au moyen de trois indices : la moyenne (la somme des valeurs divisée par leur nombre), la médiane (la valeur telle que 50% des valeurs sont au-dessus et 50% au-dessous) et le mode (valeur pour laquelle l'effectif est le plus grand). La dispersion d'une mesure correspond à la variabilité des valeurs. Plusieurs indices peuvent permettre de l'estimer : l'étendue (la différence entre la valeur maximale et la valeur minimale), la variance (la moyenne du carré des écarts à la moyenne) et l'écart-type, appelé aussi déviation standard (racine carrée de la variance) qui évaluent le degré auquel les valeurs sont resserrées autour de la moyenne.

Cette première phase de description statistique des données permet de caractériser le genre à partir de chacune des variables considérées : que l'analyse soit menée dans l'ensemble du corpus, ou dans chacun des deux groupes disciplinaires (articles de LET et articles de SCL), les effectifs (variables nominales) et les indices de tendance centrale et de dispersion (variables continues) déterminent les caractéristiques des textes ; elles rendent compte à la

fois des traits les plus représentatifs (l'effectif le plus grand, la tendance centrale) et de la diversité interne du genre de l'article.

Statistiques inférentielles

Cette visée descriptive demande à s'associer dans le cadre de notre étude à une visée inductive de l'analyse statistique : à travers les données sur le corpus, on cherche à analyser le genre de l'article et à généraliser les conclusions portées par les procédures descriptives. Cette visée inductive caractérise le domaine des statistiques inférentielles. Il s'agit d'évaluer le risque pris en affirmant que la mesure statistique calculée sur l'échantillon de données (le corpus) reste vraie pour l'ensemble de la population (le genre) ; le principe qui permet d'inférer la validité ou à l'inverse le risque de cette affirmation consiste à déterminer à quel degré la mesure faite sur l'échantillon s'écarte de ce que le hasard laisse prévoir.

Nous mobilisons les statistiques inférentielles de diverses manières. Il peut s'agir d'abord de tester des corrélations de traits textuels, autrement dit de voir en quoi le comportement d'une variable peut prédire celui d'une autre dans les textes. Nous procédons à des tests de corrélations soit de manière empirique, soit à partir d'une hypothèse que suggèrent les données et/ou la littérature existante en linguistique (c'est ainsi que nous testerons, par exemple, les corrélations entre les temps du passé, en tant que co-indices d'un mode narratif au chapitre 5). D'autre part, les statistiques inférentielles nous permettent également de tester des hypothèses de facteurs de variation du genre, i.e. les variables indépendantes que sont la discipline et le statut de l'auteur. Il s'agit de comparer l'effectif (variable nominale) ou la distribution des valeurs (variable continue) d'une variable textuelle dans les groupes (par exemple, les articles de LET *versus* ceux de SCL), et de voir dans quelle mesure l'appartenance du texte à un groupe prédit ses caractéristiques. Ces tests sont réalisés de manière systématique, contrairement à ceux de corrélations de variables.

Les tests, mis en œuvre avec le logiciel Statview dépendent à la fois de la nature des variables considérées (variables nominales ou continues), du nombre de modalités dans le cas de variables indépendantes (par exemple, deux pour la variable « discipline » : LET *versus* SCL), et de la distribution de la variable dans le cas des variables linguistiques. Ils répondent donc à des conditions strictes, mais nous avons également tenu à faire valider nos choix en demandant conseil à J.P. Chevrot, professeur au laboratoire Lidilem, que nous tenons à remercier ici. Les tests retenus sont présentés dans les paragraphes suivants, et de manière récapitulative sous la forme d'un tableau en fin de section.

Lorsque l'on a affaire à deux variables continues (les variables linguistiques), et qu'on cherche à analyser leur corrélation, nous utilisons parmi les tests proposés dans Statview la matrice de corrélation. Elle produit un résultat sous forme de tableau, qui met en correspondance les variables unes à unes en indiquant un taux de corrélation compris entre 0 et 1 (soit entre 0 et 100%). En conformité avec la plupart des analyses, et avec le manuel de référence du logiciel, nous admettons une corrélation lorsque le résultat dépasse un taux de corrélation de 0,75 entre les deux variables. La corrélation peut être positive ou négative (les variables continues sont inversement proportionnelles). C'est dans ce cas de figure que se situe l'analyse des corrélations entre temps verbaux que nous avons évoquées ci-dessus.

Pour déterminer des corrélations entre deux variables nominales, on recourt à un test dit de contingence, le test de Chi². Il consiste à mesurer à partir du tableau des effectifs des deux variables si la répartition observée dans l'échantillon diffère d'une répartition aléatoire. La probabilité p(Chi²), ou probabilité pour que la répartition soit due au hasard de l'échantillonnage (les données produites sur le corpus), est calculée à partir de la production

au hasard de tableaux d'effectifs. Nous nous référons aux modèles en vigueur en sciences humaines, confirmés dans le Manuel de Référence de Statview¹, pour admettre la significativité du résultat lorsque p est inférieur ou égal à 0,05 (soit 5%). C'est ce test que nous avons utilisé pour analyser les corrélations entre différents indices relatifs à la structure générale de l'article, au chapitre 4. : la présence ou non d'annexes, de formalismes, de titres de parties, etc.).

Le test de Chi2 concerne toute configuration où l'on cherche à déterminer en quoi la distribution d'une variable nominale peut permettre de prédire la distribution d'une autre variable nominale. Il intéresse donc non seulement les corrélations de traits textuels, mais aussi l'analyse de l'effet des variables indépendantes (i.e. l'effet de la variable « discipline » et de la variable « statut de l'auteur ») sur une variable nominale du texte (par exemple, la présence d'une bibliographie ou non). Les mêmes conditions d'acceptabilité du test sont utilisées : on conclut à un effet significatif de la variable indépendante sur la variable nominale lorsque p est inférieur ou égal à 0,05.

Pour analyser l'effet des variables indépendantes « discipline » et « statut de l'auteur » sur les variables continues, deux tests sont mobilisés, selon la distribution de la variable continue dans les groupes que compare le test (LET *versus* SCL, D *versus* Nd) : l'Anova (test paramétrique) et Mann-Whitney (test non paramétrique). Ce sont les mêmes qui servent à analyser le lien entre deux séries de données quand on a une variable nominale et une variable continue (par exemple, dans le chapitre 7, le nombre de références dans un texte, et l'annonce d'un enjeu d'opposition dans l'introduction de l'article).

Les deux tests sont tous deux utiles à l'analyse de l'effet d'une variable nominale à deux modalités sur une variable continue, mais répondent à deux conditions inverses en fonction de la distribution de la variable continue dans les deux catégories analysées (par exemple, annonce ou non d'un enjeu d'opposition dans l'article, article de LET ou de SCL, etc.). L'analyse de la variance (ou Anova) requiert qu'on ait affaire à une variance homogène, ou à des écarts-types sensiblement égaux dans les deux catégories : ces écarts-types doivent se situer dans un rapport de 1 à 2. Lorsque cette condition n'est pas remplie, cela signifie que l'on a affaire à des mesures non paramétriques (i.e. qui ne renvoient pas, par supposition statistique, à la même population), et il convient alors de se tourner vers les tests spécifiquement dédiés à ce type de comparaison.

L'Anova désigne une analyse de la variance, qui s'applique dans le cas de deux distributions de variance homogène (test paramétrique)². Elle consiste à comparer deux distributions, c'est-à-dire à la fois la tendance centrale et la dispersion des valeurs qui caractérisent chacune de ces deux distributions. Elle ne se limite donc pas à analyser les valeurs moyennes dans chaque groupe, mais part d'un principe de variabilité dans tout groupe. Cette variabilité interne à tout groupe est comprise comme l'effet de facteurs aléatoires. La variabilité entre deux groupes peut quant à elle résulter de deux types de facteurs : des facteurs aléatoires, mais également le facteur de différenciation des deux groupes.

Le rapport F de Fischer calcule le rapport entre la variabilité entre groupes et la variabilité interne au groupe ; en d'autres termes, il cherche à évaluer si la variabilité entre les groupes ne repose que sur des facteurs aléatoires, comme la variabilité au sein d'un groupe (le rapport F est proche de 1), ou si cette variabilité entre groupes est telle que les facteurs aléatoires ne sont pas seuls en jeu (le rapport F est supérieur à 1), et qu'elle doit s'expliquer par le facteur différenciant les deux groupes (i.e. le facteur dont on cherche à évaluer l'effet). L'Anova associée au rapport F de Fisher a une probabilité pour que la valeur de F ainsi calculée

¹ SAS Institute Inc., 1998.

² Cf. Le manuel de référence et l'addendum pour la version française (SAS Institute Inc., 1998).

soit le fait du hasard et le test peut conclure à un effet significatif de la variable indépendante lorsque p est inférieur ou égal à 0,05.

Quand les conditions des tests paramétriques (dans notre cas, l'analyse de la variance) ne sont pas remplies, il faut procéder à un test non paramétrique : comme les mesures sont de variance non homogène, la question est de savoir si les deux groupes indépendants (par exemple, les articles de LET et les articles de SCL) relèvent ou non de la même population parente.

Le test U de Mann Whitney permet de tester cette question d'homogénéité de deux échantillons ; pour ce faire, il transforme les données de mesure d'intervalles en mesures ordinales : l'ordre entre les valeurs a toujours un sens (une valeur peut être considérée comme inférieure ou supérieure à une autre), mais les intervalles n'en ont pas car il n'y a pas d'unité de mesure (l'exemple le plus parlant est celui des rangs d'arrivée à une course). Le test réunit les échantillons et ordonne les valeurs pour déterminer le rang de chacune. Puis il procède pour chaque échantillon à la somme des rangs et construit une statistique U pour chaque échantillon, qui permet de comparer la variance de la distribution dans chaque groupe. L'hypothèse selon laquelle les deux groupes relèvent de la même population est valide si la valeur p associée à U est inférieure ou égale à 0,05.

Un dernier test est utilisé dans le cadre de notre étude, de manière toutefois très ponctuelle, pour tester l'effet de la variable « revue » sur des variables du texte (par exemple, sa taille en nombre de mots). Il s'agit d'un test non paramétrique, qui repose sur un principe similaire au test U de Mann Whitney, mais s'applique aux cas de figure où la distribution de la variable continue doit être comparée dans plus de deux groupes (i.e. en fonction d'une variable à plus de deux modalités). Ce n'est ni le cas pour la variable « discipline » (LET et SCL) et pour la variable « statut de l'auteur » (D et Nd), mais c'est celui des revues (9 revues en SCL et 8 revues en LET). Nous recourons au test de Kruskal Wallis, et considérons qu'il valide l'hypothèse d'un effet non nul du facteur de différenciation des groupes (l'effet de la revue sur le genre) lorsque la probabilité p pour que l'effet mis en évidence soit le fait du hasard est inférieure ou égale à 0,05.

En résumé, les procédures statistiques reposent sur les variables suivantes :

Variables linguistiques :	nominales (par ex. <i>présence ou non d'une bibliographie</i>) continues (par ex. <i>taux de verbes au présent pour l'ensemble des verbes</i>)
Variables indépendantes :	discipline statut de l'auteur statut de l'auteur dans chaque discipline revues

Figure 3.5. Modélisation des variables.

Les différents tests statistiques mis en œuvre sont représentés dans le tableau suivant, qui présente les statistiques descriptives puis les statistiques inférentielles, à partir des conditions de réalisation des tests (colonne de gauche) et la description des tests (à droite, avec, pour les statistiques inférentielles, leur condition de validité (ou condition de validation d'un lien entre les séries de données)).

statistiques descriptives	Description de la distribution d'une variable			
	Variable nominale (indépendante ou linguistique)	Effectifs par catégorie		
	Variable continue (linguistique)	Indices de tendance centrale (moyenne, médiane, mode) et indices de dispersion (étendue, écart-type)		
statistiques inférentielles	Corrélations de variables linguistiques			validation d'un lien entre les séries de données
	Corrélations de variables continues	Matrice de corrélation	taux entre (-)0,75 et (-)1	
	Corrélations de variables nominales	Test de Chi2	$p(\text{Chi}2) < 0,05$	
	Effet des variables indépendantes			
	Sur une variable nominale	Test de Chi2	$p(\text{Chi}2) < 0,05$	
	Sur une variable continue			
	1. Variable indépendante à deux variantes			
1.1. variance homogène entre les deux séries de données	ANOVA (F de Fischer)	$p < 0,05$		
1.2. variance non homogène	U de Mann Whitney	$p < 0,05$		
2. Variable indépendante à plus de deux variantes (revue)	Kruskall Wallis	$p < 0,05$		

Figure 3.6. Procédures statistiques mises en œuvre.

Nous présentons en annexes les données sur lesquelles s'appuie l'analyse du genre. Elles sont présentées en deux temps, qui correspondent aux deux volets des procédures statistiques : les statistiques descriptives d'une part, et les tests de statistiques inférentielles d'autre part. Ceux-ci recouvrent comme nous l'avons dit les corrélations de traits textuels, et l'analyse des effets des variables indépendantes (ou variables groupes) sur les textes.

Pour chaque variable du texte, nous testons systématiquement l'effet de la discipline et du statut de l'auteur, et dans quelques rares cas l'effet de la revue. Par hypothèse, la discipline influence davantage les caractéristiques des textes que le statut de l'auteur, mais la spécificité des pratiques des doctorants peut être vraie au sein de chaque discipline ou indépendamment de la discipline (i.e. les deux disciplines confondues, les données sont telles qu'elles signalent que le statut de l'auteur détermine la différence de distribution).

Nous testons donc, systématiquement :

- l'effet de la discipline,
- l'effet du statut de l'auteur,
- et l'effet du statut de l'auteur dans les articles de LET puis dans les articles de SCL, ce qui permet d'affiner l'analyse et de mieux cerner la diversité du genre.

Nous verrons comment les différents résultats peuvent être mis en relation les uns aux autres, ou, parce qu'ils sont trop isolés, rester difficilement interprétables. Dans les chapitres 4 à 7, dédiés à l'étude du genre, nous présentons les résultats de l'analyse statistique de manière résumée, en nommant le test, en donnant la valeur produite et la valeur p (par exemple : *Anova*, $F=5,674$, $p < 0,001$) ; les résultats détaillés sont présentés en annexes.

L'interprétation des résultats de l'analyse statistique

L'interprétation des résultats de l'analyse statistique représente un de ses problèmes majeurs, avec la formalisation des hypothèses (de telle sorte qu'elles se prêtent aux statistiques), et le choix des méthodes (en fonction de l'objectif et du type de variable et de distribution auxquelles on a affaire). D'abord, il n'est pas possible de conclure à des différences entre groupes sur la base des seules statistiques descriptives, puisque c'est là l'objet propre des statistiques inférentielles. Des différences constatées au niveau de la moyenne, de la médiane, de l'écart-type ou de l'étendue d'une distribution ne peuvent rien indiquer d'autre en elles-mêmes que des tendances, qu'il faut alors spécifier par l'analyse inférentielle.

Menée de manière empirique ou en fonction d'hypothèses, cette dernière produit un taux de corrélation ou une valeur p qui nous sert d'indice de la significativité de l'effet de la

variable indépendante (i.e. le facteur différenciant les deux groupes de textes que l'on cherche à comparer). Pour la valeur p qui concerne la majorité de nos analyses inférentielles, nous avons dit que nous admettions la validation de l'hypothèse non-nulle (autrement dit l'hypothèse d'un effet significatif) quand p (ou probabilité pour que cette validité ne soit que le fait du hasard) est inférieur à 0,05. Si c'est cette valeur qui sert de critère en sciences humaines, il reste cependant discutable, comme le soulignent par exemple J.L. Bosson et al. (1995), d'invalider la significativité d'un résultat à $p=0,051$ et de la valider à $p=0,049$.

Comme pour le taux de corrélation, il s'agit en fait de considérer que le risque de se tromper en affirmant que ce qui est vrai pour l'échantillon de données le reste pour le genre existe toujours : en matière de textes, l'effet que l'on dit aléatoire à défaut de pouvoir mieux le cerner est bien entendu très fort. Ainsi, la comparaison des résultats produits dans l'étude de C. Poudat (2006a) et des nôtres montre les limites des tests statistiques : alors qu'ils doivent permettre de généraliser les résultats, la significativité de l'effet d'une variable sur une autre a pu être vérifiée dans un corpus mais pas dans l'autre. C'est le cas notamment de la corrélation entre le poids des exemples et le poids des citations, comme nous le verrons au chapitre 4. Alors que nos deux corpus reposent sur des critères de constitution comparables, pour ce qui est de l'article en sciences du langage il faut convenir que les revues et les thématiques qui y sont représentées ne sont sans doute pas sans effet sur les analyses du genre.

Plus largement, s'il n'est pas extrêmement compliqué de se doter de procédures de validation statistiques (statistiques inférentielles), pour généraliser les analyses descriptives, c'est toujours la question de savoir ce que l'on cherche qui resurgit. Les tests inférentiels peuvent produire des résultats significatifs qui posent d'importantes difficultés d'interprétation et il reste à déterminer si de tels résultats sont des résultats intéressants et en quoi. Il peut se poser un problème de ce type avec des résultats relativement locaux, qui concernent une variable mais ne peuvent être rapportés à d'autres différences observées dans les textes. Mais la difficulté peut intervenir aussi dans le cas où plusieurs résultats s'appuient les uns les autres.

Pour prendre un exemple concret, nous avons observé des différences significatives entre les articles de LET et de SCL d'une part, de D et Nd d'autre part, sur les parties du discours dans le cas des mots signifiants, autrement dit les substantifs, adjectifs, verbes et adverbes. Nous obtenons deux axes nets d'opposition entre d'une part les substantifs et les verbes, d'autre part les adjectifs et les adverbes, avec une différence disciplinaire sur le premier axe et une différence liée au statut de l'auteur sur le second. Cela étant, ces différences, quoique très marquées, nous semblent difficilement interprétables. Le problème relève en ce cas des catégories linguistiques elles-mêmes : l'entrée morpho-syntaxique ne suffit pas à elle seule à spécifier la fonction de ces unités (substantifs, verbes, adjectifs, adverbes), qui représentent des catégories très hétérogènes (c'est tout particulièrement le cas pour les adverbes).

Au final, si l'étude statistique doit permettre de pointer la pertinence de certaines variables pour la caractérisation du genre, nous délaissions parfois certains résultats trop problématiques, pour nous centrer dans nos analyses sur ceux qui offrent une prise véritable sur les textes : il s'agit de fonder l'analyse sur la matérialité des textes, mais de s'intéresser à des phénomènes qui disent quelque chose des enjeux du genre de l'article, en particulier au niveau du rapport à l'objet et au lecteur, de la manifestation de l'auteur dans son texte, ou encore de l'activité de recherche telle qu'elle est représentée dans les textes.

Au terme de cet aperçu des méthodes statistiques mises en œuvre, et de la question de l'interprétation des résultats, et avant de revenir sur la démarche générale d'analyse, que nous qualifions d'analyse quanti-qualitative, il nous reste à commenter un point, qui concerne la

présentation des résultats : les visualisations graphiques, qui doivent permettre d'aider à voir ce que les analyses statistiques permettent de dire.

3.2.3. Visualisations graphiques

Dans quelques cas, nous insérons des tableaux de résultats dans nos analyses. Ce cas de figure reste cependant très rare ; comme nous fournissons systématiquement en annexes les résultats des tests de statistiques descriptives et inférentielles que nous commentons dans nos analyses, nous préférons en général recourir à des visualisations graphiques plutôt qu'à des tableaux, pour rendre les résultats plus lisibles.

Un même mode de représentation peut servir des objectifs très différents, et pour un même objectif, divers modes de représentation sont possibles, mais de tels choix ne sont pas anodins : les graphiques doivent aider à voir ce que disent les données. Nous ne commenterons pas ici l'ensemble des figures utilisées, mais les deux modes de représentations principaux en regard de nos objectifs, à savoir les histogrammes pour la distribution de variables nominales d'une part, et les distributions en fréquence des valeurs d'une variable continue d'autre part.

Des histogrammes pour la distribution des variables nominales

Les histogrammes consistent, sous leur forme leur plus élémentaire, à associer pour une variable donnée une valeur (axe des ordonnées) à une catégorie (axe des abscisses) ; comme le montre l'exemple ci-dessous, un histogramme permet donc une comparaison des valeurs entre des catégories.

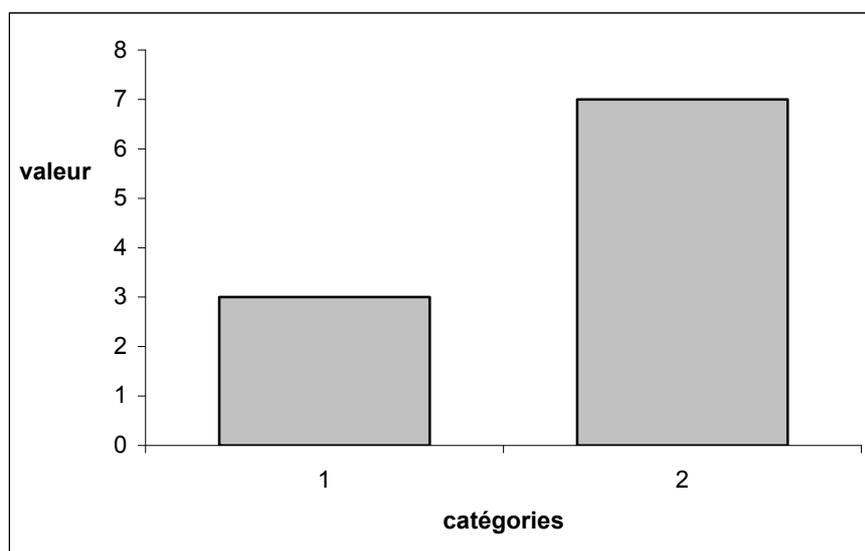


Figure 3.7. Exemple d'histogramme dans sa forme élémentaire.

Nous utilisons cette fonction de comparaison offerte par les histogrammes lorsque nous avons affaire à une variable nominale, pour comparer la distribution de cette variable dans deux catégories (par exemple, les articles de LET ou de SCL). Par exemple, nous comparons la présence dans les articles de LET et de SCL d'une bibliographie en recourant à un histogramme qui indique pour les deux catégories LET et SCL l'effectif des deux variantes de la variable linguistique (i.e. le nombre d'article comprenant une bibliographie et le nombre d'articles n'en comprenant pas). Par rapport à l'histogramme donné pour exemple, on a deux

catégories (LET et SCL), et deux séries de données (plutôt qu'une seule dans l'exemple) : les valeurs sont l'effectif d'articles sans bibliographie et l'effectif d'articles avec bibliographie.

La comparaison d'une variable continue entre deux groupes (groupes « discipline » et « statut de l'auteur ») : histogramme pour les moyennes et distributions en fréquence

Les histogrammes seront donc systématiquement utilisés pour indiquer un effectif par catégorie quand il s'agit de comparer la distribution de variables nominales. Lorsque nous avons affaire à une variable continue, les histogrammes peuvent de nouveau être utilisés, pour comparer la moyenne de la distribution dans un nombre important de catégories. Par exemple, nous comparons le volume moyen d'un article en nombre de mots dans les différentes revues de LET. De même, nous comparons le taux moyen des principaux temps verbaux dans le corpus pris dans son ensemble, et dans les différents sous-corpus (LET, SCL, D, Nd). Ici, nous avons donc un nombre important de catégories (par exemple 8 revues, ou encore 5 corpus différents), comparées sur une seule série de données (la valeur moyenne de telle ou telle variable), ou plusieurs séries de données (la valeur moyenne pour plusieurs variables, par exemple le taux de chaque temps verbal dans le texte).

Pour les variables continues, les histogrammes nous permettent de fournir un premier aperçu des différences entre catégories, aperçu basé sur la seule valeur moyenne ; en tant qu'indice de tendance centrale, celle-ci est supposée représenter au mieux toutes les valeurs, mais elle ne rend pas compte de l'ensemble des valeurs et masque une dimension essentielle d'une distribution, à savoir sa dispersion.

Dans notre étude, l'hypothèse d'un effet de la discipline sur le genre donne de l'importance à cette question de la comparaison d'une variable continue (d'ordre linguistique) dans les deux groupes LET et SCL. Il en va de même pour la comparaison entre les articles de doctorants et de non-doctorants. Pour comparer la distribution d'une variable continue dans deux (ou plus) groupes, plusieurs solutions sont offertes. S'il est possible d'identifier chacune des valeurs de manière exhaustive, une telle solution se heurte dans notre étude au problème du nombre de valeurs (110 articles par discipline). Par ailleurs, la disparité des valeurs rend la visualisation produite peu pertinente. Or, une représentation graphique doit aider à voir ce que les données permettent de dire.

Une autre solution consiste alors à identifier l'ensemble des valeurs de manière exhaustive, mais en les ordonnant. Nous avons eu parfois recours à cette méthode, en proposant une représentation des 110 articles qui classe ces derniers (par exemple, un classement de l'article le moins volumineux au plus volumineux en fonction du volume cumulé du corps du texte et des notes en nombre de mots, le graphique ayant pour objectif de mettre en évidence le poids pris par le corps du texte et des notes sur le volume total, cf. chapitre 4).

Cette mise à plat ordonnée de l'ensemble des valeurs ne sert cependant pas la visée comparative ; pour comparer les distributions à proprement parler, il importe, comme nous l'avons vu avec les tests de statistiques inférentielles, de comparer à la fois la tendance centrale et la dispersion des valeurs caractéristiques de la distribution. On peut par exemple conjointement à un indice de tendance centrale les valeurs minimales et maximales, pour rendre compte de l'étendue de la distribution (la différence entre la valeur maximale et la valeur minimale) ; mais l'étendue n'est pas le seul indice de dispersion, et ne permet pas en elle-même de rendre compte de la manière dont les valeurs sont plus ou moins resserrées autour de la tendance centrale ou s'en écartent.

Une solution intéressante est fournie avec les boîtes à moustache (box plots)¹ de J.W. Tukey (1977). Pour en expliquer le principe, nous en donnons ci-dessous un exemple, de forme élémentaire, que nous empruntons à N. Hunt (1996). Ce type de représentation associe aux valeurs minimum (min) et maximum (max) les trois quartiles (la médiane en est un, qui partage la série en deux groupes d'effectifs égaux, et ces deux groupes peuvent de nouveau être divisés par deux quartiles, q1 et q3, de manière à ce que 25% des valeurs soient comprises dans chacun des quatre groupes ainsi obtenus).

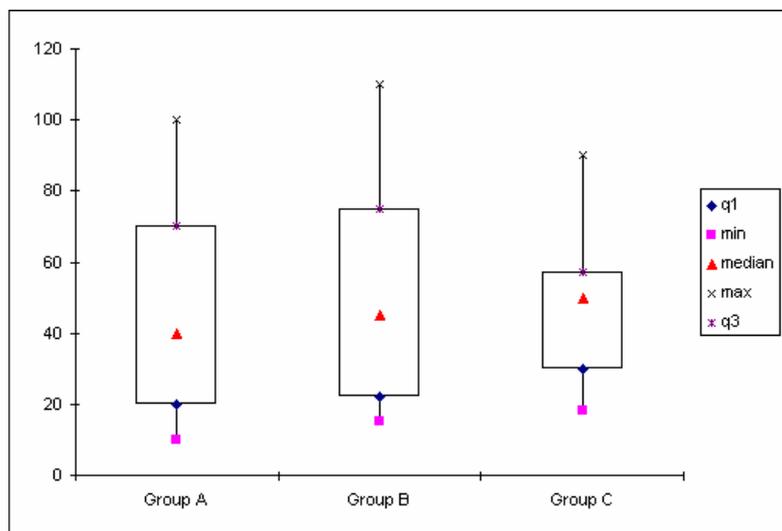


Figure 3.8. Exemple de boîte à moustache dans sa forme élémentaire (N. Hunt, 1996).

La « boîte » comprise entre le quartile 1 et le quartile 3 et ses « moustaches », qui vont du quartile 1 au minimum pour l'une, et du quartile 3 au maximum pour l'autre fournissent une visualisation qui compare à la fois la tendance centrale et la dispersion caractéristiques de chaque distribution de valeurs. C'est précisément ce que nous entendons analyser : nous voulons mettre l'accent sur la distribution des valeurs d'une variable, et à comparer cette distribution dans deux groupes (LET et SCL, D et ND)².

L'exemple nous permet cependant de signaler le principal inconvénient des boîtes à moustaches : la lecture d'un tel mode de représentation semble en première approche assez peu aisée. Comme d'autres modes de représentation graphique, la lecture de celui-ci requiert de comprendre comment il est construit. Deux problèmes peuvent expliquer la difficulté posée : d'une part, la notion de quartile, à la base de la construction, et d'autre part le fait qu'il s'agit d'un mode de représentation graphique peu répandu, en linguistique du moins³.

Cette difficulté potentielle de lecture nous a conduit à renoncer aux boîtes à moustaches et à préférer un autre mode de représentation des données, impliquant d'abord un autre mode d'organisation des données. Contrairement à un mode de représentation qui indique chacune des données d'une série (par exemple, la valeur de la variable « nombre de mots du texte » pour chaque article), la boîte à moustaches se base sur une description statistique de la distribution de valeurs en termes de quartile. Une autre manière de décrire les données pour leur donner du sens est de réorganiser la série de données avec la méthode de la distribution en fréquence (ou distribution de fréquences des données).

¹ Ce type de graphique relève des graphiques en forme de boîtes, traités par exemple dans (Howell, 1998 :60).

² Et non trois groupes comme dans l'exemple donné.

³ Les deux problèmes sont liés : la rareté de ce type de schémas peut s'expliquer par le caractère peu répandu de la notion de quartile dans les analyses statistiques (du moins en tant que telle, car la notion de médiane est fréquemment utilisée, et correspond, statistiquement, au second quartile).

La distribution en fréquence correspond à une description statistique d'une série de données. Plutôt que d'avoir affaire à une longue série de valeurs affectées à chaque individu, où l'ordre entre les valeurs n'a pas de sens, il va s'agir de rendre l'information statistique plus lisible, autrement dit de mieux faire apparaître les caractéristiques de la série de données (i.e. les caractéristiques de la distribution des valeurs). Pour ce faire, on cherche à regrouper les individus qui sont définis par des valeurs proches. On établit donc des intervalles de valeurs, et on calcule le nombre d'individus concernés, ce qui revient à déterminer la fréquence de cet intervalle de valeurs.

Si l'on reprend l'exemple de la variable « nombre de mots du texte », on peut lister chaque article en fonction du nombre de mots qu'il comprend, on peut aussi comme nous l'avons dit *supra* ordonner les articles du moins au plus volumineux en fonction du nombre de mots qu'il comprend, mais on peut aussi déterminer des intervalles (de 50 mots) et définir combien d'articles on a pour chacun des intervalles qui de 50 en 50 vont de 0 à 20000.

L'objectif est donc de rendre plus lisible encore la distribution des valeurs, d'abord pour éviter d'avoir une série de données sur 110 ou 220 articles, qui, même ordonnées, produisent une surcharge d'informations, et d'autre part pour mieux voir la tendance centrale et la dispersion de la distribution de valeurs : la distribution en fréquence organisant les données de manière à montrer comment les individus se distribuent dans les différents intervalles de valeurs. L'exemple suivant permet d'illustrer le principe de ce mode d'organisation spécifique des données : il correspond au type de visualisation de la distribution en fréquence que nous avons choisi, sous forme de courbe.

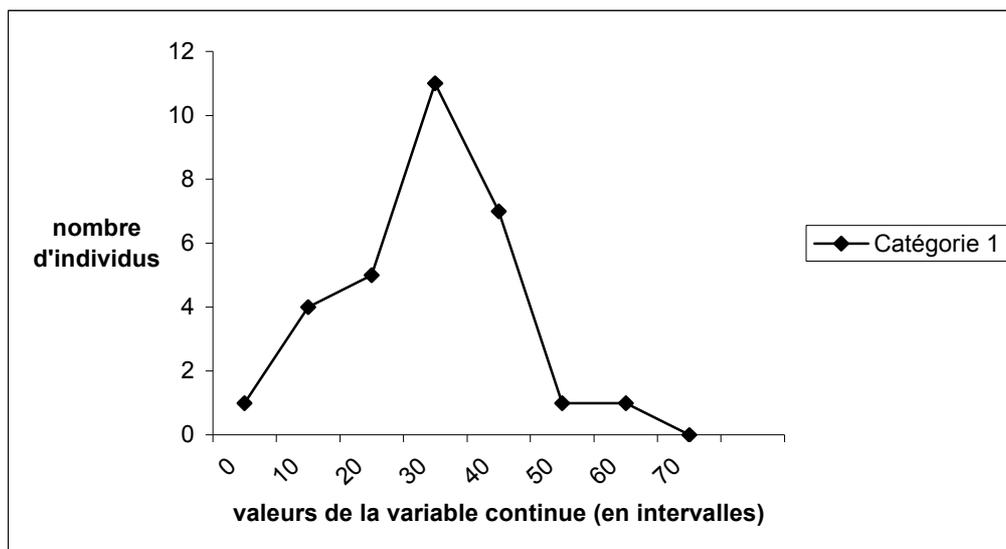


Figure 3.9. Exemple de distribution en fréquence d'une variable continue.

Le graphique permet d'associer pour chaque intervalle de valeurs ($[0-10[$, $[10-20[$, etc., sur l'axe des abscisses) le nombre d'individus (axe des ordonnées) dont le score se situe dans cet intervalle. La distribution des valeurs de la variable continue dans la catégorie 1 (ou le groupe 1, par exemple le groupe des articles de LET, le groupe des articles de doctorants, etc.) est telle qu'un individu se situe dans l'intervalle $[0-10[$, 4 se situent dans l'intervalle $[10-20[$, etc. 30 individus sont ici représentés.

La forme de la distribution en fréquence indique les tendances importantes qui caractérisent les données : les intervalles qui rassemblent le nombre le plus important

d'individus correspondent à la tendance centrale de la distribution des valeurs¹. La courbe peut être plus ou moins pointue (un grand nombre d'individus dans un espace resserré de valeurs, un nombre faible d'individus dans les intervalles à la marge, à droite et à gauche de la tendance centrale), ou plus ou moins aplatie en son sommet, lorsque les individus se distribuent de manière plus équitable entre les différentes catégories de valeurs ; dans ce cas, on aura une variabilité plus importante entre individus que dans le premier.

Les distributions en fréquence se prêtent à une lecture des données tout à fait propice à l'analyse du genre : elles favorisent une analyse de ce dernier en termes de tendances centrales (ce qui représente au mieux toutes les données, i.e. le genre) et de dispersion (la diversité irréductible du genre). Il ne s'agit pas cependant de recourir systématiquement à une telle représentation graphique. Nous l'utilisons pour comparer la distribution de valeurs dans deux groupes (notamment les articles de LET et ceux de SCL), lorsque les statistiques inférentielles ont montré que la différence est significative, c'est-à-dire que la distribution de la variable continue dépend de la variable indépendante, car ce n'est qu'à cette condition, comme nous l'avons dit, que ce que nous décrivons dans l'échantillon de données (le corpus, ou un sous-corpus) est validé pour l'analyse de la population (le genre).

Le graphique aide alors à voir ce que les données permettent de dire sur la tendance centrale et la dispersion des valeurs dans chaque groupe : dans lequel la tendance centrale est-elle la plus élevée ? Dans lequel la dispersion des valeurs est-elle la plus importante (i.e. la variabilité entre individus) ? La distribution en fréquence d'une variable continue sert ainsi de point de départ, avec l'analyse statistique des données, à une discussion sur les caractéristiques du genre de l'article, les différences entre les groupes, les régularités du genre et la diversité des textes y compris au sein d'un même groupe.

Au final, l'analyse quantitative et statistique des caractéristiques des textes requiert donc un important travail d'organisation des données.

En première étape, il s'agit de constituer une base de données qui identifie chaque article du corpus en référence aux variables indépendantes (discipline, statut de l'auteur, revue) et aux différentes variables linguistiques (des variables nominales, comme par exemple, la présence ou non d'une bibliographie), et, en nombre plus important, des variables continues, comme par exemple le nombre de mots du texte, le taux de verbes au présent par rapport à l'ensemble des verbes, etc.). Cette base de données existe en une triple version : pour le corps de l'article, pour le corps de l'article sans les exemples ni les citations, et pour les notes.

A cette étape, il est important de neutraliser la longueur des textes. Certaines variables (par exemple le nombre de phrases) dépendent du nombre de mots des textes, et comparer les textes sur la valeur absolue de la variable n'aurait pas grand sens. Il faut alors procéder à un ratio sur le nombre de mots du texte pour obtenir les valeurs de la variable en fréquence relative.

Ce ratio pose alors des difficultés pour commenter les résultats : par exemple, le nombre de phrases dans le corps de l'article LET001 est de 268, pour un nombre de mots de 6989 et le nombre de phrases dans le corps de l'article LET002 est de 327, pour un nombre de phrases de 5030. Si l'on calcule le nombre de phrases en fréquence relative, on pourra donc comparer LET001 et LET002 en disant que le premier comporte 0,04 phrases en fréquence relative contre 0,07 pour le second. Ces résultats ne sont pas très parlants, et pour que le

¹ Ici, elle se situe dans l'intervalle entre 30 et 40 ; la distribution en fréquence doit confirmer la description statistique de la distribution ; dans le cas pris comme exemple, elle correspond aux valeurs obtenues pour 30 individus, valeurs telles que la moyenne est de 32,5 et la médiane de 35,5.

nombre de phrases en fréquence relative colle mieux à la réalité des textes, nous utilisons donc un coefficient qui correspond au volume moyen de l'article¹, et que nous appliquons à l'ensemble des articles. Nous parlons de fréquence relative ou de ratio sur le nombre de mots du texte ramené à un article moyen.

En seconde étape, la base de données fait l'objet de procédures de description statistique : la distribution des valeurs est décrite en termes d'effectif (variables nominales, qu'elles soient indépendantes ou linguistiques) et de tendances centrales et de dispersion (variables continues). La description statistique permet alors de décider des tests de statistiques inférentielles à mobiliser. Nous testons systématiquement 1) l'effet de la discipline, 2) du statut de l'auteur, et 3) du statut de l'auteur dans les articles de LET et dans les articles de SCL sur l'ensemble des variables linguistiques, et parfois aussi des corrélations entre variables linguistiques (dans l'ensemble du corpus et en fonction de la discipline, du statut de l'auteur, etc.).

Une fois mis en évidence des liens de dépendance entre deux variables (une variable indépendante et une variable linguistique ou deux variables linguistiques), il faut alors revenir à la description statistique des données pour voir comment en rendre compte : par exemple, pour une différence significative de la discipline, laquelle affiche un effectif plus élevé dans le cas de variables nominales ou une tendance centrale plus élevée pour telle ou telle variable continue ? Laquelle manifeste une plus grande variabilité entre les textes ? Les outils de visualisation aident alors à voir ce que les données permettent de dire sur le genre de l'article.

Dans le cas de variables nominales (indépendantes ou linguistiques), la représentation graphique se base sur la description statistique de la variable (distribution de la variable en termes d'effectif pour chaque catégorie) et consiste en un histogramme.

Dans le cas de variables continues (linguistiques), la représentation graphique demande à réorganiser les données en produisant une nouvelle description de la série, la distribution en fréquence, pour analyser la distribution des valeurs de la variable continue dans chaque groupe (par exemple, les articles de LET *versus* ceux de SCL) en termes de tendance centrale et de dispersion des valeurs.

3.2.4. L'analyse du genre dans un va-et-vient entre échantillon de données et textes

Nous avons traité jusqu'ici d'une part de l'analyse automatique au moyen du logiciel Cordial, d'autre part de l'analyse statistique de l'échantillon de données ; ces données sont numériques, dans le sens où elles résultent d'une mesure², mais il existe comme nous l'avons vu différentes échelles de mesure (et conséquemment différents types de variables, les variables nominales et continues dans le cadre de notre étude). Ces données soumises aux procédures statistiques peuvent résulter soit de l'analyse automatique soit d'une analyse manuelle des textes.

L'analyse dite manuelle par opposition à l'analyse automatique est en fait omniprésente. Elle est nécessaire, sous la forme d'un travail d'observation et de préparation des textes, comme préalable à l'analyse automatique pour déterminer les descripteurs du genre à retenir, en évitant les risques d'analyses non pertinentes (ou le bruit). Elle la complète aussi, en permettant de se pencher sur des traits linguistiques qui ne peuvent être pris en charge par l'analyse automatique (par exemple, la présence ou non d'une bibliographie, et l'analyse des autres indices de structuration des textes, comme nous le verrons au chapitre 4). Ces traits linguistiques analysés manuellement font dans nombre de cas l'objet de quantifications ; lorsque nous quantifions les phénomènes, nous les soumettons comme nous

¹ Ou, plus précisément, au volume moyen du corps de l'article, du corps sans les exemples ni les citations, ou des notes, selon les données auxquelles on a affaire.

² La mesure associe comme nous l'avons dit un nombre à l'objet observé (Cf. D.C. Howell, 1998).

l'avons dit à une analyse statistique, à l'exception des cas où nous comptons trop peu d'occurrences (les constructions du *il* impersonnel au chapitre 7).

Comme l'automatisation, l'analyse quantitative et statistique est à envisager dans un va-et-vient avec les textes. Si l'étude statistique doit permettre de pointer la pertinence de certaines variables pour la caractérisation du genre, nous délaissions parfois comme nous l'avons expliqué certains résultats trop problématiques, pour nous centrer dans nos analyses sur ceux qui offrent une prise véritable sur les textes : il s'agit de fonder l'analyse sur la matérialité des textes, mais de s'intéresser à des phénomènes dont il est possible d'interpréter la fonctionnalité dans le texte. L'analyse recourt dans cette perspective à l'examen des phénomènes en corpus et en rend compte par des exemples.

C'est donc en référence aux observations faites dans les textes que l'on peut donner sens aux résultats de l'analyse quantitative du genre ; il s'agit de mieux déterminer quels traits sont pertinents pour caractériser le genre en regardant à quoi correspondent les différents résultats dans les textes, quitte à exclure certains phénomènes jugés problématiques. L'analyse consiste ainsi à mettre en perspective les entrées linguistiques entre elles, ou avec les profils d'articles déjà établis à partir d'autres critères (nous le verrons à propos des formalismes, qui peuvent être reliés à des modèles différents au niveau de la structure des textes, cf. au chapitre 4). Elle vise également à analyser ces entrées linguistiques en référence aux enjeux du genre de l'article, en particulier au niveau du rapport à l'objet et au lecteur, de la manifestation de l'auteur dans son texte, ou encore de l'image de l'activité de recherche donnée dans le texte.

En retour, l'analyse des fonctions prises par les unités linguistiques dans un genre montre l'intérêt de ces derniers pour la description linguistique : elle peut permettre de pallier les difficultés posées par l'objectif de description d'une langue générale, qui risque d'aboutir au final à des inventaires hétéroclites. Comme nous le verrons au chapitre 5 notamment, avec les ponctuations ou le lexique transdisciplinaire, l'analyse menée au sein d'un genre offre une prise pour rendre compte de la polysémie des unités linguistiques ; on voit là encore l'intérêt de coupler à une analyse automatique un examen plus circonstancié, car les entrées formelles ou grammaticales retenues par l'analyseur ne satisfont pas toujours une description proprement linguistique des phénomènes. Dans le cadre d'un genre, il semble envisageable d'analyser comment la polysémie peut tendre à être restreinte, ou à s'organiser de manière relativement stable (par exemple, à propos des valeurs prises par les guillemets), et d'identifier des unités sémantiquement différentes pour une même entrée formelle (par exemple, un terme comme *démonstration*, dans son usage général ou transdisciplinaire, cf. au chapitre 5).

Pour l'analyse des textes, nous mobilisons dans certains cas des outils de recherche automatisée : il s'agit de formuler des requêtes pour observer certains faits linguistiques au sein du corpus. Deux types d'outils sont utilisés, selon que nous avons besoin de faire une recherche sur le texte annoté morpho-syntaxiquement, ou que nous recherchons simplement des formes (i.e. des suites de caractères, qu'il s'agisse d'une ponctuation, d'un terme, d'un syntagme, etc.).

Dans le premier cas, nous utilisons le concordancier de Cordial, sur les textes étiquetés par ce même logiciel. Dans le second, nous utilisons parfois WConcord¹, qui permet en particulier de rechercher une liste de termes. Il produit une quantification des concordances et une modélisation sous la forme d'un tableau organisant les co-occurents du terme recherché en fonction de leur position par rapport au terme et de leur fréquence. Parfois aussi, nous utilisons un simple éditeur de texte, TextPad², qui permet d'extraire rapidement l'ensemble

¹ Version 2.0., Z. Martinek, en collaboration avec L. Siegrist, Technische Hochschule Darmstadt, 1996.

² Version 4.7.3., de Helios Software Productions, 2004.

des paragraphes où intervient l'objet recherché, et d'avoir un accès au texte concerné au moyen d'un simple lien hypertexte.

L'exploitation de ces outils rejoint ce que nous avons dit de l'automatisation de l'analyse. Qu'il s'agisse de décrire systématiquement des données quantitatives sur les textes ou de se pencher de manière plus qualitative sur certains phénomènes, la linguistique de corpus opère sur un continuum entre des procédures automatiques et des procédures manuelles d'observation. Elle ouvre la voie à une réflexion sur la manière de rendre commune une description linguistique et de l'objectiver, par exemple avec les procédures d'annotation inter-annotateurs¹.

En somme, l'analyse qualitative intervient comme préalable à l'analyse automatique d'une part et quantitative d'autre part, et comme leur prolongement nécessaire, pour spécifier et compléter les résultats statistiques (incluant ceux de l'analyse automatique) par une analyse en contexte, qui met en relation différents phénomènes, à différents niveaux de la textualité. Elle permet d'établir des critères et des catégories d'analyse, pour mieux cerner linguistiquement les phénomènes (les indices de structuration du texte au chapitre 4, la diversité des emplois des guillemets au chapitre 5, la diversité des emplois de *je* et de *nous* (chapitre 7), la diversité des modes de désignation des points de vue existants dans le champ de recherche (chapitre 7).

C'est là que l'idée de va-et-vient prend tout son sens : dans certains cas, l'analyse qualitative des textes (les emplois de telle ou telle marque) peut permettre d'aboutir à une typologie (par exemple, pour les modes de désignation de l'existant, cf. au chapitre 7., ou les types d'exemples utilisés dans les articles, cf. au chapitre 4) ; à défaut, elle sert la mise en évidence de critères multiples (par exemple, pour décrire la structure des textes, face à l'hétérogénéité des plans d'articles). La typologie ou les différents critères peuvent alors faire l'objet d'une quantification, et l'on rejoint le rôle de l'analyse qualitative en tant que préalable à la quantification. En même temps, l'analyse qualitative permet de discuter la diversité des emplois, et recourt à l'exemplification. Elle peut ainsi spécifier les résultats statistiques, en regard des enjeux des textes (par exemple, le rapport au lecteur et à l'objet au chapitre 4) et mobilise ainsi sa fonction de prolongement et de discussion circonstanciée des résultats statistiques.

Le va-et-vient entre une approche visant la quantification et une approche qualitative doit ainsi répondre à l'objectif de description globale du genre et de descriptions linguistiques fines de ses caractéristiques, qui montrent en retour l'intérêt des genres pour l'étude de la langue. Enfin, elle rend possible l'analyse de profils diversifiés de textes, sur la base de faisceau de critères, et d'opposition entre leurs caractéristiques, et rend compte aussi de la continuité existante entre ces profils.

Nous parlons donc d'analyse quanti-qualitative pour préciser que si nombre de phénomènes font l'objet de quantifications, tous cependant ne sont pas quantifiés, ne serait-ce que parce que, dans certains cas, la diversité des emplois fait obstacle à une telle quantification. Notre étude se base sur la matérialité des textes, et cherche à mettre en évidence des indices pour les

¹ Si l'objectif est d'annoter un corpus mais que l'annotation automatique n'est pas possible, et que le phénomène présente une polysémie telle qu'il peut y avoir désaccord sur cette polysémie même (cf. Véronis, 2004), on peut recourir à une méthode d'annotation par divers annotateurs humains. Celle-ci va permettre de spécifier la typologie du phénomène, et de voir sur quels points spécifiques se posent des problèmes de multiplicité interprétative (le phénomène peut faire l'objet d'un double étiquetage sur lequel tous les annotateurs humains s'accordent, ou d'un étiquetage, simple ou double, et variable en fonction des différents annotateurs. On trouvera un exemple de cette procédure dans F. Rinck, A. Tutin et C. Hermann (2004) à propos des guillemets.

analyser et dans certains cas à proposer des typologies, au titre d'outils heuristiques pour la description du genre et des faits linguistiques. Le caractère plus quantitatif ou plus qualitatif de l'étude dépend donc des faits analysés, mais l'étude du genre procède au final d'un va-et-vient entre les deux pôles, celui de données numériques (ou de mesures faites sur des objets textuels) et celui de l'observation des textes.

Notre étude du genre dans les chapitres 4 à 7 ne rendra pas compte de l'ensemble des traits examinés, mais d'une sélection de traits pertinents, pour lesquels, le cas échéant, nous fournissons en annexes le détail de l'analyse statistique des données (descriptives et inférentielles). C'est l'analyse des textes qui détermine au final quels traits sont pertinents pour l'analyse du genre. Il nous reste pour finir à préciser ici quels sont ces traits textuels concernés par notre étude.

3.3. Les traits textuels retenus

Nous proposons ici de dresser la liste des traits textuels retenus pour l'analyse. Celle-ci se base d'une part sur les variables linguistiques sélectionnées parmi le jeu de descripteurs morpho-syntaxiques du logiciel Cordial. D'autre part, elle vise à spécifier et compléter ces résultats : il s'agit d'analyser quantitativement et qualitativement des traits textuels non pris en charge par l'analyse automatique, d'ordre morpho-syntaxique, ou d'affiner la description quantitative fournie par l'analyse automatique en s'intéressant aux mêmes traits mais de manière qualitative, par l'observation de leur fonctionnement dans les textes.

Nous présentons ici les traits textuels sous la forme de trois rubriques : les 97 variables Cordial, le lexique et enfin les autres traits retenus dans l'étude et non analysés par Cordial. Les deux premières séries de descripteurs (variables Cordial et lexique transdisciplinaire) seront exploitées dans l'étude de la variation stylistique du genre, avec d'autres variables morpho-syntaxiques et lexicales. Rappelons que nous présenterons la méthodologie spécifique de l'étude des styles dans le chapitre qui leur est consacré (chapitre 6).

3.3.1. Les 97 variables Cordial

Parmi les descripteurs disponibles dans Cordial au titre de résultat de l'étiquetage morpho-syntaxique produit par le logiciel, nous avons retenu 97 variables par hypothèse pertinentes pour une analyse du genre basée sur l'unité phrase et ses composantes. Rappelons que nous étudions l'ensemble de ces variables sur le corps de l'article, le corps sans les exemples ni les citations et sur les notes, et ce dans l'ensemble des 220 articles du corpus, par discipline, en fonction du statut de l'auteur, et en fonction du statut de l'auteur dans chaque sous-corpus disciplinaire.

Toutes les données ainsi produites ne seront pas commentées dans les chapitres suivants, consacrés à la caractérisation du genre. Par exemple, nous avons vu avec les parties du discours (noms, adjectifs, verbes et adverbes) que l'interprétation des différences observées entre groupes était problématique, car ces catégories linguistiques sont trop hétérogènes ; elles ne nous paraissent donc pas finalement tout à fait pertinentes pour une description des régularités et de la variabilité du genre qui tienne compte de la fonction dans les textes des unités linguistiques, quoiqu'elles aient montré leur intérêt dans la comparaison inter-générique¹. La plupart des variables sont cependant traitées, en particulier au chapitre 4, à propos de la structure des textes, et au chapitre 5, qui s'intéresse à la phrase, aux ponctuations et au lexique.

¹ Cf. notamment D. Malrieu et F. Rastier (2001), C. Poudat (2006a).

Nous listons ces 97 variables telles qu'elles sont présentées et classées dans le logiciel. Bien entendu, l'analyse de l'article ne saurait se calquer sur ce classement largement discutable, et demande donc à réorganiser les variables en fonction des pistes d'analyses suggérées par l'observation des données statistiques et des textes.

TOTAUX
Nombre total de caractères, y compris les espaces et ponctuations
Nombre total de caractères, y compris les ponctuations mais sans les espaces et tabulations
Nombre total de mots
Nombre total de phrases
Nombre total de phrases verbales
Nombre total de paragraphes
Nombre total de ponctuations
MOYENNES
Nombre moyen de lettres par mot
Nombre moyen de mots par proposition
Nombre moyen de mots par phrase
Nombre moyen de mots par paragraphe
Nombre moyen de propositions par phrase
Nombre moyen de phrases par paragraphe
MOYENNES GRAMMATICALES
Nombre moyen de substantifs par proposition
Nombre moyen d'adjectifs par proposition
Nombre moyen d'adverbes par proposition
Nombre moyen de pronoms par proposition
PHRASES
% des phrases comportant au moins une proportion subordonnée par rapport à l'ensemble des phrases
% de phrases interrogatives par rapport à l'ensemble des phrases
% de phrases exclamatives par rapport à l'ensemble des phrases
PONCTUATIONS
% de points par rapport à l'ensemble des ponctuations
% de points de suspension par rapport à l'ensemble des ponctuations
% de points d'exclamation par rapport à l'ensemble des ponctuations
% de points d'interrogation par rapport à l'ensemble des ponctuations
% de points virgules par rapport à l'ensemble des ponctuations
% de deux-points par rapport à l'ensemble des ponctuations
% de virgules par rapport à l'ensemble des ponctuations
% de parenthèses par rapport à l'ensemble des ponctuations
% de tirets par rapport à l'ensemble des ponctuations
% de crochets et accolades par rapport à l'ensemble des ponctuations
MORPHOLOGIES
% de mots-outils par rapport à l'ensemble des mots
% de mots significatifs (c.à-d. non mots-outils) par rapport à l'ensemble des mots
% de substantifs par rapport au total des mots significatifs (substantifs, adjectifs, verbes, adverbes)

% d'adjectifs par rapport au total des mots significatifs (substantifs, adjectifs, verbes, adverbes)
% de verbes par rapport au total des mots significatifs (substantifs, adjectifs, verbes, adverbes)
% d'adverbes par rapport au total des mots significatifs (substantifs, adjectifs, verbes, adverbes)
DETERMINANTS
% d'articles définis par rapport à l'ensemble des déterminants
% d'articles indéfinis par rapport à l'ensemble des déterminants
% d'articles définis par rapport à l'ensemble des articles
% d'adjectifs démonstratifs par rapport à l'ensemble des déterminants
% d'adjectifs cardinaux par rapport à l'ensemble des déterminants
% d'adjectifs interrogatifs par rapport à l'ensemble des déterminants
% d'adjectifs indéfinis par rapport à l'ensemble des déterminants
% d'adjectifs possessifs par rapport à l'ensemble des déterminants
% d'adjectifs possessifs à la 1e personne du singulier par rapport à l'ensemble des adjectifs possessifs
% d'adjectifs possessifs à la 2e personne du singulier par rapport à l'ensemble des adjectifs possessifs
% d'adjectifs possessifs à la 3e personne du singulier par rapport à l'ensemble des adjectifs possessifs
% d'adjectifs possessifs à la 1e personne du pluriel par rapport à l'ensemble des adjectifs possessifs
% d'adjectifs possessifs à la 2e personne du pluriel par rapport à l'ensemble des adjectifs possessifs
% d'adjectifs possessifs à la 3e personne du pluriel par rapport à l'ensemble des adjectifs possessifs
PRONOMS
% de pronoms démonstratifs par rapport à l'ensemble des pronoms
% de pronoms indéfinis par rapport à l'ensemble des pronoms
% de pronoms relatifs par rapport à l'ensemble des pronoms
% de pronoms personnels par rapport à l'ensemble des pronoms
% de pronoms possessifs par rapport à l'ensemble des pronoms
% de pronoms personnels à la 1e personne du singulier parmi l'ensemble des pronoms personnels sujets
% de pronoms personnels à la 2e personne du singulier parmi l'ensemble des pronoms personnels sujets
% de pronoms personnels à la 1e personne du pluriel parmi l'ensemble des pronoms personnels sujets
% de pronoms personnels à la 2e personne du pluriel parmi l'ensemble des pronoms personnels sujets
VERBES
% de verbes au présent par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes à l'imparfait par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au passé simple par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au futur par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au conditionnel par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au subjonctif présent par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au subjonctif imparfait par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au subjonctif passé par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au subjonctif plus-que-parfait par rapport à l'ensemble des verbes conjugués (tps simples et composés)
% de verbes au passé composé par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au plus-que-parfait par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au passé antérieur par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au futur antérieur par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes au conditionnel passé par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes à l'impératif par rapport à l'ensemble des verbes conjugués (temps simples et composés)
% de verbes à la 1e personne du singulier parmi l'ensemble des verbes conjugués (temps simples et composés)
% de verbes à la 2e personne du singulier parmi l'ensemble des verbes conjugués (temps simples et composés)
% de verbes à la 3e personne du singulier parmi l'ensemble des verbes conjugués (temps simples et composés)

% de verbes à la 1e personne du pluriel parmi l'ensemble des verbes conjugués (temps simples et composés)
% de verbes à la 2e personne du pluriel parmi l'ensemble des verbes conjugués (temps simples et composés)
% de verbes à la 3e personne du pluriel parmi l'ensemble des verbes conjugués (temps simples et composés)
% des verbes pronominaux parmi l'ensemble des verbes
% des verbes intransitifs parmi l'ensemble des verbes
% des verbes transitifs directs parmi l'ensemble des verbes
% des verbes transitifs indirects parmi l'ensemble des verbes
% des verbes prenant l'auxiliaire « avoir » parmi l'ensemble des verbes
% des verbes prenant l'auxiliaire « être » parmi l'ensemble des verbes
% des verbes à COD obligatoire parmi l'ensemble des verbes
% des verbes à COI obligatoire parmi l'ensemble des verbes
% des verbes précédés d'une négation parmi l'ensemble des verbes
PROPOSITIONS
% de propositions indépendantes parmi l'ensemble des propositions
% de propositions principales parmi l'ensemble des propositions
% de propositions relatives parmi l'ensemble des propositions
% de propositions subordonnées parmi l'ensemble des propositions
% de propositions coordonnées parmi l'ensemble des propositions
% de propositions ayant un complément d'objet direct
% de propositions ayant un complément d'objet indirect
% de propositions ayant un attribut du sujet

Figure 3.10. Liste des 97 variables Cordial

La première rubrique, intitulée « totaux » demande, comme nous l'avons évoqué *supra*, à établir un ratio sur le nombre de mots du texte pour neutraliser l'impact de leur longueur sur les caractéristiques que rassemble cette rubrique (nombre de phrase, etc.). Pour les autres, le fait que Cordial fournisse des nombres moyens et des pourcentages permet précisément de contourner ce problème de longueur variable des textes.

3.3.2. *Le lexique : restriction au lexique transdisciplinaire*

Le logiciel Cordial nous permet également une étude des textes. Comme nous l'expliquons au chapitre 5, c'est plus spécifiquement un lexique transdisciplinaire qui retiendra notre attention (de type hypothèse, méthode, résultat, etc.). Nous présentons la liste de termes retenus et les modalités d'analyse dans ce chapitre 5, et fournissons en annexes (Annexes 3) les descriptions statistiques et les tests de statistiques inférentielles sur les 100 termes retenus.

3.3.3. *Les traits analysés manuellement*

Parmi les traits linguistiques analysés manuellement, certains font l'objet d'une première caractérisation quantitative avec le logiciel d'analyse. D'autres ne sont pas pris en charge par ce dernier, et sont soumis à une analyse quantitative et qualitative, ou exclusivement qualitative dans certains des traits analysés au chapitre 7.

Nous listons ci-dessous ces traits, en les classant en fonction du chapitre d'analyse qu'ils intéressent et en précisant le corpus concerné par la quantification : 1) le corpus de 220 articles, soit 110 par disciplines, ou 2) un sous-corpus de 70 articles, soit 35 par disciplines, ou encore 3) un sous-corpus de 44 articles de SCL et enfin 4) un sous-corpus de 40 articles soit 20 par disciplines.

Quand il s'agit de relayer la quantification par un examen qualitatif, c'est ce même corpus qui est privilégié, mais dans le chapitre 7 en particulier, on se réfère également aux autres textes du corpus. L'étude peut, selon les traits, concerner le corps de l'article et/ou le corps de l'article sans les exemples ni les citations, et dans certains cas aussi les notes. Nous fournissons cette précision en début de chaque section d'analyse.

	Corpus concerné par la quantification
STRUCTURE (chapitre 4.)	
Nombre de notes	220
Bibliographies	220
Annexes	220
Introductions : titrées (1), identifiables (2), non identifiables (3)	220
Conclusions : titrées (1), identifiables (2), non identifiables (3)	220
Titrage des parties	220
Numérotation des sections	220
Sous-parties	220
Indices de structuration internes au paragraphe (tirets, puces, numéros)	220
Présence de formalismes	220
Poids des exemples et des citations dans les textes	70
Inventaire des exemples par discipline et quantification des types d'exemples	70
PHRASE, PONCTUATION, LEXIQUE (chapitre 5.)	
Nombre de passages entre guillemets	40
Fonction des guillemets	40
SUBJECTIVITE ET POLYPHONIE (chapitre 7.)	
Occurrences de je (<i>j', je et dérivés, m', me, moi, mes, mien, etc.</i>) : nombre et emplois	40
Occurrences de nous et dérivés : nombre et emplois	40
Nombre de <i>on</i>	40
Nombre de <i>il</i> impersonnels	40
Constructions du <i>il</i> impersonnel	40
Nombre de passifs	40
Les discours représentés	40
Les références dénominatives (à une personne ou un texte)	40
Les points de vue se présentant comme attestés	44 (SCL seulement)
Objectifs de l'article annoncés dans l'introduction	40
Lexique axiologique	40

Figure 3.11. Liste des traits analysés manuellement.

Cette liste de traits analysés manuellement montre la complexité de ce que l'on peut entendre par trait textuel : si, dans tous les cas, on se réfère à la matérialité des textes, certaines entrées listées, en particulier pour ce qui concerne la subjectivité et la polyphonie (chapitre 7.) recouvrent en elles-mêmes plusieurs traits (par exemple, pour les objectifs de l'article annoncés dans l'introduction), et c'est parfois leur analyse qui a pu permettre de déterminer la désignation même de l'entrée (par exemple, les points de vue se présentant comme attestés).

Si notre étude n'exploite pas les outils de traitements automatiques des langues les plus novateurs, elle se caractérise par une méthodologie relevant de la linguistique de corpus et mobilise les possibilités offertes par l'automatisation, au niveau de l'étiquetage morpho-syntaxique et des analyses qu'il permet, et au niveau des requêtes dans le corpus pour l'extraction de concordances (exhaustive ou non). Elle privilégie la quantification pour rendre compte statistiquement des régularités et de la variabilité du genre, et procède d'un va-et-vient entre les données numériques de l'échantillon et l'observation qualitative des textes.

L'exploitation de corpus et de méthodes en linguistique de corpus (automatisation de l'analyse, quantification, statistiques) pose au final nombre de questions complexes, qui ont trait à l'instrumentation en linguistique. Cette problématique demande à ce que se développe une réflexion spécifique, et l'on en trouve des jalons particulièrement intéressants dans les « portraits de linguiste(s) à l'instrument » que propose d'interroger B. Habert (2005), et dans la rubrique « corpus et méthodes » du site « Texto ! » où est paru ce texte ; les nouvelles possibilités offertes en linguistique appellent à dépasser les oppositions entre étude qualitative et quantitative, et entre une analyse traditionnelle des textes et le recours à l'informatique, pour questionner le rapport du linguiste à son objet.

Pour conclure ce chapitre, nous résumerons les principes de l'analyse que nous allons à présent développer. Le genre nous sert de point de départ pour constituer le corpus et caractériser les textes. Dans ce cadre, d'autres entrées sont mobilisées pour rendre compte des régularités observables entre les articles, tout en mettant l'accent sur leur diversité. Les traits linguistiques retenus doivent dans la mesure du possible être mis en relation entre eux et analysés en référence à leurs fonctions dans les textes ; en cela, ils peuvent montrer en retour l'intérêt d'être étudiés en tant que caractéristiques d'un genre, pour pallier les limites d'une description des unités linguistiques et de la textualité menée sans référence aux normes génériques.

Nous présentons dans les chapitres suivants les analyses menées sur le genre de l'article.

- Le chapitre 4 vise une macro-analyse du genre et la mise au jour de modèles de textes en s'intéressant aux composantes sémiographiques et structurelles des textes.
- Le chapitre 5 propose des éléments de description linguistique du genre ; si l'ensemble des chapitres est concerné par la description linguistique des textes, il faut comprendre ce choix de titre en référence aux catégories linguistiques traditionnelles. Les résultats issus de l'analyse automatique sont mobilisés pour décrire certains aspects de la phrase, les temps verbaux, les marques de personne et de non personne, ainsi que les ponctuations et le lexique transdisciplinaire.
- Le chapitre 6 requiert une méthodologie spécifique pour une analyse multidimensionnelle de la variation stylistique de l'article (i.e. les styles ou usages singuliers que chaque auteur fait du genre) et de la diversité interne du genre, avec comme point de départ des variables morphosyntaxiques issues de deux jeux d'étiquettes (dont celui qui sert l'étude menée au chapitre 3), ainsi que le lexique, et notamment le lexique transdisciplinaire étudié au chapitre 2.
- Enfin, le chapitre 7 aborde les marques de soi et de l'autre observables dans l'article, pour analyser la subjectivité et la polyphonie caractéristiques du genre et proposer ainsi des pistes pour une analyse énonciative de la rhétorique scientifique.

Pour chacun des traits analysés, nous commençons par préciser rapidement la manière dont ont été produites les données : le corpus utilisé, et notamment la question de savoir combien de textes, et quelle partie du texte sont concernés (corps de l'article, notes, corps de

l'article sans exemples ni citations), et la démarche générale (automatisation, quantification, examen qualitatif des textes).

Concernant les analyses statistiques, nous indiquons la plus grande partie des résultats significatifs des tests en notes de bas de page, et on trouvera en annexes le détail des descriptions et des tests réalisés pour chacune des variables traitées dans les quatre chapitres de l'analyse empirique.

Le caractère réflexif de notre écrit (un écrit de recherche sur l'écrit de recherche) impose des contraintes au niveau des exemples tirés du corpus auxquels nous nous référons. Pour éviter les risques d'ambiguïté, nous reproduisons les extraits tels qu'ils se présentent dans les textes. Les exemples courts, intégrés au texte sont mis entre guillemets et en italiques et les exemples de plus de 20 mots sont placés dans un paragraphe spécifique.

La seule modification qui nous semble essentielle consiste à mettre en gras les marques principales sur lesquelles s'appuie notre analyse, pour rendre aussi plus lisibles les relations éventuelles entre les traits. A défaut de numérotation dans le cas des exemples longs, nous recourons pour référencer les articles au code de l'article, inséré à la fin des exemples pour indiquer de quel texte ils sont tirés. Ce code associe la discipline (LET ou SCL) et un numéro de 001 à 110. On peut se référer aux annexes 1 pour retrouver à partir du codage les références bibliographiques de chaque article du corpus.

Le corpus « auteurs » traité dans le chapitre dédié à l'étude des styles (chapitre 6) fait l'objet d'un codage différent, qui associe le numéro du texte à l'identification de son auteur (A1 à A15-0 à 122). Les textes du corpus « auteurs » sont également référencés en annexes (annexes 4).

Chapitre 4.

Modèles de textes et rapports contrastés au lecteur et à l'objet

Nous proposons ici une première approche du genre de l'article, que l'on peut qualifier de macro-analyse en raison du palier textuel concerné¹. Elle s'intéresse aux caractéristiques structurelles de l'article sur la base de deux principes constitutifs de sa textualité : le fait que l'article est un texte fait de plusieurs textes, et le fait que l'article est un texte fait de plusieurs systèmes sémiotiques.

L'objectif est de mettre au jour des modèles de textes, et les aspects qui y contribuent sont les suivants :

- le volume du texte en nombre de mots,
- la manière dont l'article investit les différents espaces péri-textuels (notes, bibliographies, annexes, etc.),
- la structuration du corps du texte (ou plans de texte), et les indices de structuration de bas niveau, internes au paragraphe,
- les formalismes,
- les exemples et les citations.

L'analyse de ces caractéristiques montre en premier lieu la diversité des pratiques d'écriture de l'article. Des modèles contrastés de textes se manifestent, et avec eux des identités disciplinaires fortes. La description des modèles de textes repose sur des faisceaux de critères qui mettent en relief des différences au niveau de l'exploitation des espaces textuels, de la lisibilité de la structure et de la pluri-sémiotité de l'article.

La description ouvre des perspectives pour discuter de la fonctionnalité des textes en termes de guidage du lecteur : la structure des textes peut tendre à être explicitée de manière maximale, ou requérir à l'inverse davantage d'inférences de la part du lecteur. La tendance à l'explicitation maximale, que nous retrouverons avec d'autres dimensions linguistiques du genre (cf. chapitre 7) ne représente pas une norme partagée par les deux champs disciplinaires et révèle une diversité interne à chacun d'eux.

Les modèles de textes rendent également tangibles des images différentes de l'activité de recherche. Les identités disciplinaires reposent en particulier sur la question des

¹ F. Rastier (2001) propose en effet de distinguer les paliers micro-, méso et macro-textuels pour l'analyse sémantique. Le premier va du morphème à la lexie, le second du syntagme à la période, le dernier de la période au texte.

formalismes, et sur la gestion des sources documentaires et empiriques. Les caractéristiques structurelles du genre offrent ainsi une prise sur l'analyse de tendances épistémologiques plurielles, qui se positionnent diversement par rapport à une tentation de scientificité. L'analyse des exemples montre à titre programmatique leur intérêt pour la question du statut de la preuve dans les champs considérés, et celle de la construction des savoirs disciplinaires.

4.1. Le volume du corps du texte et des notes : premier indice d'une diversité insoupçonnée du genre

4.1.1. Volume total : proximité inter-disciplinaire et disparité intra-disciplinaire

Nous présenterons les données sur le volume des articles en nombre de mots ; le comptage est limité au volume constitué par le corps du texte et les notes. Nous avons donc exclu les bibliographies, annexes et résumés en raison des difficultés que cela pose : les résumés et les bibliographies peuvent être intégrés à l'article, mais ils peuvent aussi, dans certaines revues ou dans certains numéros, être réunis dans une partie spécifique. Les annexes demeurent quant à elles difficilement quantifiables en nombre de mots car certaines sont constituées de figures, de graphes ou d'autres types d'éléments à caractère pictographique.

Ainsi calculé, le volume du corpus est de 1 535 253 mots. Ceux-ci se répartissent de manière presque équivalente dans les deux disciplines : le corpus d'articles de Lettres contient 784713 mots, celui de Sciences du langage un peu moins, avec 750540 mots. La différence des mesures faites dans les deux disciplines n'est pas statistiquement significative et en moyenne, un article comporte donc un nombre de mots comparable dans les deux disciplines, quoique légèrement plus élevé en LET (7133,8 mots en LET contre 6823,1 en SCL).

Cette proximité entre les articles des deux disciplines ne doit pas masquer la variabilité interne à chacun des deux champs considérés. Le volume maximal et minimal en fournit un bon indice : si le volume moyen avoisine les 7000 mots dans les deux disciplines, un article d'études littéraires peut contenir entre 1451 et 15796 mots, un article de sciences du langage entre 1945 et 14021. Forte dans les deux disciplines, la variabilité des articles à ce seul niveau du volume relativise donc considérablement l'unicité supposée d'un modèle d'écriture de l'article : que peut-il y avoir de commun entre un texte de 1500 et un texte de 15000 mots ?

L'étude de la distribution des mesures atteste cependant d'une certaine stabilité du genre : dans les deux disciplines, les écarts sont importants entre les articles, mais une tendance centrale se manifeste clairement par une convergence entre la moyenne et la médiane¹ aux alentours de 6000 à 7000 mots

C'est de cette stabilité du genre et de sa diversité insoupçonnée dont rend compte le graphique ci-dessous : la distribution en fréquence du nombre de mots par article montre pour chaque discipline combien d'articles correspondent à l'intervalle de valeurs pour le volume en nombre de mots.

¹ C'est-à-dire la valeur se situant au milieu de la série de sorte que 50% des valeurs soient au-dessus et 50% au-dessous.

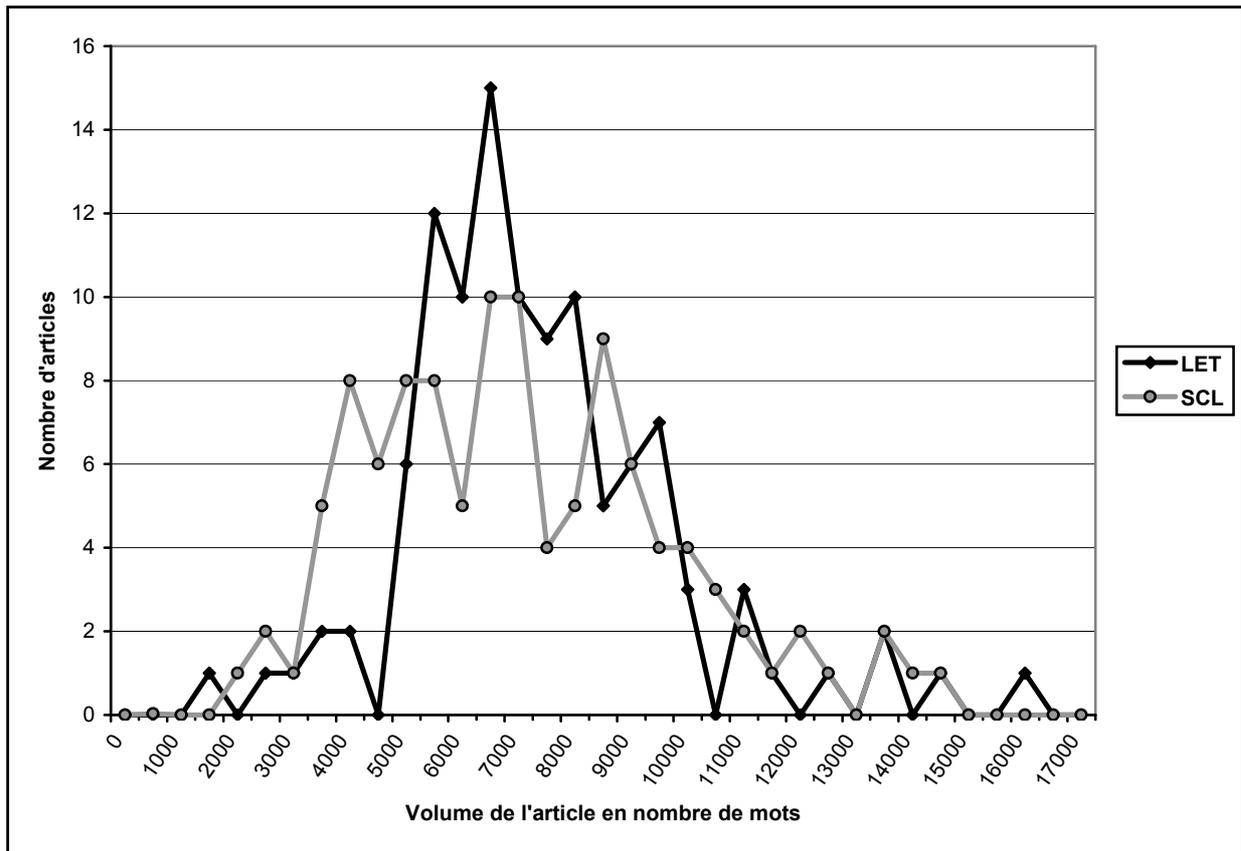


Figure 4.1. Distribution en fréquence du nombre de mots par article : comparaison disciplinaire.

La distribution des valeurs montre qu'entre les articles contenant le moins de mots et les articles en contenant le plus le rapport est de plus de 10 en Lettres et de 7,2 en Sciences du Langage. Un article peut ainsi atteindre plus du double du volume moyen dans les deux disciplines, et, au minimum, représenter 3,5 fois moins que le volume moyen en Sciences du langage et jusqu'à 5 fois moins en Lettres.

L'écart-type s'avère cependant légèrement supérieur en SCL, ce qui se traduit par une forme pyramidale moins accentuée de la distribution en fréquence des mesures. On observe donc une tendance centripète plus marquée des articles de LET, où le troisième indice de tendance centrale, le mode (c'est-à-dire la valeur dont l'effectif est le plus grand) rejoint les deux premiers, la moyenne et la médiane, cette égalité tendancielle des trois indices représentant une caractéristique des distributions normales ou gaussiennes.

Le volume cumulé du corps du texte et des notes des articles du corpus montre qu'à ce niveau, le genre de l'article se caractérise dans les deux disciplines par une extrême diversité des formats textuels. En dépit de cette diversité insoupçonnée, l'homogénéité du genre se manifeste : les mesures se distribuent de manière relativement régulière autour d'une tendance centrale se situant aux alentours de 6000 à 7000 mots dans les deux champs considérés. La différence entre les deux disciplines n'est pas statistiquement significative quoique le volume moyen soit légèrement supérieur en LET, de même que la tendance centripète du genre. Mais nous nous sommes restreints ici au nombre de mots dans le corps et les notes, et il nous reviendra donc de tenir compte d'autres éléments, en traitant notamment des annexes et des représentations graphiques présentes dans le corps du texte.

4.1.2. Contribution du corps du texte et des notes au volume total : différence disciplinaire dans l'usage des notes

Si l'on s'intéresse au volume du corps du texte, à celui des notes, et à la contribution respective du corps du texte et des notes au volume total, on observe à nouveau la même variabilité entre les articles. C'est ce que montrent les deux graphes ci-dessous, qui présentent le volume croissant des articles de LET puis de ceux de SCL, et la contribution respective du corps du texte et des notes au volume total de chaque article.

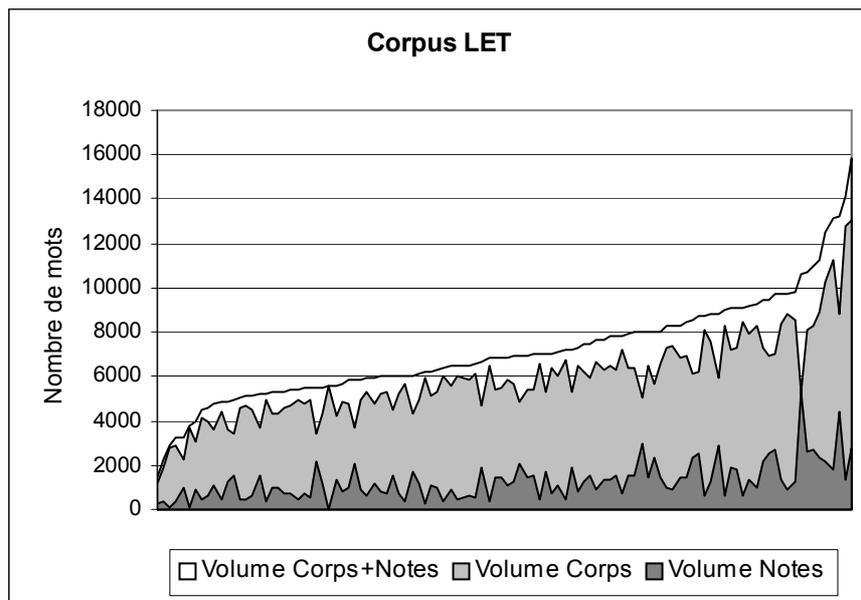


Figure 4.2a. Contribution du corps et des notes au volume total : articles de LET.

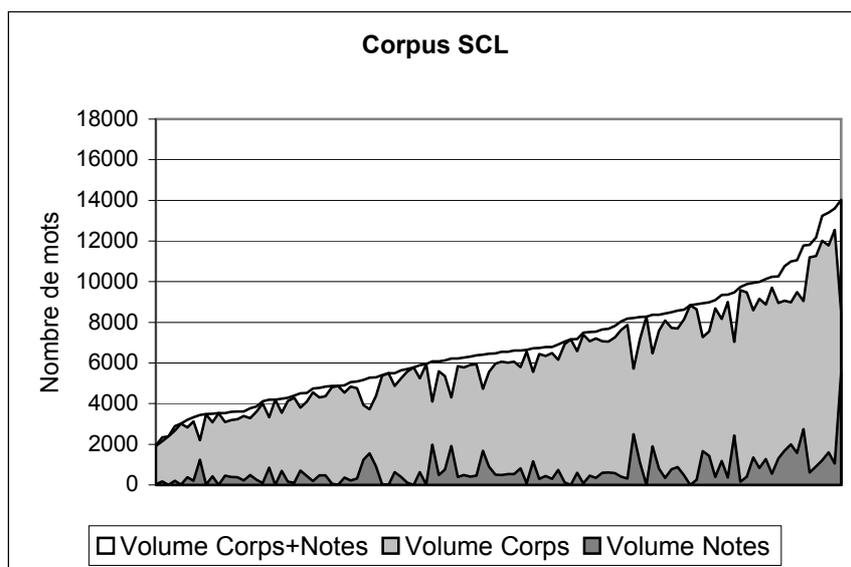


Figure 4.2b. Contribution du corps et des notes au volume total : articles de SCL.

Le volume du corps du texte peut se limiter à 1142 mots et atteindre 13033 mots en LET ; en SCL, il est compris entre 1912 et 12547 mots, mais l'écart-type est à nouveau plus important en SCL qu'en LET où la tendance centripète du genre serait donc plus marquée. Les moyennes et médiane indiquent que cette tendance centrale des mesures faites sur le volume du corps du texte est sensiblement inférieure en LET : le volume moyen du corps du texte, en nombre de mots, atteint 6130 mots en SCL, contre 5875 en LET.

La différence disciplinaire n'est pas statistiquement significative en ce qui concerne le volume du corps du texte. Elle l'est par contre pour les notes : comme le suggèrent les deux graphes ci-dessus, la part prise par les notes dans les articles de LET est significativement plus importante que dans ceux de SCL.

Bien qu'un article de SCL se distingue par un volume maximal des notes supérieur à celui enregistré en LET (5529 mots pour l'article présentant les notes les plus volumineuses en SCL vs 5290 en LET), les valeurs en SCL se caractérisent par une tendance centrale relativement faible par rapport aux valeurs en LET. La distribution en fréquence du volume des notes représentée ci-dessous rend compte de cette différence disciplinaire dans l'usage des notes.

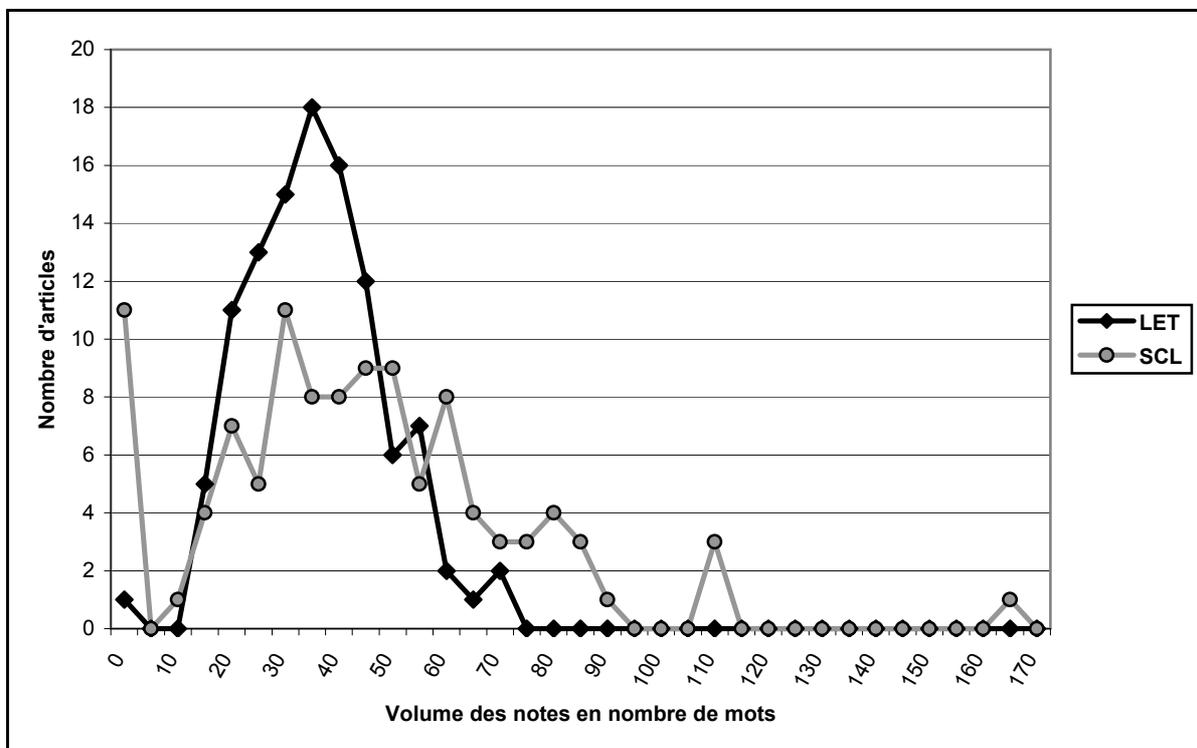


Figure 4.3. Distribution en fréquence du volume des notes : comparaison disciplinaire.

Le volume moyen des notes d'un article de SCL (693 mots) est près de deux fois inférieur au volume moyen des notes en LET (1259 mots). C'est en SCL que l'on observe une tendance centripète plus grande, mais la variabilité intra-disciplinaire du volume des notes reste vraie dans les deux disciplines. Elle s'explique à la fois par des valeurs particulièrement élevées, et par la possibilité pour un article de ne comporter aucune note ; cette possibilité reste toutefois très marginale en LET (un seul article ne comprend aucune note), mais n'est pas négligeable en SCL où elle représente près de 7,3% (8 articles) du corpus.

A propos de la contribution respective du corps du texte et des notes au volume total, on notera d'abord que, contrairement au volume des notes, le volume du corps du texte présente une distribution proche de celle observée pour le volume cumulé du corps du texte et des notes (cf. figure 4.1. *supra*). Le test de corrélation entre les deux séries de mesures montre que le volume cumulé du corps du texte et des notes dépend étroitement de celui du corps du texte (avec un taux élevé de 0,945), mais qu'il n'est pas corrélé au volume des notes. Ces résultats n'ont rien d'étonnant : ils sont conformes à l'image que l'on peut se faire du genre de l'article, comme texte composé principalement de ce que l'on appelle ici le corps du texte, et secondairement des notes.

La contribution des notes, parfois nulle on l'a dit, représente en moyenne 10,16% de l'article en SCL et 17,65% en LET. Elle se limite à moins de 8% du volume total pour plus de la moitié des articles de SCL (57 exactement, contre 15 articles de LET). En LET, elle peut aller jusqu'à déplacer en partie l'image que l'on se fait du genre de l'article, et l'idée d'un texte principal (le corps du texte) et d'un texte quantitativement secondaire (les notes) : comme le suggèrent les figures 4.2a. et b. *supra*, qui rendent compte de la contribution respective du corps et des notes au volume total, un phénomène tout à fait singulier se manifeste en LET, pour l'article dont le volume des notes est le plus important : il atteint 49,99% de l'article, et se trouve donc être quasiment identique à celui du corps du texte à 3 mots près. Unique dans le corpus, ce phénomène ne contredit donc pas complètement l'image du genre, mais il reste qu'il n'est pas observé en SCL, si ce n'est en tant que tendance dans un article, le plus volumineux du corpus (soit sur la droite du graphe 4.2b. *supra*), où les notes représentent tout de même 39,4% du volume total.

Si le genre de l'article est indépendant de la discipline pour ce qui concerne le volume du corps du texte – qui intéresse principalement nos analyses – et pour le volume cumulé du corps du texte et des notes, qui dépend étroitement du premier, chaque discipline investit cependant l'espace des notes de manière très différente ; elles prennent une place plus importante dans les articles de LET, où elles représentent en moyenne 17,6% de l'article contre 10,2% en SCL, mais elles sont soumises à une importante variabilité à l'intérieur de chaque discipline. Quoique de manière moins atypique, le volume du corps et, corrélativement, le volume cumulé du corps et des notes varient également beaucoup au sein de chaque discipline, jusqu'à un rapport de 10 en LET et de 7,2 en SCL entre les articles contenant le moins de mots et ceux en contenant le plus.

Cette variabilité n'est sans doute pas sans lien avec les revues : c'est d'abord au niveau du dispositif éditorial que sont édictées les normes relatives au volume de l'article ; c'est même là bien entendu une caractéristique forte des contraintes éditoriales que de systématiquement spécifier le nombre de mots maximum attendu. Il importe donc de déterminer, depuis le corpus, dans quelle mesure le volume des articles varie avec la revue dont ils relèvent.

4.1.3. La variabilité entre revues

Le volume moyen des articles varie d'une revue à l'autre, comme le montrent les deux graphiques suivants, où les revues de chaque discipline sont classées en fonction de la taille moyenne des articles parus dans ces revues (pour les abréviations des noms de revues, se reporter au chapitre 3 et à la partie 3.1., relative au corpus).

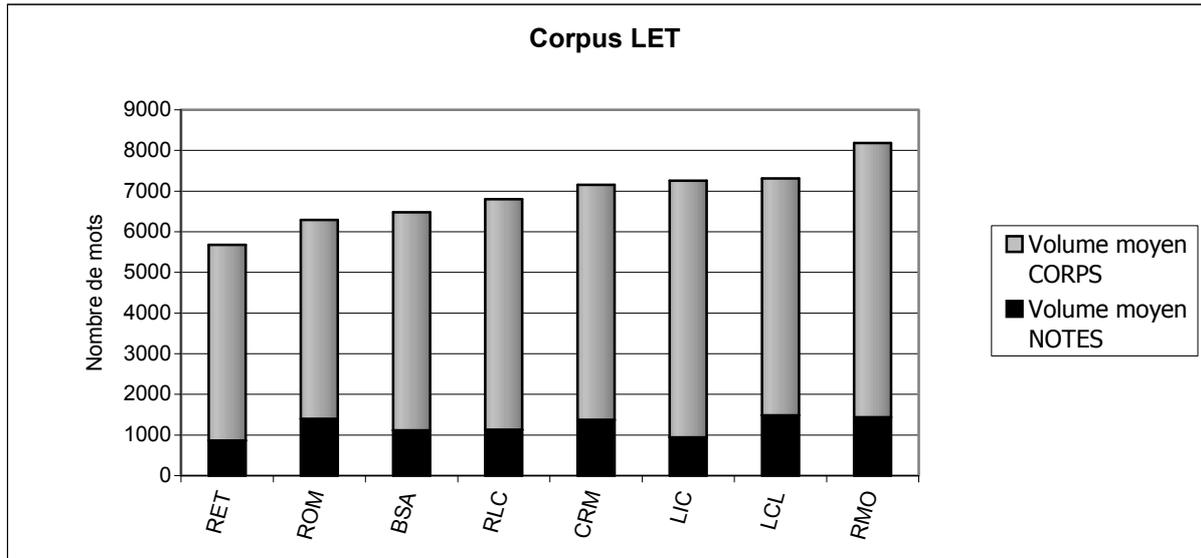


Figure 4.4a. Volume moyen des articles dans les revues de LET.

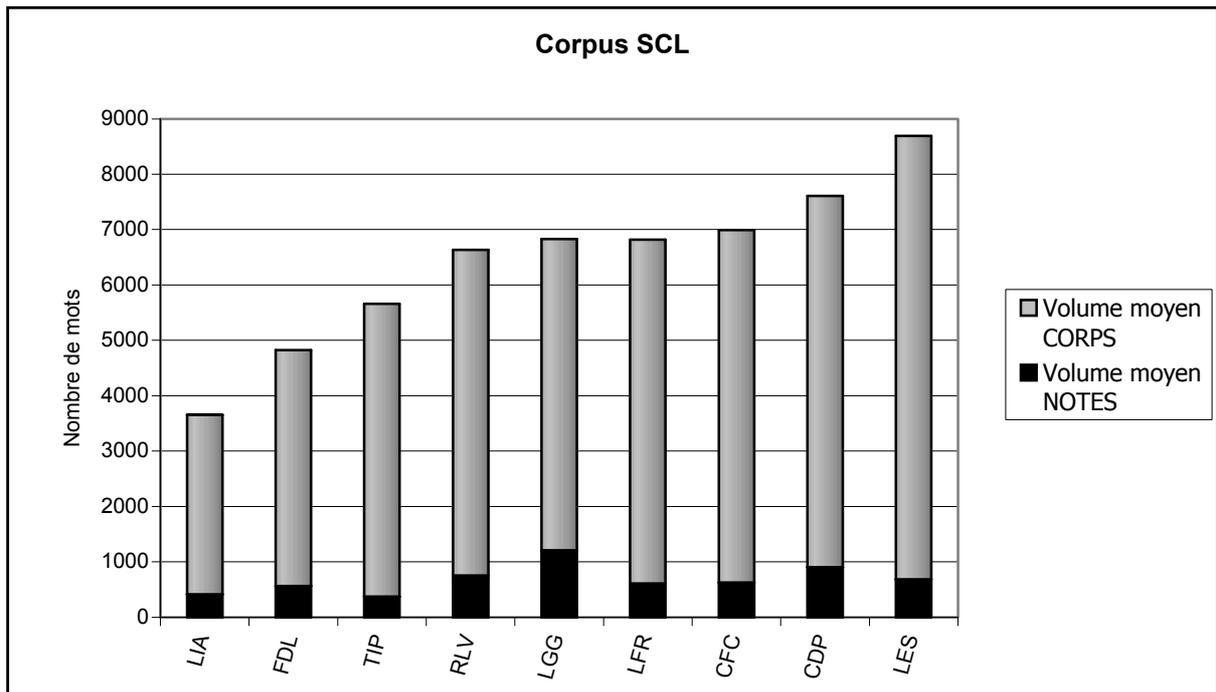


Figure 4.4b. Volume moyen des articles dans les revues de SCL.

Ces deux graphiques rendent compte de la variation d'une revue à une autre au sein d'une discipline : en LET, la revue *Recherches et Travaux* (RET) se distingue par des articles dont le volume moyen est le plus bas avec 5909,7 mots. Inversement, la *Revue de Montesquieu* (RMO) représente dans notre corpus celle dont le volume moyen des articles est le plus élevé avec 8557,6 mots. La *Revue de Littérature Comparée* présente de ce point de vue un profil

intermédiaire, avec le volume moyen le plus proche du volume moyen de l'ensemble des articles de LET.

En SCL, c'est la revue *Liaisons-Airoe* qui se distingue par un volume moyen des articles extrêmement faible par rapport aux tendances centrales dans les deux disciplines (3884,9 mots). Au pôle opposé, la revue *Langages et Société* présente un volume moyen très élevé, là encore par rapport aux autres revues de la discipline et à l'ensemble du corpus (9159,3 mots). Les revues *Recherches Linguistiques de Vincennes*, *Langages* et *Langue Française* sont celles qui représentent le mieux le volume moyen des articles en SCL et dans l'ensemble du corpus.

Ces quelques résultats n'ont pour vocation que de souligner la variabilité observable entre les revues. Cela étant, ils permettent d'envisager l'existence d'un lien entre les revues et le positionnement des articles par rapport aux tendances centripètes et centrifuges du genre. Ils suggèrent en effet que ce sont des revues locales, à tirage plus limité, qui présentent sur cette question du volume moyen les résultats les plus à la marge (supérieure ou inférieure), tandis que ce sont des revues nationales, à plus large diffusion, qui se trouvent représenter au mieux les caractéristiques communes du genre.

Si ces caractéristiques communes montrent le rôle joué par le dispositif éditorial et les exigences économiques qui le sous-tendent¹, il y aurait là une hypothèse à approfondir, quant à l'influence potentielle exercée par les revues les plus diffusées sur la constitution des normes du genre².

Par ailleurs, il convient de souligner que la variabilité inter-revues du volume des articles se comporte de manière différente dans les deux champs considérés : la revue influe de manière significative sur le volume des articles en SCL³, mais n'a statistiquement pas de poids sur le volume des articles de LET. Cet effet des revues propre aux articles de SCL rejoint l'observation faite *supra* d'une tendance centripète du genre plus importante en LET, et d'une tendance davantage centrifuge du genre en SCL, au moins pour ce qui concerne le volume du corps du texte et de l'article.

Nous concluerons cette analyse de la variabilité du volume des articles entre les différentes revues en notant qu'elle ne doit pas masquer la variabilité observable à l'intérieur de chacune d'elles. Autrement dit, la variable revue n'épuise pas la variabilité inhérente du genre. Prenons pour exemple de cette forte variabilité la revue de LET *Bulletin de la Société des Amis de Montaigne* : le volume moyen de ses contributions, de 6759,5 mots, est assez proche de la tendance centrale observée en LET ; cela étant, la revue est représentée dans notre corpus par un article de moins de 1500 mots et un de plus de 14000, et par 4 articles se situant entre 2000 et 4000, 9 entre 4000 et 8000, et 7 entre 8000 et 10000 mots. En SCL où la revue a un effet significatif sur la distribution des mesures faites sur le volume, cette dispersion des valeurs, quoique moins probante si l'on s'en tient au volume effectif de chaque texte, est plus importante encore statistiquement.

Cette variabilité relativise le poids des contraintes éditoriales : on sait que celles-ci fixent tout au plus une limite supérieure, qui ne préjuge pas d'un volume moyen ni d'une limite inférieure, et semble elle-même pouvoir fluctuer. Il faudrait alors analyser la contrainte exercée au niveau du numéro, et évaluer par ailleurs s'il existe une relation entre la taille d'un texte et son statut dans l'économie du numéro ; pour les numéros thématiques en particulier,

¹ Le volume de l'article dépend évidemment de contraintes liées au coût de la livraison, au rapport coût/profit et à la survie de la revue qu'il autorise.

² Un corpus plus important serait vraisemblablement nécessaire, et l'on pourrait tenter de mieux cerner la diffusion d'une revue en associant à l'indice du tirage, qui semble le plus pertinent, d'autres données sur la présence de la revue dans les bibliothèques universitaires, le fait qu'elle soit ou non citée, etc.

³ Kruskal-Wallis, $H = 56,71$, $p < 0,0001$.

on peut penser que certaines études, offrant un cadrage théorique ou une synthèse des travaux existants pourraient au niveau éditorial faire l'objet d'une demande explicite de textes plus longs ou encore d'une autorisation spécifique accordée à leur auteur.

4.1.4. La variabilité liée au statut de l'auteur

Pour finir, il importe de se centrer sur un dernier facteur de variation du volume des articles, le statut de l'auteur. On peut relier cette question à la remarque que nous venons de faire sur le lien entre la taille de l'article et le type de recherche qui y est présenté : en raison de leur statut spécifique, il se pourrait que les doctorants produisent des recherches d'un type particulier, plus ou moins propres à ce groupe au regard de l'ensemble des textes considérés.

Sans permettre de conclure à cela, les résultats montrent toutefois que le statut de l'auteur a effectivement du poids sur le volume des textes qu'ils produisent, mais ce indépendamment de la discipline ; autrement dit, il n'y pas, statistiquement, de différence de volume entre les articles de doctorants et ceux des non doctorants en LET, ni en SCL, mais globalement, le volume cumulé du corps du texte et des notes est plus faible que celui des autres scripteurs : les deux disciplines confondues, on enregistre un volume moyen de 6210,63 mots pour les articles de doctorants et de 7123,68 mots pour les autres.

Si les indices de tendance centrale du volume du texte sont moins élevés chez les doctorants, la variabilité de leurs articles à ce niveau est cependant relativement importante, si bien que la dispersion des mesures est plus grande entre leurs articles qu'entre ceux des autres scripteurs. Rien ne permet de dire que c'est dans le corps du texte ou dans les notes que cette différence du statut de l'auteur se joue, et comme elle ne se vérifie pas dans chacune des disciplines, elle reste finalement difficilement interprétable autrement que dans les termes proposés *supra* d'une spécificité du type d'articles produits par les doctorants.

Cela étant, on pourrait aussi envisager, de manière intuitive, qu'en raison de leur statut, les doctorants produiraient des articles davantage conformes aux contraintes imposées aux auteurs par les revues. Or, la variabilité importante du volume des articles des seuls doctorants interdit une telle hypothèse ; on soulignera en particulier que l'article le plus long produit par un doctorant, noté SCL071, contient 13595 mots ou 85765 caractères espaces compris, alors que la revue où il est publié, *Langage et Société*, demande aux contributeurs de respecter un maximum de 46000 caractères espaces compris, soit près de la moitié moins de l'article concerné. Si ce dernier s'avère être la traduction française d'un article en portugais, il reste difficile de dire si cet écart par rapport aux recommandations est le fait de la seule traduction, ou également de l'original lui-même.

On retiendra donc de l'examen de ce facteur de variation que le statut de l'auteur influe sur le volume total de l'article : cela ne signifie pas nécessairement que les doctorants produisent des textes davantage conformes aux recommandations éditoriales ; par contre, cet effet du statut de l'auteur est peut-être lié à une spécificité quant au type de recherche caractérisant les textes des doctorants. Leurs textes sont en moyenne plus courts, mais il convient de souligner que le statut du scripteur est sans effet sur l'usage – quantitativement parlant – des notes, ni sur la contribution respective du corps et des notes au volume total.

L'analyse du volume du corps du texte et des notes nous a permis de montrer que le corpus dans les deux disciplines est comparable en nombre de mots pour un même nombre d'articles. De manière prévisible, le volume dépend prioritairement de celui du corps du texte, et varie sensiblement sous l'effet du statut de l'auteur, car les doctorants tendent à produire des textes plus courts. Une différence disciplinaire se manifeste au niveau des notes, qui prennent une

place beaucoup plus importante dans les articles de LET. Un tel constat invite à aller en chercher la cause du côté de la fonction de cet espace textuel.

La diversité insoupçonnée des pratiques d'écriture de l'article au niveau de son volume relativise par ailleurs le poids des prescriptions éditoriales sur les usages. Ces prescriptions définissent une limite supérieure qui autorise donc une variabilité importante en deça, mais qui s'avère en outre dépassée dans nombre de cas, y compris chez les doctorants où ce décalage entre les normes prescrites et les normes d'usage est peut-être plus surprenant. Il resterait cependant à déterminer dans quelle mesure ce dépassement, autorisé puisque effectif, résulte d'une demande éditoriale (des responsables du numéro et/ou des relecteurs), et/ou d'une autorisation accordée à la demande de l'auteur.

4.2. Des espaces péri-textuels investis différemment dans les deux disciplines

Nous avons étudié jusqu'ici la question du volume des textes, en nombre de mots compris dans le corps du texte et les notes. Il nous faut cependant aller plus en avant dans cette macro-analyse du genre pour pondérer éventuellement les équivalences et les différences observées entre les deux disciplines et traiter de la manière dont les articles investissent les différents espaces textuels. Nous avons en effet laissé de côté les annexes et les bibliographies notamment, et si nous avons abordé les notes, il est possible de compléter ce que nous en avons dit en associant à leur volume la prise en compte de leur nombre.

4.2.1. Texte et péri-texte : introduction

Le constat selon lequel tout texte est fait de plusieurs textes qui entrent en relation les uns avec les autres est à la base de la notion de « péri-texte » que G. Genette introduit dans *Palimpsestes* (1982) et qu'il développe dans *Seuils* (1987) : elle désigne les titres, sous-titres, notes, dédicaces, etc. autrement dit les éléments qui entourent directement le texte, et s'inscrit dans la paratextualité, c'est-à-dire dans les relations que le texte entretient avec les autres textes qui constituent ses à-côtés. La paratextualité est elle-même une des composantes des relations transtextuelles, ou ensemble des relations du texte avec d'autres, que l'objet de G. Genette est de tenter de mieux cerner.

Cette tentative de décrire les différentes composantes du texte et l'ensemble de ses relations avec d'autres est remarquable, et sert aujourd'hui encore de socle commun à la critique littéraire. Centré sur le texte littéraire, le cadre d'analyse semble cependant aisément transposable, si l'on en juge le large recours aux notions initiées par G. Genette dans diverses études textuelles. Si la question se pose de savoir dans quelle mesure un modèle de textualité et de relations transtextuelles peut être proposé abstraction faite des discours et des genres, il importe pour le moins de voir comment discours et genres s'approprient de manière spécifique ces composantes et ces relations.

Par ailleurs, la référence aux propositions de G. Genette appelle une précision : dans la manière dont elles ont été diffusées, en particulier dans une visée didactique à travers les manuels scolaires, c'est principalement l'identification des différentes composantes du texte qui semble avoir été retenue. Cette identification comporte des limites propres à toute entreprise classificatoire : la démultiplication des instances qu'elle induit nuit à une compréhension unifiée de la textualité. Or, c'est de ce risque que G. Genette vise à prémunir son approche, en entendant suppléer à une analyse des composantes celle de leurs relations.

Notons à ce propos à quel point l'opposition assez communément répandue entre la lecture du document papier et la lecture numérique demande à être nuancée : si l'hypertexte, en tant que structure faite de renvois (ou liens hypertextuels) entre des nœuds signifiants

caractérise le document numérique, les seuls composants péri-textuels, comme les notes par exemple montrent que la lecture du document papier ne saurait être considérée restrictivement comme linéaire.

La structure spécifique du document numérique trouve ses fondements dans une pratique de renvois intra- et intertextuels qui caractérise le texte quel que soit son support de diffusion. Les développements récents des langages de programmation dynamiques, exploités en particulier dans les weblogs et les wikis attestent eux aussi d'une continuité des pratiques de lecture et d'écriture dans l'histoire : la convivialité de production des documents numériques qu'ils autorisent donne en effet lieu à une pratique de glose et d'« entreglose » pour reprendre les mots de Montaigne ; par exemple, le blog se caractérise ainsi par une double composante, le texte de l'auteur du blog et les éventuels commentaires des contributeurs à ces textes, commentaires qui se trouvent repris alors, de manière explicite ou plus implicite dans les productions postérieures de l'auteur. Le lecteur parcourt les liens entre ces espaces au moyen de simples clics, mais on reste proche de la pratique de la glose telle qu'elle s'est développée avec les premiers manuscrits.

Le document numérique permet donc de bénéficier de nouvelles fonctionnalités qui cependant ne font que poursuivre des caractéristiques fondamentales de l'activité d'écriture et de lecture. Plutôt que de concevoir comme allant de soi l'abstraction qui consiste à envisager l'écriture ou la lecture d'un texte pris isolément, il pourrait être intéressant de définir l'activité herméneutique sur la base de cette idée d'entreglose qui caractérise le document numérique dans sa structure, et plus largement le phénomène de l'intertextualité.

Une telle approche permettrait d'éviter les limites de l'opposition hâtive entre document numérique et papier, et les confusions qu'elle peut engendrer entre le texte et son support. Le fonctionnement du document numérique, dans ses constantes évolutions techniques, pourrait alors s'avérer utile pour reconsidérer la notion de texte et certaines distinctions utilisées pour le décrire. Ainsi l'idée d'un intérieur et d'un extérieur au texte, et l'opposition entre des liens intra-textuels et des liens inter-textuels qui en découle est-elle mise en péril par la co-présence virtuelle de l'ensemble des documents numériques mis en réseau. Comment isoler un texte dans un réseau de relations entre textes ?

On refermera cette parenthèse en soulignant que la seule notion de péri-texte n'échappe pas à cette difficulté : elle implique un texte auxquels viennent s'adjoindre des à-côtés ; or, ce « texte » - le corps du texte, comme nous avons choisi de l'appeler – est un élément défini négativement, comme ce qu'il reste de l'article une fois soustraits l'ensemble des éléments péri-textuels. La structuration du texte en différents espaces, et en réseaux de relations entre ces espaces demeure cependant capitale, dans l'optique d'une analyse de la textualité qui tienne compte des différents paliers qui la composent du niveau global au niveau local.

C'est dans cette perspective que nous proposons de rendre compte de la manière dont l'article investit les différents espaces péri-textuels qui entrent en relation entre eux et avec que nous avons nommé le corps du texte. Ce faisant, nous reprenons le terme de « péri-texte » en référence à G. Genette, mais nous limitons grandement l'empan de la notion : il s'agira principalement pour nous d'étudier les annexes, les bibliographies et les notes. Mais nous dirons auparavant quelques mots des résumés d'une part, et du titrage et de la signature d'autre part.

4.2.2. Résumés et abstracts

Les résumés (et/ou abstracts) ont un statut ambivalent au regard de la théorie de G. Genette : ils relèveraient moins du péri-texte à proprement parler que du paratexte de manière plus générale, dans la mesure où ils ne sont pas inclus dans le texte de l'article, mais se situent

dans une partie spécifique du numéro, dédiée à une présentation monolingue ou multilingue des résumés des différentes contributions.

En tant qu'éléments du paratexte, ils posent le problème de savoir comment les considérer en regard de la bipartition proposée par G. Genette entre ce qu'il appelle le paratexte auctorial, i.e. placé sous la responsabilité de l'auteur, et le paratexte éditorial, i.e. placé sous la responsabilité de l'éditeur ; ils suggèrent, comme le note G. Genette à propos d'autres éléments textuels, une dilution de la frontière entre ces deux types de paratexte, et attestent en cela de la complexité du régime de l'auctorialité, du moins dans le cas du texte scientifique qui intéresse la présente étude.

Par ailleurs, comparativement à d'autres éléments paratextuels, les résumés ont ceci de spécifique qu'ils se présentent comme un texte autonome : alors que le résumé a pour fonction de rendre compte sous une forme succincte du contenu de l'article, et d'en fournir ainsi une première approche susceptible d'inciter à la lecture du texte intégral, il n'y a aucune nécessité intrinsèque au résumé de se référer à l'article, dans le sens où le résumé ne comporte aucune instruction à cela. Réciproquement, l'article ne comporte aucun renvoi à son résumé.

En raison de ce statut spécifique d'une part, et d'autre part parce que le mode de constitution du corpus ne nous a pas toujours permis d'accéder à ces textes, nous avons exclu les résumés de nos analyses. De fait, les éléments pris en compte s'inscrivent bien dans ce que G. Genette appelle le péri-texte, au sens d'« à côtés » du texte. Pour autant, nous gardons en vue que si la notion a une assise spatiale, une interprétation adéquate au cadre formulé par G. Genette impose de l'aborder en termes de relations textuelles.

4.2.3. Titre et signature

Le titre et la signature ressortissent, depuis les propositions de G. Genette, de ces éléments péri-textuels¹. Ils s'inscrivent dans l'espace de l'en-tête de l'article, la signature pouvant également être apposée en fin de texte, quoique ce soit manifestement moins souvent le cas, du moins dans notre corpus. Nous n'avons pas retenu cette entrée pour notre analyse du genre ; les travaux de D. Pontille (2001, 2004, 2006) en ont déjà bien décrit le fonctionnement dans le texte scientifique.

O. Ducrot (1984 : 194) montre que la signature a pour fonction d'indiquer à qui est imputée la responsabilité de l'énoncé (dans ses termes, quel est le LOC désigné par *je*) et qu'elle assure une identité entre le LOC et un individu empirique. Dans notre corpus de textes signés par un seul nom d'auteur, elle ne présente pas la richesse analysée par D. Pontille dans des disciplines de sciences dures. L'auteur propose une sociologie pragmatique de l'*authorship* en science sociologique, et interroge la question de la paternité des textes. Si la signature comporte un nom d'auteur, son rôle d'attribution de la responsabilité auctoriale montre toute la complexité de cette notion de responsabilité en sciences : loin de ne renvoyer qu'à un individu, elle peut d'abord, du moins dans certaines disciplines, comporter plusieurs noms d'auteur, et leur ordre indique alors le partage de la responsabilité², dans la production du texte mais aussi de la recherche dont il rend compte³.

Plus largement, le nom d'auteur ne va pas sans mention de son rattachement institutionnel, de sorte que la caution ainsi signée se présente comme nécessairement

¹ Nous ne traiterons pas ici des titres de parties : bien qu'ils relèvent également, selon les propositions de Genette, des éléments péri-textuels, nous les aborderons *infra*, lorsqu'il sera question de la structuration interne du texte.

² Et, partant, des bénéfices de cette responsabilité en termes de visibilité et de reconnaissance dans le champ.

³ Les descriptions de D. Pontille montrent que la signature regroupe des noms d'auteurs dont les fonctions sont extrêmement variées, de celui qui a contribué à l'expérimentation, à celui qui sert de caution scientifique en termes de paternité intellectuelle, en passant par scripteur effectif du texte.

collective, et c'est cette thèse d'une auctorialité scientifique définie sous le sceau du collectif que défend D. Pontille (2004).

Notre corpus vérifie cette règle : comme le titrage, ces éléments que représentent le nom d'auteur et les précisions quant à son rattachement (en général, laboratoire ou université) sont des caractéristiques figées du genre, communes à tous les articles. De cette première propriété commune aux titres et aux signatures en découle une seconde, qui tient à la spécificité du rapport de dépendance entre ces espaces et le reste du texte.

Ils interviennent dans un en-tête, souvent repris d'ailleurs à chacune des pages, ce qui montre l'importance prise au niveau éditorial par la question de l'identité du texte, et celle pointée par O. Ducrot d'une identité entre le responsable de l'énoncé tel qu'il s'y présente dans son sens même (dit LOC), et un individu empirique. Ces éléments de titre et de signature représentent donc un espace textuel à part entière, dont l'importance est mise en avant par la publication des textes, qui en fait une composante obligatoire.

Le rapport de dépendance qu'ils entretiennent avec le reste du texte est plus spécifiquement un rapport d'identité ; dans le cas spécifique de la signature, les renvois qu'elle appelle depuis le reste du texte tiennent presque exclusivement à la marque du *je* du locuteur (LOC), et à l'identité qu'elle implique entre le responsable d'un énoncé particulier et la responsabilité auctoriale du texte. Les formes d'auto-citation sont une autre forme de renvoi, qui se fonde sur l'identité entre le nom de l'auteur cité et le nom de l'auteur du texte, identité basée sur la propriété de désignateur rigide des noms propres.

4.2.4. Annexes

Si le titrage et les signatures sont un trait immuable du genre de l'article, imposé par son éditorialisation et sa diffusion, les annexes au contraire représentent une caractéristique optionnelle. Rien ne permet pour autant de dire qu'elles relèvent de la seule responsabilité auctoriale, et de choix propres au scripteur : il est possible en effet qu'elles résultent d'une demande faite par les relecteurs, ou encore d'une contrainte dictée par les responsables éditoriaux dans le cadrage thématique du numéro et/ou dans la commande faite à l'auteur le cas échéant.

Cette possibilité d'annexes offerte dans les articles de LET comme dans ceux de SCL est cependant davantage exploitée par les seconds. La proportion d'articles comprenant des annexes est ainsi de 7,3% en LET (soit 8 articles), alors qu'elle représente près d'un quart du corpus de SCL (24,5% des articles, soit 27).

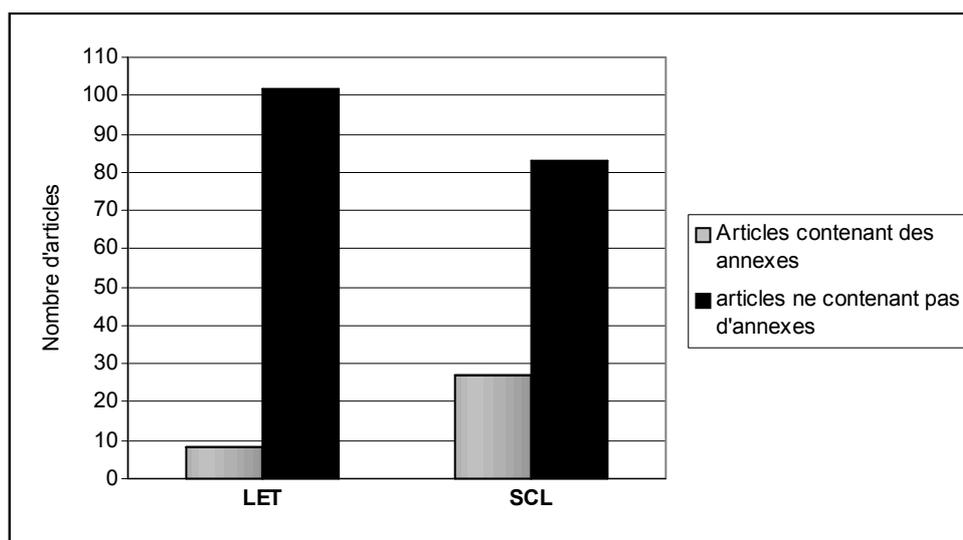


Figure 4.5. L'usage des annexes dans les deux disciplines.

Comme le suggère le graphique ci-dessus, l'option textuelle que représentent les annexes reste donc assez peu en usage en LET, et d'un usage modéré en SCL. La différence disciplinaire s'avère ici statistiquement significative. Le recours plus abondant aux annexes en SCL invite alors à relativiser les résultats évoqués à propos du volume des articles de LET et de SCL. Nous avons montré que les articles des deux disciplines comprenaient un nombre de mots comparable, quoique légèrement supérieur en LET, en excluant cependant les annexes de ce comptage pour s'en tenir au corps du texte et aux notes. La présence plus importante d'annexes en SCL doit ainsi compenser la différence, ou engendrer un volume plus important des articles de SCL.

Une autre différence, de nature qualitative, se manifeste entre les deux disciplines. L'apport documentaire des annexes dans les articles des deux disciplines se traduit par la présence de textes ou de documents iconographiques en LET, et par la présence en SCL de matériaux plus variés (productions d'élèves, transcriptions d'entretiens, textes, liste d'occurrences d'un morphème ou d'un syntagme en corpus, documents iconographiques – à hauteur d'une seule occurrence comme en LET). On remarque donc une variabilité plus forte au sein du corpus de SCL, déjà observée sur d'autres dimensions structurelles du texte.

Notons pour clore ce rapide examen des annexes que leurs fonctions se rattachent dans les deux disciplines aux « matériaux » de la recherche, au sens de données produites dans le cadre ou non de la recherche qui les prend pour objet. Il peut s'agir de données langagières (textes, entretiens, etc.) ou iconographiques d'une part, et de métalangage sur ces données d'autre part, notamment quand sont consignés en annexes des tableaux de résultats.

Il nous reste à présent à analyser les notes et les bibliographies, qui manifestent également des différences disciplinaires fortes, dans la manière dont les articles investissent ces espaces et organisent les relations entre ces espaces et le corps du texte.

4.2.5. Notes

L'analyse du volume des articles en nombre de mots nous a permis de mettre en évidence une différence importante entre les articles de LET et ceux de SCL. Nous avons montré que les notes représentent une part du texte beaucoup plus volumineuse en LET, mais il est possible d'approfondir la question de l'investissement que chaque discipline fait de cet espace textuel en traitant également du nombre de notes.

Comme nous l'avons dit, seul un article de LET ne contient aucune note, contre 8 articles de SCL. L'absence de notes est donc tout à fait marginale en LET puisqu'elle n'est pas même le fait de 1% des articles du corpus, alors qu'elle est moins négligeable en SCL où elle atteint 7,3%. Cela étant, il apparaît clairement que dans une discipline comme dans l'autre, le genre de l'article se caractérise par cette pratique des notes.

En nombre de mots les notes représentent rappelons-le 17,6% du texte en moyenne en LET, contre 10,2% en SCL. Le nombre de notes quant à lui est en moyenne de 38,9 notes en LET et peut atteindre 117 ; en SCL, il est beaucoup plus faible puisqu'il se situe en moyenne à 14,7 notes, et peut aller jusqu'à 85.

Ces résultats sur le nombre de notes vérifient ceux mis en valeur à partir de leur volume. C'est ce que montre le graphique suivant, qui rend compte de la distribution en fréquence du nombre de notes par articles dans les deux disciplines.

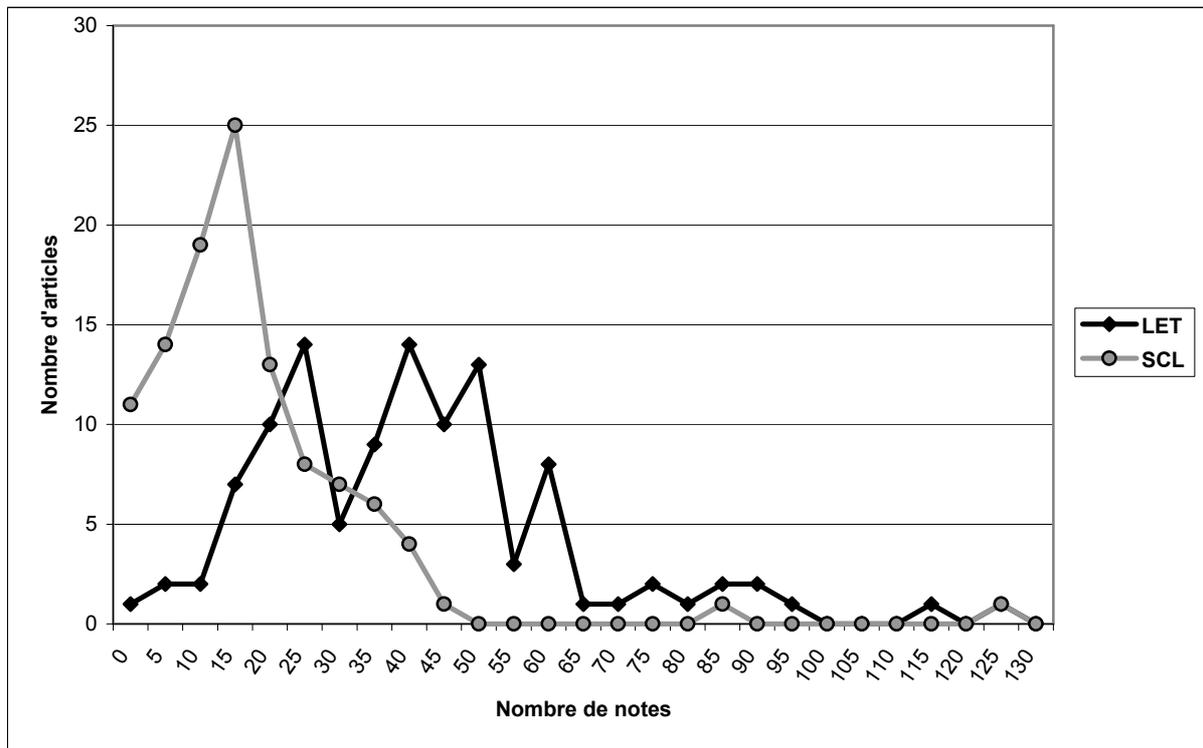


Figure 4.6. Distribution en fréquence du nombre de notes : comparaison disciplinaire.

Si pour des raisons de lisibilité nous n'avons pas intégré à ce graphique les courbes concernant le volume des notes en nombre de mots (cf. la distribution en fréquence du volume des notes *supra*), on remarque que la distribution des mesures concernant le nombre de notes s'avère relativement proche de la distribution concernant le volume. La corrélation entre les deux séries de valeurs est d'ailleurs relativement importante puisqu'elle dépasse 75%.

Les valeurs pour le nombre de notes en SCL se caractérisent par une tendance centrale relativement faible par rapport aux valeurs en LET : le pic en SCL se situe entre 10 et 15 notes et les articles présentant plus de 40 notes ne sont qu'au nombre de 6. En LET, les valeurs se répartissent de manière plus régulière, autour d'une tendance centrale qui s'étend entre 20 et 55 notes ; la proportion d'articles qui comporte plus de 40 notes est de 54,5%, soit 10 fois plus qu'en SCL.

L'étude du nombre de notes confirme donc les résultats obtenus à propos du volume : en nombre de notes aussi, la différence entre les deux disciplines s'avère significative¹. Les notes constituent donc un espace textuel investi très différemment dans les deux disciplines, et davantage dans les articles de LET où les notes sont beaucoup plus nombreuses, et où elles représentent, en nombre de mots, une proportion beaucoup plus importante du texte ; il nous faudra donc mettre en relation ce constat avec la gestion du renvoi aux sources.

Les données croisées du volume et du nombre de notes permettent de mieux cerner encore l'espace des notes, puisqu'il est possible de calculer un volume moyen de la note dans l'article de LET et de SCL. Bien entendu, ce volume moyen est, comme toute moyenne, à manier avec précaution : il ne doit pas masquer la variabilité qui existe d'une part d'une note à l'autre – ne serait-ce que parce que les fonctions que peuvent remplir les notes sont multiples, et d'autre part entre les articles y compris au sein d'une même discipline.

Cela étant, il fait apparaître un résultat tout à fait intéressant, car il montre qu'une note est en moyenne plus longue dans les articles de SCL que dans ceux de LET : alors que les notes représentent une part du texte plus importante et qu'elles sont plus nombreuses dans les articles de LET, une note totalise en moyenne 47,3 mots en SCL contre 32,4 mots en LET.

De manière peu surprenante, on ne trouve que très peu d'articles de LET ou de SCL pour lesquels le volume moyen d'une note reste au-dessous de 15 mots. Hormis ce trait commun aux deux disciplines, la distribution des mesures montre que les valeurs en LET sont davantage resserrées autour d'un volume moyen des notes compris entre 30 et 35 mots ; en SCL, on observe une plus grande disparité des valeurs, qui tient à l'absence de notes dans 11 articles, mais aussi à la présence de valeurs élevées. Le volume moyen d'une note par article peut atteindre 164,7 mots en SCL alors que le maximum en LET est de 70 mots, et si deux articles de LET dépassent un volume moyen de 65 mots par notes, c'est le cas de 18 articles en SCL. Ces valeurs élevées jouent sans doute beaucoup sur la tendance centrale, et sur l'effet significatif de la discipline sur ce volume moyen de la note dans l'article.

L'étude des notes d'un point de vue quantitatif nous permet au final de formuler un triple constat : la pratique des notes est d'usage dans les deux disciplines, mais il s'agit d'un espace davantage investi en LET, à la fois au niveau de la présence ou non de notes, (puisque en LET l'option que représente l'absence de notes est extrêmement marginale), du nombre de notes dans l'article, et de part prise par les notes dans l'ensemble du texte, en nombre de mots. Cela étant, la note est en moyenne de longueur plus importante dans l'article de SCL. Ces différences disciplinaires, statistiquement significatives, suggèrent que les notes ne remplissent pas la même fonction dans chacune des deux disciplines. L'examen des bibliographies va pouvoir nous en dire plus long.

Auparavant, soulignons que les choix éditoriaux attestent d'une nette préférence pour les notes infrapaginales, plutôt que pour les notes situées en fin de texte, ce que l'on interprétera comme un souci ergonomique : par définition, les notes se caractérisent par une dépendance à l'égard du corps du texte et, sans nier qu'elles puissent être ignorées ou qu'à l'inverse le lecteur puisse choisir de ne lire que cet espace spécifique, leur positionnement en bas de page est un choix facilitateur pour une lecture *a priori* fondée sur des va-et-vient.

4.2.6. Bibliographies

Les bibliographies ont une fonction plus limitée et plus aisément identifiable que les notes : elles s'inscrivent dans le système de références intra-textuelles, système qui sert le renvoi aux

¹ Anova, $F = 100,937$, $p < 0,0001$.

sources ; leur fonction de compilation des références ne représente cependant qu'une alternative à la référencement exhaustive des sources, l'autre consistant à les faire figurer en notes.

Or, l'usage des bibliographies fait état à nouveau d'une différence disciplinaire statistiquement significative¹, dont rend compte le graphique suivant.

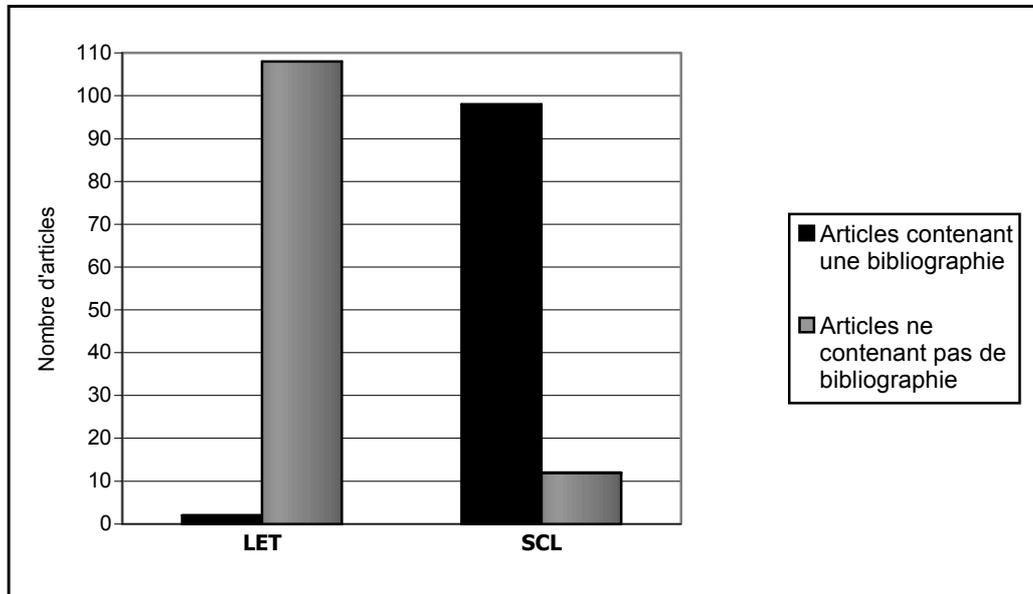


Figure 4.7. L'usage des bibliographies dans les deux disciplines.

Les tendances dans les deux disciplines sont ici diamétralement opposées : seuls 1,8% des articles de LET comprennent une bibliographie, alors que c'est le cas de 88,2% des articles de SCL.

En SCL, la bibliographie représente donc une contrainte forte du genre : son absence relativement marginale, quoique non négligeable (11,8%) peut s'expliquer soit par le choix alternatif d'apposer les références exhaustives d'un document en notes, soit par une absence de référence explicite ; cette dernière serait étonnante, puisqu'il s'agit là d'une caractéristique attendue du discours scientifique, mais elle n'est cependant pas exclue.

En LET, c'est l'absence de bibliographie qui se présente comme une norme du genre : même si l'on admet qu'un certain nombre d'articles peuvent ne pas contenir de références, cela ne saurait suffire à expliquer la proportion de 98,2% d'articles ne comprenant pas de bibliographie. Un examen du corpus le confirme : c'est bien du côté de l'alternative entre notes et bibliographie que l'on peut trouver une explication à cette différence disciplinaire.

¹ Test de Chi2 = 165,748, p < 0,0001.

4.2.7. Une alternative disciplinaire entre notes et bibliographies dans la gestion des références

L'usage opposé des notes et des bibliographies pour la gestion des références dans les deux disciplines permet d'abord de revenir sur la question du volume : restreint au corps du texte et aux notes, l'article comporte un nombre de mots comparable en LET et SCL, et légèrement supérieur en LET, mais avec la prise en compte des bibliographies, on peut penser que cette différence est abolie, voire renversée, comme nous l'avons dit aussi des annexes.

Le constat de l'usage diamétralement opposé des bibliographies permet par ailleurs d'approfondir ce que nous avons dit des notes : à défaut de bibliographie en LET, ce sont les notes qui remplissent une fonction de référencement exhaustive des sources. Cette fonction supplémentaire des notes de LET peut expliquer pour partie au moins qu'elles soient plus nombreuses qu'en SCL, et plus volumineuses en termes de part prise par les notes dans l'ensemble du texte. Cela permet également de comprendre qu'elles sont en moyenne moins longues en nombre de mots, car cette fonction ne requiert vraisemblablement pas autant de développements que d'autres fonctions remplies par les notes dans les deux disciplines.

En SCL, les notes peuvent également remplir cette fonction de référencement des sources : ce serait le cas dans les articles comportant des références mais pas de bibliographie, mais ce constat se vérifie aussi plus largement, dans des modalités de renvoi aux sources qui associent appel de notes dans le texte, référence minimale en notes (du type nom d'auteur et date), et référence exhaustive en bibliographie.

Les différentes stratégies d'exploitation des espaces textuels pour la gestion des références peuvent être représentées à l'aide du schéma suivant. Il rend compte de l'indexation des propos à une source bibliographique au moyen de la chaîne de renvois qui va du corps du texte aux références exhaustives, placées soit en notes soit en bibliographie. Deux types d'indicateurs de renvois peuvent intervenir : les appels de notes, qui ne comportent pas d'autre indication que le renvoi à cet espace du texte, et les indicateurs de références, ou références minimales, typiquement l'association du nom d'auteur et de la date de l'ouvrage.

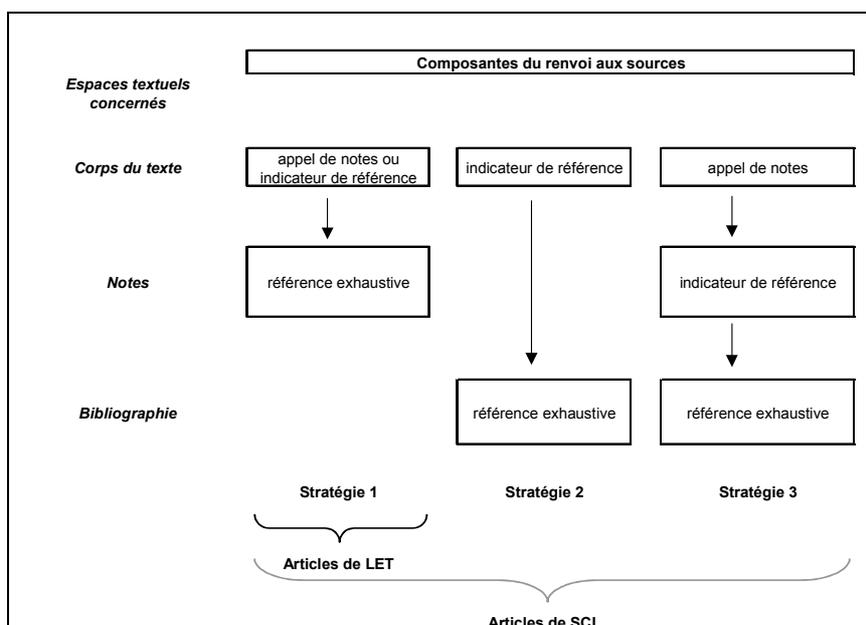


Figure 4.8. Les stratégies d'exploitation des espaces textuels pour la gestion des références.

Les articles de LET recourent de manière presque exclusive à la première stratégie, représentée sur la gauche du schéma, tandis que les articles de SCL, qui adoptent massivement la bibliographie, exploitent l'ensemble de la palette de stratégies. La stratégie privilégiée en LET est en effet également attestée en SCL, où l'absence de bibliographie, quoique rare puisqu'elle ne concerne que 11,8% des articles, est une possibilité moins marginale que sa présence en LET.

Le schéma montre ainsi comment se structurent des réseaux entre espaces du texte pour la gestion des références, au-delà de la seule différence quant à l'utilisation disciplinairement marquée des notes et des bibliographies concernant les références exhaustives. L'opposition entre les deux disciplines sur la fonction de référencement bibliographique alternativement remplie par les notes ou la bibliographie et la mise en évidence de ces réseaux appelle cependant à approfondir leur analyse. Nous traiterons donc pour finir de la dimension fonctionnelle des relations entre les espaces textuels en fonction de leur emplacement.

Nous avons dit que les notes se trouvaient très majoritairement placées en bas de page. Cette norme du genre s'impose au niveau éditorial en raison de son caractère facilitateur dans la lecture :

- d'une part pour leur fonction de référencement bibliographique, qu'il s'agisse de signaler par un appel de note des références minimales, qui renvoient à leur tour à la bibliographie, comme le nom d'auteur et la date, ou qu'il s'agisse d'indiquer les références exhaustives, qui remplissent alors une fonction de référencement bibliographique comme c'est très souvent le cas en LET, où la bibliographie n'est un espace du texte que de manière extrêmement marginale ;

- d'autre part pour une fonction plus étendue et que nous appellerons de « commentaire », dont la première fonction de référence ne serait qu'un cas, en tant que commentaire méta-énonciatif (Authier-Revuz, 1995) consistant à signaler l'indexation des propos à une source autre.

Il est à souligner que les deux ne sont pas exclusives au sein d'une même note, et que le commentaire apporté dans la note peut comporter des références¹. La note ouvre alors à son tour un espace de commentaires et de prolongements d'elle-même, de même que cette possibilité reste inscrite dans le corps du texte d'inclure un commentaire - de type méta-énonciatif par exemple - sans recourir à l'espace des notes.

Concernant l'emplacement des bibliographies, leur fonction compilatrice de référencement exhaustive *via* un classement alphabétique requiert un espace spécifique, qui dans la grande majorité des cas succède au corps du texte et, le cas échéant, précède les annexes.

Cela étant, si la bibliographie appartient à l'article en tant qu'espace péri-textuel dédié à la gestion des sources, il arrive qu'il y ait une seule bibliographie commune pour l'ensemble du numéro. Ce choix semble pouvoir se justifier dans le cas de numéros thématiques, où un univers de référence partagé existe entre les articles et se manifeste par des références bibliographiques communes ; la compilation des différentes bibliographies en une seule participe de l'économie du numéro, et permet d'éviter les redondances et de mettre l'accent sur l'unité thématique.

Cette alternative laissée aux responsables éditoriaux du numéro de faire de la bibliographie un espace commun aux différents textes révèle le statut spécifique de l'article

¹ Il faudrait ajouter à cela qu'elles peuvent elles-mêmes être à l'origine de commentaires.

par rapport au numéro dans lequel il s'inscrit, au moins dans le cas des contributions à des numéros thématiques. Sans remettre en question la frontière entre les textes, la bibliographie commune montre la difficulté qu'il y a à penser le texte comme un espace fermé clos sur lui-même. Elle valide l'idée que les espaces péri-textuels représentent des à-côtés du texte, qui cependant peuvent être investis par d'autres textes et demandent donc à être analysés en termes de réseaux entre textes plutôt que comme composantes d'un texte pris isolément.

En LET où la bibliographie n'est pas en usage, un réseau similaire peut se constituer entre les différentes contributions d'un numéro : chaque texte recourt à un code commun à la livraison, qui consiste à utiliser une abréviation pour référencer de manière économique le titre des ouvrages centraux dans la thématique du numéro.

Qu'en est-il finalement de la fonctionnalité de ces diverses stratégies que sont la référencement en notes, la référencement en bibliographie, ou le recours à un système de renvois intratextuels plus complexe, qui associe notes et bibliographie ?

Le fait d'indiquer les références exhaustives en notes de bas de page doit faciliter leur lecture, par rapport au recours à la bibliographie en fin d'article ou, dans certains cas plus rares, en fin de numéro. A contrario, ces indications en bas de page n'autorisent pas une consultation par entrées alphabétiques de la liste constituée par l'ensemble des références bibliographiques.

L'utilisation des notes pour la référencement exhaustive des sources peut également poser un problème de redondance, quand l'éloignement entre deux notes interdit le recours à une formule du type « ibid. » ; cela étant, les articles de LET contournent généralement ce problème, en recourant à des abréviations des titres des ouvrages consultés¹. La référence exhaustive n'intervient donc qu'une fois, comme c'est le cas avec l'usage des bibliographies ; elle est relayée ensuite par un codage plus économique, qui représente dans les relations entre notes l'équivalent des indicateurs de référence dans les relations entre corps du texte et notes, ou entre notes et bibliographie.

Le fait de préciser des indications minimales (typiquement, nom d'auteur et date) en notes plutôt que dans le corps du texte permet d'éviter une surcharge dans ce dernier, mais produit un système relativement coûteux de renvois qui vont du texte aux notes et des notes à la bibliographie. Cela étant, ce renvoi en trois étapes présente l'avantage de concilier pour le lecteur les deux fonctionnalités des notes de bas de pages et de la bibliographie : l'aisance d'un renvoi aux sources sur la même page, sans la surcharge occasionnée par leur emplacement dans le corps du texte, et la consultation alphabétique de la liste de l'ensemble des références.

Nous reviendrons au chapitre 7 sur la gestion des sources. Abordée ici sous l'angle de la structure des textes, elle montre à partir de l'usage des notes et des bibliographies dans les articles de LET et de SCL qu'une identité disciplinaire de l'article se manifeste dans la manière de gérer la textualité, en termes de les renvois entre les différents espaces textuels. Ne serait-ce qu'au niveau formel de la notation des indicateurs de références et des renvois, il est possible de repérer des stratégies propres à certains domaines ou revues², sinon aux champs disciplinaires eux-mêmes, comme on l'a rapidement évoqué avec l'utilisation en LET

¹ Nous avons vu que ce système d'abréviations pouvait également consister en un code communément utilisé par les diverses contributions d'un même numéro.

² A titre d'exemple, notons qu'en SCL, la revue d'informatique linguistique « Traitement des Langues », publiée par l'ATALA, recourt à des indicateurs de référence formés des trois premières lettres du nom d'auteur mises entre crochets. En physique, dans un article de P.G. de Gennes paru dans les *Comptes-Rendus de l'Académie des Sciences de Paris* et cité dans l'étude de A. Barberousse et E. Guyon (2003), un appel de note renvoie à une bibliographie qui se présente sous la forme de notes de fin d'article. L'appel de note peut se composer d'un seul numéro et renvoyer à une référence, ou une série de numéros (1-3 dans l'exemple donné) pour indiquer un renvoi à plusieurs sources.

d'abréviations des titres d'ouvrage, pour un renvoi rapide entre notes. Le numéro, en particulier dans le cas de numéros thématiques, et la revue déterminent également les caractéristiques de l'article en ce qui concerne les différents espaces du texte et leurs relations.

La fonction de gestion des références à l'œuvre dans le texte appelle une ultime remarque. Nous avons en effet mis jusqu'ici de côté la prise en compte du type de sources référencées dans l'article. Or, il est à noter qu'en SCL, les bibliographes ne recensent que les textes de référence. C'est du moins leur rôle commun et prioritaire, car il arrive aussi, quoique très rarement qu'une seconde bibliographie s'ajoute à la première, avec pour objet de référencer les matériaux de la recherche (par exemple, des textes littéraires). Ceux-ci peuvent également être référencés dans le corps du texte (la partie exposant la méthodologie de la recherche se prête à des précisions de ce type), ou encore en notes de bas de page. On rejoint alors la fonction de référencement exhaustive des notes exploitée en LET, où à l'inverse, hormis dans un article, toutes les sources font l'objet d'un traitement commun.

De telles données sur la macro-structure du genre suggèrent que les articles des deux disciplines confèrent un statut différent aux sources, selon que cette distinction entre types de sources est ou non inscrite dans la textualité. Il nous reviendra d'approfondir ce constat ; c'est ce que nous ferons *infra* à la fin de cette macro-analyse du genre, en traitant des exemples et des citations.

4.2.8. Bilan sur le péri-texte de l'article : un contraste disciplinaire dans les degrés d'éclatement du texte

L'étude du péri-texte fait apparaître qu'au-delà de caractéristiques figées du genre que représentent le titre et la signature, les disciplines s'approprient de manière spécifique les différents espaces textuels qui composent l'article. Le texte de l'article en LET se caractérise principalement par un usage important des notes, à hauteur en moyenne de 17,6% des mots, et par le recours faible aux annexes. Les bibliographies quant à elles sont très rares, les notes remplissant la fonction de référencement exhaustive des sources.

En SCL, la textualité se caractérise par une plus grande hétérogénéité des espaces textuels : les annexes représentent une possibilité attestée dans 24,5% des articles, les bibliographies sont de mise, et les notes occupent une place importante, quoique moindre qu'en LET. Les 3 stratégies de référencement des sources sont représentées, i.e. l'utilisation de la bibliographie, des notes, ou d'un dispositif de renvois qui va du texte aux notes et à la bibliographie.

L'article en SCL se caractérise par une textualité faite d'espaces éclatés, avec des réseaux complexes de relations dans la gestion des références. En LET, la place prise par le corps du texte et les notes tend à épuiser le volume du texte ; recentrée sur le corps et les notes, la textualité est davantage homogène, entre les parties d'un article et entre les articles du corpus.

Ces résultats montrent le poids des différences disciplinaires sur les pratiques d'écriture de l'article, leurs implications sur les modes de lecture (la multiplication des espaces péri-textuels en SCL, ou la centration sur le corps et les notes en LET), et la possibilité d'identifier dans le texte des données d'ordre épistémologique (la distinction spatiale de deux types de sources en SCL, celles produites par les pairs, et le matériau sur lequel se fonde l'analyse et l'absence de distinction dans le traitement des sources primaires et secondaires en LET).

Si la manière dont le genre de l'article investit le péri-texte est marquée par des identités disciplinaires fortes, elle est soumise également à l'effet du statut de l'auteur mais de

manière moindre toutefois. Le fait qu'il soit ou non doctorant joue que sur la présence de bibliographie et d'annexes : leur usage est significativement plus important chez les doctorants. Cela étant, cet effet du statut ne se vérifie que les deux disciplines confondues, et reste assez difficilement interprétable ; mais il est possible de voir dans cet effet du statut de l'auteur des types d'études spécifiques, qui appelleraient un recours aux annexes, et l'indice d'une évolution possible du genre vers un abandon – partiel et sans doute progressif – d'un modèle de texte caractérisé par l'absence de ces éléments péritextuels.

Nous retiendrons donc que du point de vue du péritexte, deux modèles de textualité opposés se manifestent, l'un plus recentré sur le corps et les notes comme en LET, l'autre plus éclaté ; le fait que ce dernier soit préféré par les doctorants pourrait être le signe d'une évolution possible du genre vers ce modèle. Les articles de SCL se situent pour certains à ce pôle, mais caractérisent par une diversité telle qu'ils se répartissent entre les deux selon des configurations variées.

Cette macro-analyse de la textualité demande alors à être complétée par la prise en compte de paliers inférieurs, internes au corps du texte. Il s'agit de voir comment celui-ci est structuré, en traitant de la question des plans de texte.

4.3. L'irréductible diversité des plans de texte : vers un continuum entre deux tendances opposées et disciplinairement marquées

4.3.1. Quelle grille pour l'analyse des plans de texte ?

Dans un certain nombre de recherches consacrées à l'article sa structure est décrite comme canonique (par exemple, dans Heslot, 1983 ou De Nuchèze, 1998) : elle respecterait le plus souvent le plan « IMRD » (*Introduction, Materials and Methods, Results and Discussion*), ou un autre type de plan assez proche. Dans les travaux que nous venons de citer, ce modèle anglo-saxon du plan de texte est décrit comme une caractéristique de l'article de recherche en contexte français.

Or, un examen rapide des articles de notre corpus suffit à s'en convaincre : une telle structure apparaît très rarement. On l'observe sous une forme plus ou moins canonique dans 7 articles de SCL (soit une proportion de 6,4% des articles du corpus), et jamais en LET. Mais ce n'est pas tout : les plans de texte se caractérisent par une extrême diversité. Cette diversité est telle qu'une description commune des articles n'est pas envisageable à ce niveau et rend problématique le seul balisage logique des documents.

Le problème qui se pose est de déterminer quels critères retenir pour décrire la structure de l'article. A défaut de pouvoir utiliser des entrées notionnelles (du type « matériels », « méthodes », « résultats », « analyses » comme dans les formats dits IMRD), il nous faut prendre appui sur des entrées formelles.

Les indices de structuration qu'il semble possible de retenir pour définir la structure de l'article dans les deux disciplines concernées sont d'abord, le titrage des parties et des sous-parties : un élément verbal annonce le contenu de la section ou son statut dans l'ergonomie du texte. De telles indications pourraient servir d'appui à une analyse notionnelle de la structure, si du moins elles s'avéraient aussi régulières que dans les formats dits IMRD.

Le second indice de structuration est la numérotation des sections. Un nombre ou un autre symbole à valeur numérique, sans rien dire à lui seul du contenu même de la partie, la signale en tant que telle et précise la manière dont elle se situe dans la structure hiérarchique du texte. Il est à noter que, dans notre corpus, cette indication peut apparaître indépendamment du titrage de la section.

Le troisième et dernier indice est fourni par la dimension typographique, plus précisément les sauts de ligne qui peuvent, on va le voir, être les seuls indices de structuration. En ce cas, se pose un problème d'ambiguïté : a-t-on affaire au passage d'une section à une autre, ou d'une sous-partie à une autre au sein d'une même section ? Cette ambiguïté peut être levée par l'utilisation dans certains articles d'astérisques, qui fonctionnent comme séparateurs de sections.

La structure des articles a donc été décrite au moyen de ces trois indices : il s'agissait de voir lesquels étaient présents, étant entendu que le premier et le second impliquent le troisième, mais que chacun des trois indices peut intervenir sans les autres. On obtient donc les possibilités suivantes : titrage et numérotation, titrage seul, numérotation seule, ou séparateurs typographiques exclusivement.

4.3.2. Pratique des titres et de la numérotation : la division du texte en composantes structurelles

Nous avons appliqué cette typologie du marquage de la structure aux introductions, aux conclusions, et enfin aux autres sections et sous-sections du texte.

Le premier constat est que les sections et sous-sections utilisent un marquage commun, qui ne préjuge pas de celui choisi pour l'introduction et la conclusion. Ces dernières ne peuvent pas toujours être isolées de parties connexes plus larges : l'introduction se détache cependant plus aisément que la conclusion, si la partie qui la suit est titrée ou numérotée, alors qu'une conclusion non signalée comme telle ni au moyen d'astérisques ne permet pas de décider où se situe son commencement. Dans un cas comme dans l'autre se pose donc parfois le problème d'ambiguïté évoqué *supra* de la segmentation au seul moyen des sauts de ligne.

En dépit de ces difficultés pour les seules introductions et conclusions, il est possible d'identifier de telles parties dans l'article, soit dans le cas dit 1 où ces parties introductives et conclusives sont titrées comme telles, soit dans le cas 2 où, sans titres, leurs bornes initiales et finales sont cependant repérables.

Pour les introductions, le repérage est plus aisé que pour les conclusions on l'a dit, et s'avère plus évident en SCL, où elles sont titrées dans près de 50% des cas comme on le voit sur le graphe ci-dessous. Elles ne portent au contraire jamais de titre en LET, mais restent repérables à partir des indications des parties suivantes. Dans 14 articles de chacune des deux disciplines, rien ne permet cependant de dire où s'arrête l'introduction (cas 3).

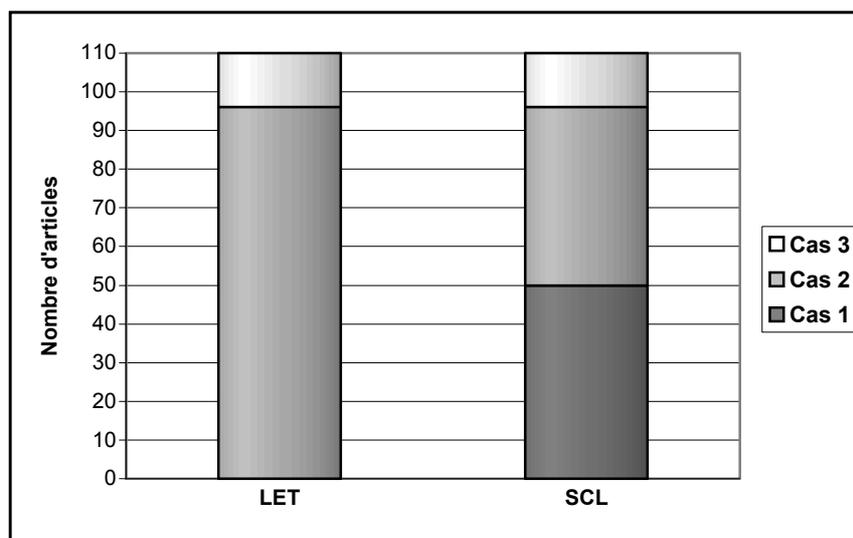


Figure 4.9. Les introductions dans les deux disciplines : titrées (cas 1), identifiables (cas 2), non identifiables (cas 3).

Concernant les conclusions, une différence disciplinaire se manifeste là aussi dans leur titrage : 78 articles de SCL comportent une conclusion titrée comme telle, contre 5 en LET. Elles restent identifiables à l'aide des indications typographiques par exemple, mais dans 38 articles de LET, de même que dans 13 articles de SCL, aucun indice n'est fourni pour déterminer à coup sûr où commence la conclusion.

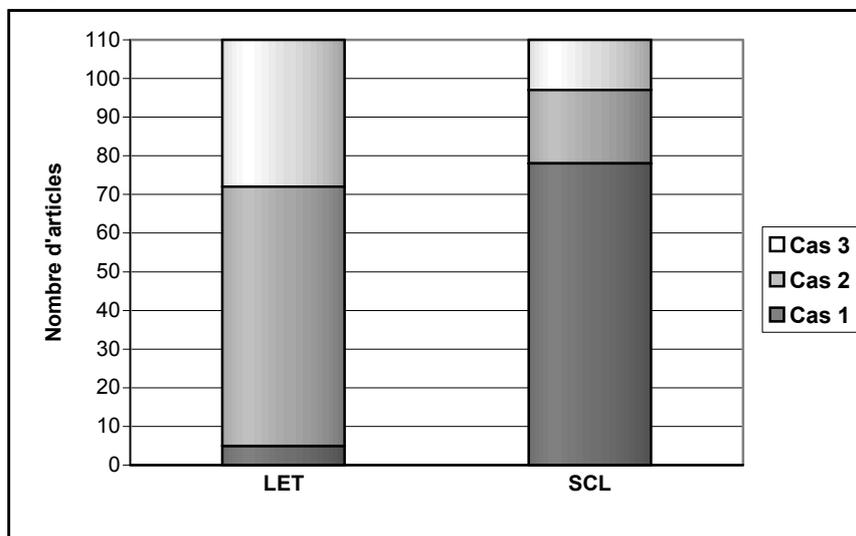


Figure 4.10. Les conclusions dans les deux disciplines : titrées (cas 1), identifiables (cas 2), non identifiables (cas 3).

Ces différences disciplinaires dans l'identification des introductions et les conclusions sont statistiquement significatives¹ (pour les introductions comme pour les conclusions).

Les sections (ou parties de premier niveau) et sous-sections (ou parties de second, troisième, n^{ième} niveau) font état de cette même différence disciplinaire statistiquement marquée², en particulier au niveau du titrage des parties, comme le suggère le graphe suivant :

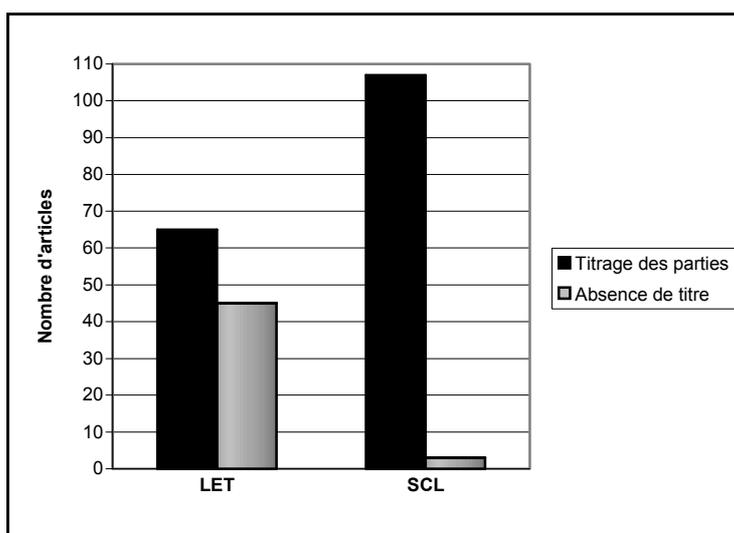


Figure 4.11. Le titrage des sections dans les deux disciplines.

¹ Test de Chi2, $p < 0,0001$ pour les introductions comme pour les conclusions.

² Test de Chi2, $p < 0,0001$.

On voit ici clairement que la pratique des titres pour les sections de premier niveau est d'un usage massif en SCL où elle est le fait de 98% des articles, alors qu'en LET les deux possibilités d'indiquer ou non un titre sont en concurrence, quoique le titrage l'emporte à hauteur de 60%.

La numérotation des sections, qui représente une alternative au titrage ou s'associe à lui renforce l'idée que les sections ne sont pas tant sémiographiquement marquées en LET qu'en SCL :

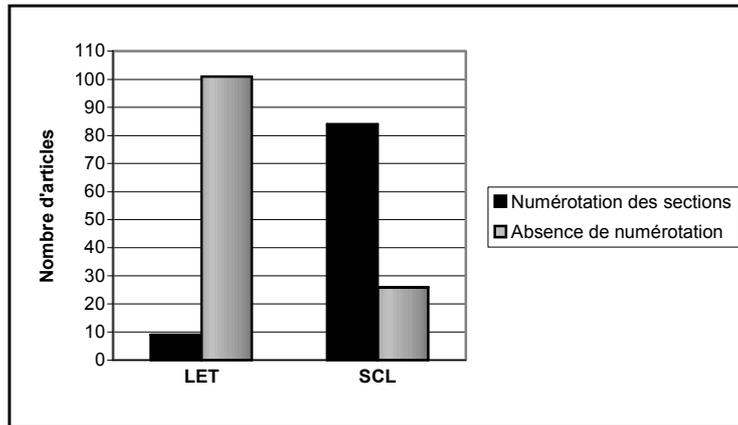


Figure 4.12. La numérotation des sections dans les deux disciplines.

Moins fréquente que le titrage, la numérotation des sections est un usage bien représenté en SCL, avec 76,4% des articles, alors qu'elle est extrêmement rare en LET (8,2%). Si l'on se réfère aux données sur le titrage, on remarque donc que les articles de SCL se caractérisent par des sections titrées, le titre s'accompagnant dans plus de trois-quarts des cas d'une numérotation, et que les articles de LET se caractérisent par des sections titrées dans plus de la moitié des cas, mais le plus souvent non numérotées. Dans 40% des cas, les sections ne sont donc introduites que par des sauts de ligne ; dans ces cas-là, la hiérarchisation des sections se limite à un niveau, et rien n'est dit sur le contenu des sections.

On remarquera ainsi que la pratique des sous-sections (ou parties de second, troisième et etc. jusqu'au $n^{\text{ième}}$ niveau), repérée à partir de leur titrage et/ou de leur numérotation, voire par de simples sauts de ligne est d'un usage très répandu en SCL puisque 82% des articles sont concernés. Elle est au contraire très faible voire marginale en LET où elle ne concerne qu'un peu plus de 7% des articles.

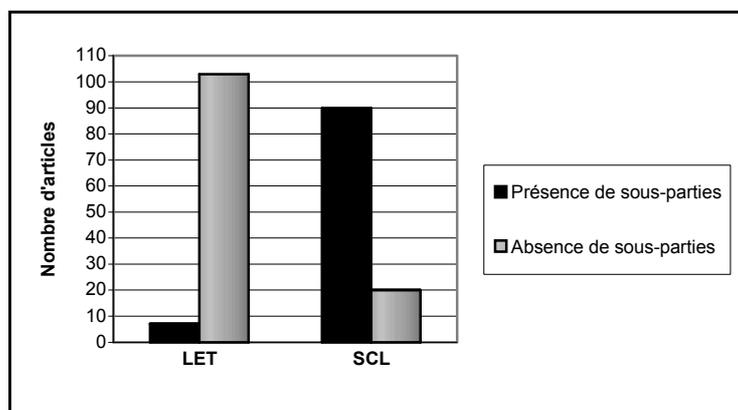


Figure 4.13. L'usage des sous-parties dans les deux disciplines.

Bien entendu, cela ne signifie en rien que les sections des articles de LET ne sont pas structurées en sous-composantes, y compris quand elles ne contiennent pas même de sauts de ligne. La séparation en paragraphes reste utile au guidage du lecteur dans la structuration du contenu, mais c'est là le seul élément autre que le contenu verbal sur lequel il peut s'appuyer. On soulignera que si l'éditorialisation de l'article et les principes « reader-friendly » revendiqués en particulier dans les manuels d'aide à la rédaction scientifique imposent souvent des paragraphes brefs, cette norme ne se vérifie pas toujours ; en particulier, quand les seuls indices de structuration sont des paragraphes, on observe des textes avec des paragraphes extrêmement longs, qui peuvent atteindre plusieurs pages.

Cette possibilité de paragraphes longs est confirmée par l'analyse statistique. Celle-ci manifeste une différence disciplinaire au niveau du nombre de paragraphes dans le texte (étudié à l'aide d'un ratio sur le nombre de mots du texte, pour neutraliser l'effet de cette variable) et de la longueur moyenne d'un paragraphe en nombre de mots. Que ce soit dans le corps du texte, dans le texte (corps et notes), ou dans le corps sans les citations et exemples, le nombre de paragraphes est globalement plus important en SCL, et la longueur du paragraphe est plus restreinte¹. Là encore, les résultats sont plus homogènes en LET qu'en SCL où les modèles de textualité semblent davantage diversifiés.

4.3.3. Un continuum entre deux modèles de texte opposés

L'analyse des indices de structuration aux niveaux des introductions, conclusions, sections et sous-sections permet au final de situer des profils d'articles le long d'un continuum de modèles entre deux pôles opposés, dont nous rendrons compte *infra* au moyen d'un schéma.

A l'un des pôles, interviennent des articles avec des sections titrées et numérotées qui se conforment tendanciellement au plan IMRD : quelques articles du corpus d'articles de linguistique présentent des titres notionnels très proches de ce format (introduction, méthodes, résultats et analyse, discussion et/ou conclusion) ; il s'agit de 7 articles de SCL, et plus précisément de 7 des 13 articles de la revue « Travaux de phonétique de l'Institut de Strasbourg ».

Au pôle opposé, la textualité n'est pas faite de sections ; au contraire, elle constitue une totalité, divisée en paragraphes, ou en séries de paragraphes séparées par des sauts de ligne seulement, et où la longueur parfois des paragraphes suggèrent qu'ils peuvent remplir à eux seuls le rôle de section ; aussi s'éloigne-t-on d'un modèle de textualité se caractérisant par une subdivision en de nombreux niveaux, balisée en sections, sous-sections et paragraphes, en vigueur dans d'autres champs disciplinaires comme le suggère l'utilisation de cette architecture du document dans sa rédaction au moyen d'un logiciel dédié à la rédaction scientifique comme Latex.

Cela étant, ce modèle de textualité reste au regard des autres pratiques d'écriture de l'article de recherche difficile à définir autrement que de manière négative ; on pourrait voir dans la structure de ce type d'articles un format assez proche de celle du genre de la dissertation (cf. les travaux d'I. Delcambre (1990) sur la dissertation scolaire), où l'annonce de plan et l'aménagement de transitions remplacent la pratique de titrage, quand elles ne sont pas déconseillées pour être réputées trop scolaires.

Entre ces deux pôles d'une structure IMRD et d'une structure plus dissertative, et c'est là que se situent la majorité des textes, l'approche présentée suit un plan à entrées thématiques, généralement titrées mais numérotées ou non. On soulignera que les entrées thématiques ne reproduisent que très peu celles que l'on retrouve dans le format IMRD, car hormis pour l'introduction, seule peut-être la partie de méthodologie est affichée en tant que

¹ Cf. les annexes 2.

telle, et dans quelques cas très rares, la partie d'analyse ou de discussion des résultats. De tels intitulés ne se vérifient toutefois qu'en SCL ; mais en SCL y compris, les titres proposés restent extrêmement variés, et si le contenu pourrait dans certains cas être envisagé au moyen des entrées notionnelles propres au format IMRD, il nous faut retenir que ce ne sont pas ces entrées-là qui sont choisies pour titrer les parties.

Les deux pôles ne doivent pas être assimilés à une opposition radicale entre les deux disciplines, quoique les deux disciplines se distinguent clairement quant au format de texte privilégié. Les articles de LET se situent plutôt entre le milieu de l'axe et le second pôle, tandis que ceux de SCL se répartissent entre les deux pôles sur l'ensemble du continuum des modèles de texte, comme en rend compte le schéma suivant.

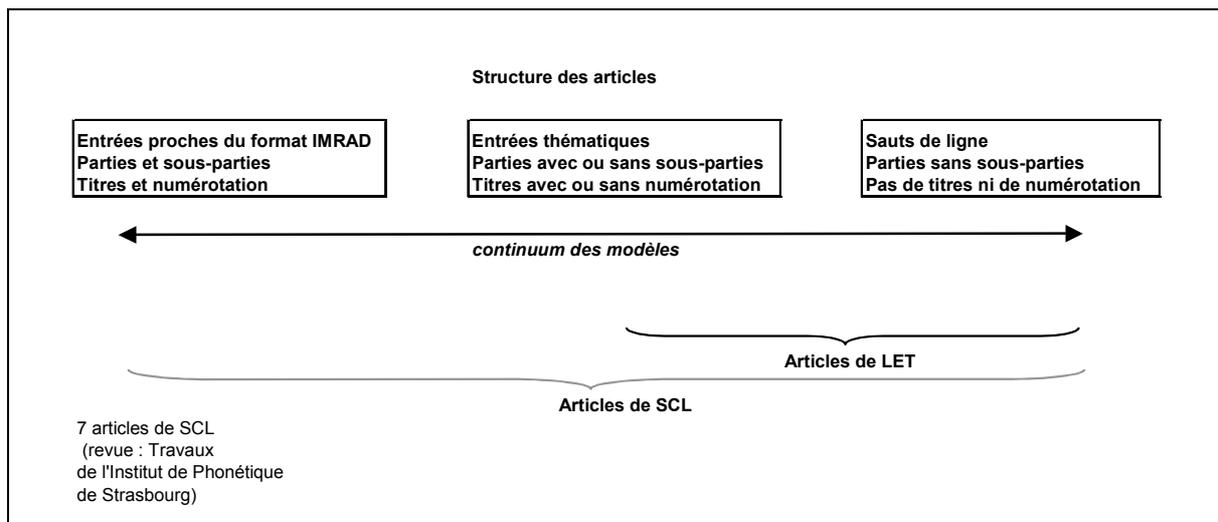


Figure 4.14. La structure des articles : continuum des modèles et comparaison disciplinaire.

Comme pour le périphrase, la structure interne de l'article rend donc tangible non seulement des modèles de textualité différents dans les deux disciplines, mais également une plus grande variabilité de l'article en SCL : il peut prendre les traits de la tendance plus homogène en LET, ou toutes les formes du continuum qui va de cette tendance à la tendance opposée, fondée sur le modèle jugé canonique en sciences du plan IMRD. Celui-ci reste toutefois très peu représenté, et cède la place à une structure interne balisée à l'attention du lecteur, qui comporte des sous-sections, mais où les titres sont extrêmement variés et la numérotation pas systématique.

En LET prédomine un profil de textualité qui serait en quelque sorte le contre-pied du modèle IMRD et plus largement de la subdivision du texte et du guidage du lecteur en vigueur en sciences ; la recherche sur les textes littéraire se présente comme une approche unifiée, peu propre à être caractérisée au moyen des divisions pré-établies de type « méthode », « résultats », « analyse », etc.

On voit donc que le seul plan IMRD est loin de rendre compte des tendances observables dans les deux disciplines. En l'absence de cette structure type, un balisage logique commun aux documents est difficile, car les marques formelles sur lesquelles s'appuyer font état d'une grande diversité dans les modes de subdivision du texte, plus ou moins développés, et dans leur mise en évidence à l'attention du lecteur, à l'aide de titres, de numéros, ou uniquement de séparateurs typographiques.

Aussi la solution adoptée dans le projet KIAP, qui balise la structure de l'article dans trois disciplines en opposant un format IMRD à un format dit non IMRD est-elle la seule envisageable pour décrire l'article à ce niveau. Cela étant, elle ne permet pas de rendre compte de la diversité subsumée par le format non-IMRD, et montre donc la difficulté qu'il y a à ne pas reproduire l'idée d'un format canonique de l'écriture de l'article à partir duquel les autres formats ne peuvent être décrits que négativement.

4.3.4. Structuration de plus bas niveau, interne au paragraphe

Il est intéressant de croiser alors ces éléments concernant la structure de l'article et la visibilité de cette structure avec d'autres éléments sémiotiques remplissant une fonction de structuration de plus bas niveau, interne au paragraphe. On observe en effet la présence dans les articles de tirets, de puces ou de numéros, qui ne servent pas la structuration du texte en sections, mais la présentation structurée d'une liste de propositions par exemple, comme on le voit dans les deux extraits suivants :

Et toutes ces lettres sont traversées par des effets de réversibilité similaires, qui infléchissent notre compréhension des héros :

- Réversibilité du mouvement même de notre lecture (...)
- Réversibilité du point de vue choisi (...)
- Réversibilité de la fiction (...)
- Réversibilité du contenu de la lettre (...) [LET026]

Ainsi, la division de la première partie de L'Esprit des lois en livres et en chapitres, est le résultat de la mise en œuvre de trois ensembles de distinctions :

1. La typologie tripartite ;
2. La distinction entre la nature (la constitution politique) et le principe (l'ensemble des dispositions d'un peuple qui permettent à un gouvernement d'exister) d'un gouvernement ;
3. La distinction entre le droit public et le droit civil ou criminel. [LET048]

Le scripteur a ainsi divers recours pour mettre en valeur ses propositions en exploitant la dimension visuelle du texte : il peut s'agir simplement de les distinguer à l'aide de retours-chariots éventuellement couplés à des tirets comme dans le premier extrait, ou d'utiliser des symboles qui associent à la fonction démarcative une valeur ordinale, comme a), b), c), i, ii, iii, ou encore des nombres comme dans le second extrait.

Cette mise en forme des propositions par des indices de structuration permet d'assurer le guidage du lecteur de la même manière que l'usage de connecteurs de type « *d'une part, [...] d'autre part, [...] enfin* », et favorise une visualisation spatiale de la liste des propositions qui montre que le texte ne se réduit pas à son matériau verbal.

De tels indices de structuration sont présents dans les deux disciplines, et concernent à la fois les propositions de l'auteur et les exemples ou les extraits de texte étudié. Cependant, la différence disciplinaire est là encore significative, comme le suggère le graphe ci-dessous, qui analyse ces indices tous types confondus.

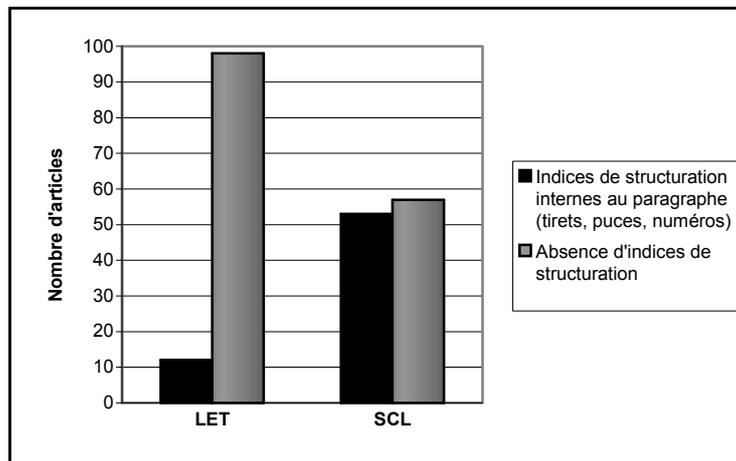


Figure 4.15. L'usage des indices de structuration de bas niveau dans les deux disciplines.

Il est particulièrement intéressant de remarquer que les indices de structuration internes au paragraphe se trouvent nettement corrélés à des caractéristiques structurelles de plus haut niveau, comme le montrent les tests Chi2 de contingence : la présence d'indices de structuration va de pair avec le titrage des parties¹, la numérotation des sections², la présence de sous-parties³.

En cela, de tels éléments peuvent permettre d'affiner les modèles de textualité précédemment décrits, notamment au niveau du pôle où le texte prend le contre-pied d'un modèle en vigueur en sciences : à ce pôle caractérisé par des subdivisions qui se limitent à un niveau, et qui ne sont marquées que par des sauts de ligne, les textes ne contiennent pas même de tirets ni de symboles ordinaux qui serviraient la mise en relief des propositions. Le guidage du lecteur tend ainsi à se réduire au matériau verbal lui-même.

Il importe alors de tenir compte de la question des formalismes : l'analyse de la pluri-sémioticité peut en effet permettre d'approfondir ce constat d'un texte réduit ou non à son matériau en langue naturelle, et de mieux cerner encore ces différences quant aux composantes du texte et aux modes de guidage du lecteur qu'elles rendent tangible.

4.4. Pluri-sémioticité de l'article et identités disciplinaires des modèles de texte

4.4.1. Les formalismes observables dans le corps des articles

La pluri-sémioticité est une caractéristique essentielle de l'écrit scientifique. Elle fait souvent l'objet de traitements spécifiques dans les manuels et guides d'aide à la rédaction⁴. Nous l'aborderons ici par une étude sommaire des formalismes présents dans les corps de texte, à titre de piste pour l'analyse des genres ; une telle question mériterait à elle seule un traitement approfondi, mais il s'agira principalement pour nous de voir en quoi elle est utile à la description des identités disciplinaires manifestes dans l'écriture de l'article.

Les deux pôles de textualité mis en évidence à partir des indices de structuration suggèrent deux tendances inverses à un texte réduit ou non à ce que nous pourrions appeler du

¹ Test de Chi2 = 13,271, $p = 0,0003$.

² Test de Chi2 = 11,881, $p(\text{Chi}2) = 0,0006$.

³ Test de Chi2 = 13,211, $p(\text{Chi}2) = 0,006$.

⁴ Elle est également thématiquée du côté des travaux de recherche dans le numéro 20 de la revue *Genesis* (Pinon et Barberousse eds, 2003), qui traite de l'écriture scientifique d'un point de vue génétique, et notamment dans le travail d'A. Barberousse et E. Guyon sur l'écriture et la création scientifique chez Pierre-Gilles de Gennes, et dans celle de S. Jane Blevins, E.-O. Le Bigot sur la genèse de résultats théoriques en physique.

« texte brut », en référence au traitement informatique des données textuelles. Le texte brut se caractérise en effet par le fait que l'information codée se limite aux caractères et aux tabulations, les premiers comprenant les ponctuations mais ne dépassant pas le nombre de 256. Pour être plus juste, il semble qu'on puisse voir dans le modèle de textualité où n'intervient aucun des indices de structuration retenus une tendance à ne pas ajouter au matériau linguistique en langue naturelle d'autres systèmes sémiographiques que la segmentation en paragraphes.

La question des formalismes est éminemment complexe : la centration sur le matériau linguistique explique sans doute que des éléments comme les flèches, les accolades, et plus largement tout ce qui relève de l'iconographique, décrits notamment par J. Anis (1988) soit assez rarement réinvesti dans les analyses de textes et des discours, hors d'une perspective communicationnelle ou rhétorique, comme par exemple dans les études sur l'image dans la publicité) alors que ce sont là des caractéristiques centrales de l'écrit en général. Si l'on peut y voir aussi l'effet du primat épistémologique accordé par la linguistique à l'oral, les tentatives de descriptions de ces formalismes montrent les lacunes au niveau du métalangage disponible, la difficulté qu'il y a à rendre compte des différents types de formalismes existants et de leur fonction, mais aussi leur intérêt pour l'analyse de certaines pratiques textuelles (cf. l'étude de la flèche dans le travail mené sur la prise de notes par F. Boch, 1999).

L'analyse des écrits disciplinaires a fort à gagner à leur étude. Un premier ensemble de formalismes observables dans les textes concerne en effet l'appareillage symbolique constitutif du métalangage propre à un champ disciplinaire. Nombre d'articles de notre corpus contiennent ainsi des symboles utiles à la représentation linguistique : par exemple, les abréviations pour les parties du discours (« N » pour nom, « V » pour verbe...), les symboles de l'alphabet phonétique, les tirets pour signaler les particules en morphologie (« sem- dans semestre »), «), », les crochets utilisés pour la transcription phonétique, ou encore pour les traits sémantiques (« les deux traits [+bipède] et [-plume] »), les astérisques (*) pour l'agrammaticalité d'un énoncé, et les points d'interrogation (?) pour son inacceptabilité.

On trouve également des systèmes de signes (alphabétiques ou non) utilisés pour les transcriptions d'entretiens (// pour une pause par exemple), et qui pourraient en cela n'être que pour partie spécifiques à la linguistique ; ainsi le marquage de certaines liaisons n'est-il peut-être que le fait des transcriptions linguistiques des entretiens.

D'autres symboles semblent davantage transversaux que disciplinaires. C'est le cas par exemple des barres obliques que l'on trouve dans les analyses lexicales et sémantiques pour marquer l'opposition (« anciens / modernes, fertilité / stérilité ») ; l'usage disciplinaire semble ici codifié, mais la fonction oppositive du symbole est attestée au-delà du seul discours scientifique. Interviennent également des signes mathématiques, les signes « + » et «-» notamment, qui comme on l'a vu avec l'exemple « les deux traits [+bipède] et [-plume] » ne sont pas tant là pour rendre compte d'analyses mathématiques que pour signaler la présence ou l'absence d'une unité, d'un trait sémantique, etc. D'autres formalismes mathématiques sont également en usage, comme les rapports d'inclusion, d'implication et les ensembles (« l'ensemble {2, 4, 6, 8, 10,...} »).

Ces symboles peuvent entrer dans des formalismes complexes, qui intègrent des effets de mathématisation, comme les équations, les règles logiques et les algorithmes. Les représentations phonologique et syntaxique mobilisent ainsi une grande diversité de symboles qui rendent compte des unités linguistiques et de leurs relations de dépendance. Les deux extraits suivants en sont un exemple parmi nombre d'autres :

Syllabe
 ((s)(,°(l)) = ((attaque)(noyau (coda))) = ((adjonction)(tête (complément))) [SCL054]

Every cellist played some variations.

D'après May (1977), (8) peut être interprété seulement de deux manières différentes, en fonction de la portée assignée aux deux QP, c'est-à-dire (i) *some* > *every* et (ii) *every* > *some*. Mais Diesing soutient que l'interprétation (ii) doit être elle-même subdivisée en deux suivant la (non-) présuppositionnalité de « *some variations* », alors qu'en (i), le DP objet « *some variations* » est toujours présupposé. **Voici les représentations correspondant aux trois interprétations de (8)**, données par Diesing (1992, p. 66) :

[Every_x [cellist (x)] Some_y [variation (y)] x played y]
 [Every_x [cellist (x)] 3_y variation (y) A x played y]
 [Some_y [variation (y)] Every_x [cellist (x)] x played y] [SCL052]

Ces représentations peuvent aller jusqu'à exploiter une configuration spatiale non exclusivement linéaire, comme c'est le cas en particulier dans les deux articles cités, parus dans un numéro de la revue *Recherches Linguistique de Vincennes* consacré à la linéarité.

(12) (I) Pierre-ka (II) Marie-lîl (III) Paris-e (IV) ponàs'ta
 -Nom -Acc -Loc envoya
 'Pierre a envoyé Marie à Paris.' [SCL052]

(14) [_N a- -N° [_V d-e-z-Nég [_V [a-N-PL] [-V° [f-y-m-R]
 1 5 4 2 3
 1 : a°()
 2 : a° a°() ~ a° a°()
 3 : (f_y) (m(a³ 3)
 4 : a° (N) (f_y) (m(a° 3)
 5 : (d e°) (z (a° (N)) (f_y) (m(a° Z) = [dez ~fyma 3]
 [SCL054]

On se rapproche alors d'un point de vue structurel des formalismes constitués par les tableaux, mais c'est regrouper sous une même dénomination de « formalismes » des modes de représentations extrêmement variées, qui exploitent plus ou moins les possibilités offertes par les colonnes, les lignes, leurs divisions et leurs intersections.

Notons à ce propos que l'utilisation de tableaux peut correspondre à une manière de contourner les difficultés posées par l'alignement dans les logiciels de traitement de texte. Ainsi leurs quelques occurrences en LET servent-elles parfois la mise en forme d'un texte en vers. Il peut également s'agir d'aligner deux textes, comme dans cet extrait d'un article de LET où l'auteur compare l'Ovide moralisé avec le Roman de Fauvel :

L'*Ovide moralisé* ne se limite pas à expliciter le *sensus moralis* caché sous le voile de la fable, il en suggère les implications référentielles. S'il pense, à l'instar des traités que nous venons d'évoquer, la politique en termes de morale, il s'en distingue par le mouvement qui le conduit du récit mythologique à son allégorisation, puis à l'application de la glose à l'actualité du moment. Ce mouvement est celui même d'une satire comme le *Roman de Fauvel* (1310) qui se sert du récit allégorique (le triomphe du cheval fauve) pour dénoncer le renversement des valeurs à la cour de France. La parenté est plus profonde qu'on est tenté de le croire à une première lecture : les deux textes voient dans l'abaissement de l'Église et dans le triomphe des vices les signes avant-coureurs de la fin des temps. Ils trouvent, pour le dire, des mots, des rimes, qui se font écho :

Saint Yglise flote et chancele.
 Se Diex ne garde sa nacele,
 Toute est descloëe et rompue.
 (OM, vv. 4052-4054)

(...) c'est une trop grande merveille
 Que saint Pere ne s'apareille
 De tost secourre a sa nacele,
 Qui si horriblement chancele.
 (Fauvel, vv. 371-374)

[LET042]

Cette mise en correspondance des lignes est vraie aussi au niveau des colonnes, comme dans cette représentation syntaxique tirée article de SCL

a. (fDr)(ta)(ks a~)	f- -r-Adj.-t	ni T ni L (non taxique inerte)
b. (fDr)(a)(ks ~)	f- -r-Adj.-t	L (non taxique, actif)
c. (fD)(ra)(ks ~)		L et resyllabation

[SCL054]

Le tableau peut ainsi fonctionner comme une instruction à une lecture conjointe de ses différents espaces ; cet outil de visualisation dérivé des pratiques de liste s'associe dans le second extrait à un autre type de formalisation, qui sert une symbolisation en langage formel des données syntaxiques et montre la diversité des formes et des fonctions de ce que nous traitons sous le terme de formalismes.

Outre les tableaux, d'autres représentations graphiques sont utilisées dans les articles, notamment les schémas et les graphes. Ceux-ci peuvent permettre de relativiser les résultats obtenus quant au volume de l'article en nombre de mots : la taille comparable voire inférieure constatée pour les articles de SCL ne tient vraisemblablement qu'au fait qu'ils sont davantage concernés par la pluri-sémioticité, et par des éléments graphiques volumineux mais non quantifiables en des termes verbaux.

Comme pour les tableaux, la diversité observable en matière de graphes et de schémas inviterait à discuter plus en avant ces catégories « iconiques »¹ : la question se pose par exemple de savoir si les représentations syntaxiques que constituent les arbres, mêmes restreints à une branche, peuvent être considérées comme des formes de schéma, ou ce qui fait la spécificité d'un tableau ou d'un graphe pour mériter d'être traités indépendamment d'autres types de schémas.

Il reste que ce recensement des différents formalismes en usage dans notre corpus offre nous semble-t-il une ouverture aux analyses textuelles ; il est à souligner que C. Poudat (2006a) intègre ainsi à son étude automatisée du genre de l'article une annotation de certains des symboles linguistiques que nous avons évoqués (l'astérisque et le point d'interrogation), dans un double objectif de description de ces marqueurs proprement scientifiques et de désambiguïsation (l'analyse des points d'interrogation doit tenir compte du fait qu'ils ne signifient pas tous qu'on a affaire à une modalité interrogative, et ne pas nettoyer ou annoter les textes à ce niveau parasite immanquablement les résultats).

Sans procéder à une quantification systématique de l'ensemble des phénomènes évoqués ci-dessus, nous avons tenté de rendre compte de leur diversité dans notre corpus et il convient à présent de montrer qu'il n'en va pas de même dans chacune des deux disciplines.

4.4.2. Formalismes et modèles de texte : le contraste disciplinaire

Dans le corps des articles de LET, on relève des symboles de représentation linguistique d'une part, des représentations sous forme de schémas d'autre part, qui mettent en évidence la progression thématique d'un texte ou la structure d'un poème, et enfin des tableaux qui ont une fonction proche des schémas, de structuration des contenus. Cependant ces éléments sont loin de concerner la majorité des articles ; au contraire, ils restent faiblement représentés. Soulignons ainsi que seuls 6 articles de LET contiennent un tableau ou plus, contre 37 articles en SCL. Si l'on tient compte de l'ensemble des formalismes évoqués, 8 articles de LET en

¹ C'est en ces termes que U. Eco traite des diagrammes et dessins dans sa *Sémiotique et philosophie du langage* (1988).

contiennent, contre 83 articles de SCL, soit une proportion marginale en LET de 7,3% des articles, et très importante en SCL avec 75,5% des articles concernés.

En SCL, l'ensemble des formalismes évoqués est présent, pour la présentation des données, leur analyse et la visualisation des résultats mais également dans les cadrages théoriques, alors que cet usage est complètement absent en LET.

Par ailleurs, il est remarquable de constater la présence en SCL d'effets de mathématisation : ces formalismes spécifiques interviennent assez rarement (8 articles), mais pas uniquement pour la formulation de règles en logique et pour certains formalismes syntaxiques ; on trouve par exemple des équations en analyse du discours pour décrire les rapports entre langue et discours ou encore entre écrit et oral comme dans cet extrait :

l'équation *écrit = oral surmarqué*, c'est-à-dire *écrit = oral + effet de spontanéité + expressivité + subjectivité*, soit la formule :

$$É = O + [S/+Ex+Sub].$$
[SCL091]

Si la tendance à la formalisation est un trait fort de l'écriture de l'article en SCL, il est possible de se demander en quoi l'équation sert ici véritablement le propos. Elle joue certes une fonction d'économie verbale, déjà remplie cependant par un texte introductif comportant lui-même des symboles, et le recours à cette formule calquée sur un modèle mathématique semble finalement presque caricatural.

Il rend pour le moins tangible une volonté de l'auteur de se poser, à peu de frais, comme auteur « scientifique » : le formalisme ici est porteur d'une représentation de la science que l'on peut dire scientifique, dans le sens où elle se fonde sur le modèle dominant des sciences dures et expérimentales et ne retient en quelque sorte qu'un appareil de scientificité, à défaut semble-t-il de servir une véritable fonctionnalité. L'équation pourrait même, d'un point de vue scientifique, être discutée quant à sa pertinence pour rendre compte des rapports complexes entre écrit et oral, et quant au degré de formalisation et de non-ambiguïté des concepts mis en scène au moyen des abréviations et des signes mathématiques.

Cet examen rapide des formalismes présents dans le corps de l'article permet comme pour les aspects structuraux du genre de mettre en évidence deux tendances opposées, l'une consistant en une réduction maximale des formalismes avec l'absence même de symboles de représentation linguistique, l'autre consistant en leur multiplication et en l'usage de formalismes iconiques de visualisation¹ (la schématisation, qu'elle prenne la forme de dessins, de graphes ou de présentations dans des tableaux) et de formalismes linguistiques complexes, comme les représentations phonologiques ou syntaxiques évoquées *supra*, et les formes de mathématisation de certaines des représentations linguistiques.

Ces deux pôles de texte, avec ou sans formalismes, sont élaborés avec comme point de référence une tendance qui reproduirait un modèle de dominant de l'écriture en sciences, et une tendance inverse définie négativement à partir de l'absence de composantes en vigueur dans ce modèle. Un profil intermédiaire de textes se situe en particulier dans l'usage de symboles de représentation linguistique répandus (*SN*, *SV*, etc.), mais dans l'absence d'autres types de formalismes.

Les deux tendances adverses peuvent alors permettre d'affiner les profils de textualité mis en évidence à partir de la structure des textes. Il apparaît en effet que le pôle des textes proches de ceux des sciences dures dans leur structure s'associe à un grand nombre de

¹ Nous parlons ici de formalismes de visualisation mais il convient de souligner leur fonction heuristique, pour le lecteur comme pour le scripteur dont le raisonnement se construit pour beaucoup à travers ces schématisations, comme le montre l'étude de A. Barberousse et E. Guyon (2003) consacrée à l'étude de l'écriture scientifique de P.G. de Gênes.

formalismes de visualisations (notamment les schémas, tableaux). Toutes catégories de formalismes confondues, le test de Chi2 atteste de corrélations entre les formalismes d'une part et la pratique du titrage des parties, de la numérotation des sections, la présence de sous-parties et d'indices de structuration¹.

Au pôle opposé, les textes qui ont pour seuls indices de subdivision des sauts de ligne ne contiennent donc - tendanciellement au moins – aucun formalisme, et pas non plus de tirets ni de numéros qui serviraient la mise en relief des propositions. La textualité tend à se suffire de l'usage des caractères alphabétiques, des ponctuations et des tabulations.

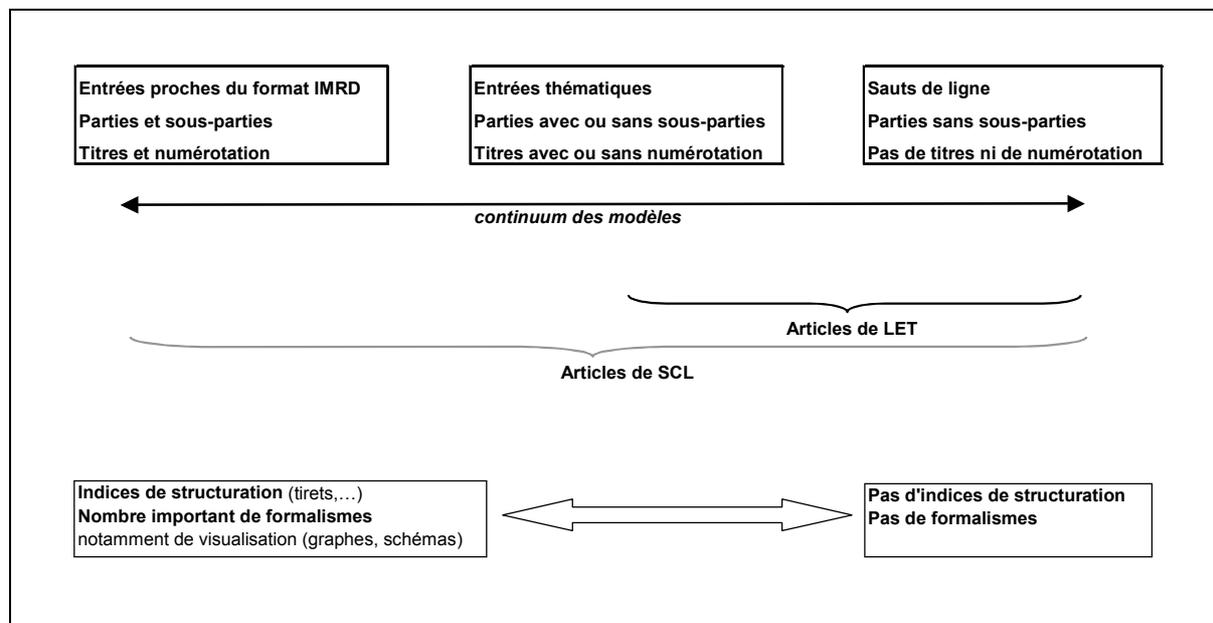


Figure 4.16. Corrélations entre structure et formalismes : deux modèles de textualité opposés, des identités disciplinaires contrastées.

Là encore, les deux disciplines ne se superposent pas aux deux tendances opposées. On l'a dit, les articles de SCL peuvent se répartir tout le long de l'axe en ce qui concerne leur structuration, mais c'est vrai également en y associant les formalismes : la pratique de formalisation est un trait fort de l'article en SCL, mais 19,1% des articles de SCL ne contiennent pas de formalismes. Il existe donc entre les articles des deux disciplines une zone commune, même si les articles de LET ne tendent que dans une proportion très marginale vers le pôle caractéristique d'un modèle dominant en sciences – dominant au moins dans les représentations que l'on en a, mais aussi dans certaines descriptions qui en sont faites, dans les travaux de recherche ou dans les guides et manuels.

4.4.3. Bilan sur les modèles de texte et les identités disciplinaires qu'ils manifestent

Les deux pôles de textualité mis en évidence à partir de traits opposés et de corrélation de traits correspondent aux caractéristiques intrinsèques de quelques articles, mais plus largement à des tendances fondées sur des caractéristiques textuelles contraires et qui rapprochent ou éloignent les textes de modèles de texte et de modèles épistémologiques. Elles font apparaître un contraste disciplinaire dans la mesure où les articles de LET s'échelonnent sur une moitié de l'axe et tendent plutôt vers le pôle droit du schéma, tandis que ceux de SCL

¹ p (Chi2) <0,0001 dans les trois premiers cas, et p (Chi2)= 0,0003 dans le dernier.

se caractérisent par une plus grande disparité des pratiques et peuvent couvrir l'ensemble du continuum.

Ainsi les corrélations entre les indices de structuration et les formalismes se vérifient-elles à l'échelle du corpus comme dans celui de SCL, où elles signalent l'existence de profils diversifiés en fonction de ces traits. Par contre, ces variables ne sont qu'extrêmement peu soumises à l'effet du statut de l'auteur. Hormis le recours plus important aux bibliographies et annexes, les textes des doctorants n'ont pour autre spécificité que le titrage plus massif des introductions (à hauteur de 42,9% des articles, contre 18,9% chez les non-doctorants) et plus largement, le fait qu'elles soient plus aisément identifiables dans leurs textes.

Cette unique différence suggère un guidage plus important du lecteur à ce niveau, mais elle n'est valable que les deux disciplines confondues ; comme elle ne s'associe pas à d'autres traits, il serait peu pertinent de vouloir positionner les textes des deux groupes des doctorants et non doctorants sur le continuum schématisé ci-dessus : les tendances décrites à partir des caractéristiques formelles des textes renverraient plutôt à des types d'étude et des épistémologies disciplinaires et sous-disciplinaires qu'à une forme d'acculturation spécifique aux normes en usage.

Elles manifestent ainsi des images différentes de la recherche : d'un côté, une recherche se subdivisant en composantes, et dont l'appareillage conceptuel, les matériaux et les résultats peuvent faire l'objet de formalismes et de représentations visuelles ; de l'autre, une activité de recherche qui se présente de manière unifiée, par le faible recours aux espaces péri-textuels autres que les notes, aux subdivisions internes du texte, au titrage, à la numérotation, aux symboles autres que le lexique de la langue naturelle et à la formalisation. Le métalangage disciplinaire de ce côté-là n'autorise que le texte en langue naturelle, défini négativement par l'absence de dispositifs propres à la tendance opposée (principes « reader-friendly » de marquage de la structure, appareillage péri-textuel, métalangage formel, visualisations graphiques, etc.), et qui pourraient même être caractéristiques du discours scientifique pris dans son ensemble au regard d'autres champs de pratiques.

Si les caractéristiques textuelles peuvent ainsi signaler des rapports différents au lecteur et à l'objet de recherche, se pose la question de leur fonctionnalité. Celle-ci est délicate tant les identités disciplinaires doivent influencer sur ce que l'on peut considérer comme étant plus ou moins exigeant pour le lecteur dans la subdivision du texte et ses composantes, et plus ou moins pertinent dans le métalangage utilisé.

Cela étant, il est des éléments qui relèvent de l'efficace de la présentation, mais d'autres dont l'usage sert prioritairement un positionnement dans des paradigmes épistémologiques, la monstration semblant alors pouvoir l'emporter sur la fonctionnalisation. Le scientisme qui semble ainsi se manifester dans certaines caractéristiques des textes, au niveau de la formalisation en particulier montre au moins l'intérêt qu'il y aurait à mieux connaître les usages effectifs des dispositifs textuels propres aux sciences dures et expérimentales, au discours scientifique en général, et à leur circulation et leur mise en scène dans d'autres champs¹.

¹ On voit ainsi aussi l'intérêt qu'il y aurait à penser en termes de genres le continuum entre la science et sa vulgarisation mis en évidence d'un point de vue socio-communicatif (Jacobi, 1999), pour décrire en quoi et comment les caractéristiques textuelles de genres *a priori* distincts attestent de ce continuum.

4.5. Les exemples et les citations : pratiques textuelles et épistémologies

Les modèles de textualité mis en évidence à partir des corrélations de variables structurelles et sémiotiques attestent de différences disciplinaires et de tendances épistémologiques, et ce sont encore avec ces dernières que nous terminerons cette macro-analyse de l'article en traitant des exemples et des citations.

La prise en compte des exemples et des citations est de toute importance dans le cadre d'une analyse attentive aux composantes structurelles du genre et à l'idée que tout texte est fait de plusieurs textes : les exemples et les citations se présentent comme des espaces textuels à part entière, d'un double point de vue, en tant qu'espaces spécifiques du texte de l'article, et en tant que textes spécifiques, au statut de « discours autres » qui s'intègrent à l'article, et relèvent en cela de sa dimension polyphonique.

Les exemples et les citations se caractérisent par une rupture avec le texte qui les introduit et les commente. Cette rupture peut être signalée au moyen de sauts de ligne et tabulations, dans le cas des exemples et citations détachés du reste du texte dans un paragraphe à part :

En organisant sa soirée de gala dans l'atmosphère balnéaire de Lucque qui rappelle celle de Baden, Montaigne aurait donc pu tout aussi bien évoquer les réflexions philosophiques qu'avait inspirées cent ans auparavant au Pogge, en route vers le Concile de Bâle, la fréquentation des bains de Baden:

Ces bons Allemands auraient fait *d'excellents citoyens de la République de Platon*, où tout devait être commun; bien que fort ignorants de la théorie, ils eussent été, n'en doutons pas, très expert dans la pratique (...) Cette petite société se rapproche de la secte d'Epicure.

Le Pogge n'avait pas manqué de souligner le rôle primordial de l'eau dans le processus de sociabilité.
[LIT004]

ou au moyen des guillemets et/ou italiques, dans les cas des exemples et des citations que nous dirons « intégrés » au texte comme dans l'extrait suivant :

Le passage suivant montre que Montaigne est conscient du problème de l'interprétation du pyrrhonisme: "J'exprime cette fantaisie [*sc. le pyrrhonisme*] autant que je puis, par ce que plusieurs la trouvent difficile à concevoir; et les auteurs mesmes la représentent un peu obscurément et diversement" (II,12,505a). Montaigne indique que les sources sont parfois difficiles à concilier et qu'on trouve des problèmes d'interprétation.
[LIT006]

Au vu des prescriptions éditoriales adoptées par certaines revues, et des guides et manuels d'aide à la rédaction scientifique, la différence entre les citations détachées et intégrées doit tenir à leur longueur ; les exemples devraient quant à eux être détachés. Mais un examen du corpus montre que les pratiques restent variables et que d'autres critères semblent pouvoir entrer en jeu.

Par ailleurs, il n'est pas rare que, tout en se caractérisant par une démarcation typographique, ces composantes, omniprésentes, se trouvent intégrées à la syntaxe de la phrase qui les introduit ou les commente sous la forme d'« îlots textuels » (Authier-Revuz, 1984, Boch et Grossmann, 2002) :

Inversement, deux nouveaux dérivés sur *-on* sont introduits, mais aucun d'entre eux n'a suivi la règle conseillée qui préconise l'emploi de la consonne simple : *jargonneux* (également dans le *PR 2001* mais il est vrai que *jargonner* n'a pas été rectifié) au lieu de *jargoneux*, et surtout *cohabitationniste* au lieu de *cohabitationiste*... [SCL012]

Cette intrication a pour conséquence que les choix à faire pour supprimer ces éléments dans une perspective d'analyse automatique sont périlleux¹ (cf. au chapitre 3, la question du nettoyage et du traitement des textes préalable à leur analyse). Aussi, quelle que soit la solution adoptée, l'analyse automatique intègre nécessairement les caractéristiques de ces composantes, plus ou moins marquées par rapport à celles du texte analysé (au niveau de la langue, du genre de texte, etc.)².

En prenant comme point de départ une approche des exemples et des citations en tant que composantes structurales, nous procéderons d'abord à une estimation du poids de ces éléments dans le texte. Dans la manière dont les exemples et les citations se présentent dans le texte, il s'agira alors de voir ce qu'ils peuvent dire de types d'études et d'orientations épistémologiques dans les deux disciplines.

4.5.1. Le poids des exemples et citations dans les textes

Nous avons étudié le poids des exemples et des citations dans 35 articles en retenant les exemples et les citations décrochés du texte, et les exemples et les citations intégrés au texte et comptant plus de 20 mots ou 100 caractères (espaces non compris). Leur poids a été évalué en nombre de mots avec un ratio sur le nombre de mots du texte.

Poids des exemples

En ce qui concerne le poids des exemples, le premier résultat à souligner est que près d'un tiers des articles de SCL ne contient pas d'exemples – ou plus précisément, pas d'exemple détaché dans un paragraphe séparé, et pas d'exemple intégré de plus de 20 mots). En LET, seuls 5 articles présentent ce profil. Cependant, les résultats sont proches dans les deux disciplines comme le suggère la distribution en fréquence du poids des exemples représentée ci-dessous.

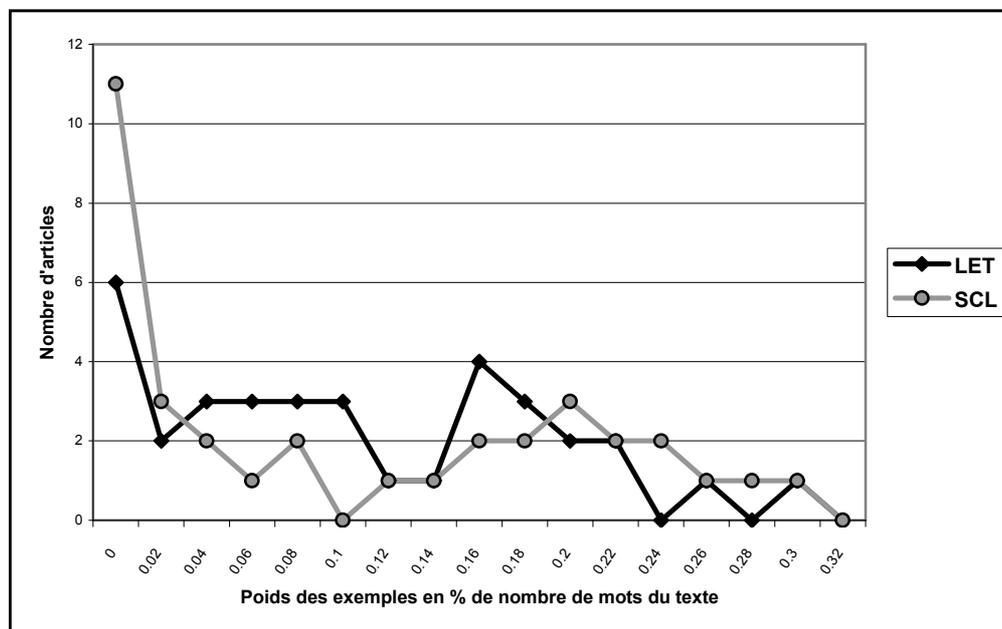


Figure 4.17. Distribution en fréquence du poids des exemples : comparaison disciplinaire.

¹ Enlever uniquement les éléments détachés produit un bruit trop important dans l'analyse des caractéristiques du texte, mais chercher à enlever tous ces éléments, jusqu'à ceux qui sont intégrés à la syntaxe vide le texte d'une part importante de sa substance et entrave la reconnaissance des unités phrastiques et de leurs constituants.

² C'est en ce sens que les citations, *a priori*, perturberaient moins l'analyse que les exemples.

En moyenne, les exemples représentent 10,5% du texte en LET et 9,8% en SCL. Les valeurs varient entre 0 et 29,2% en LET, et entre 0 et 29,7% en SCL. Les tests statistiques vérifient que la discipline est sans effet sur le volume des exemples : en tant que composante structurelle, l'exemple occupe une place équivalente dans les articles des deux disciplines et constitue un trait commun du genre. Soulignons cependant que les exemples sont en général numérotés en SCL et non en LET, ce qui rejoint ce que nous avons dit des indices de structuration internes au paragraphe, plus utilisés en SCL qu'en LET, et des modèles de textualité en vigueur dans les deux disciplines.

Le poids des exemples n'est pas soumis non plus à l'effet du statut de l'auteur. Par contre, il est étroitement corrélé au poids cumulé des exemples et des citations, dans l'ensemble du corpus avec un taux élevé de 92,4%, de même que dans chaque discipline et chaque groupe d'auteurs. Ce résultat suggère que le poids des exemples est globalement bien supérieur à celui des citations.

Poids des citations

Les citations représentent une composante structurelle assez marginale du genre : 30 articles de LET et 13 articles de SCL n'en contiennent aucune. Ces résultats corroborent les analyses de F. Boch et F. Grossmann (2002) et F. Grossmann (2002) sur les pratiques expertes de la citation dans l'article de recherche : les auteurs soulignent en effet que si la pratique citative est très prisée des néophytes dans leurs premiers écrits de recherche, les experts lui préfèrent d'autres modes d'introduction des références, comme la reformulation ou l'évocation et l'allusion.

Le volume moyen des citations est extrêmement faible comparativement à celui des exemples : il représente 2,3% de l'article en SCL et 0,4% en LET. Le caractère marginal de la citation est donc plus vrai encore en LET qu'en SCL ; en LET, le volume maximal attesté est de 8,8% de l'article, alors qu'on remarque en SCL un article où les citations occupent une place importante, qui atteint 28,3%. Ce cas reste cependant isolé, puisque la tendance centrale pour les articles contenant des citations se situe plutôt entre 2 et 5%. La distribution en fréquence du poids des citations représentée dans le graphique suivant rend compte à la fois de ce caractère marginal de la citation, et de sa place moindre encore en LET.

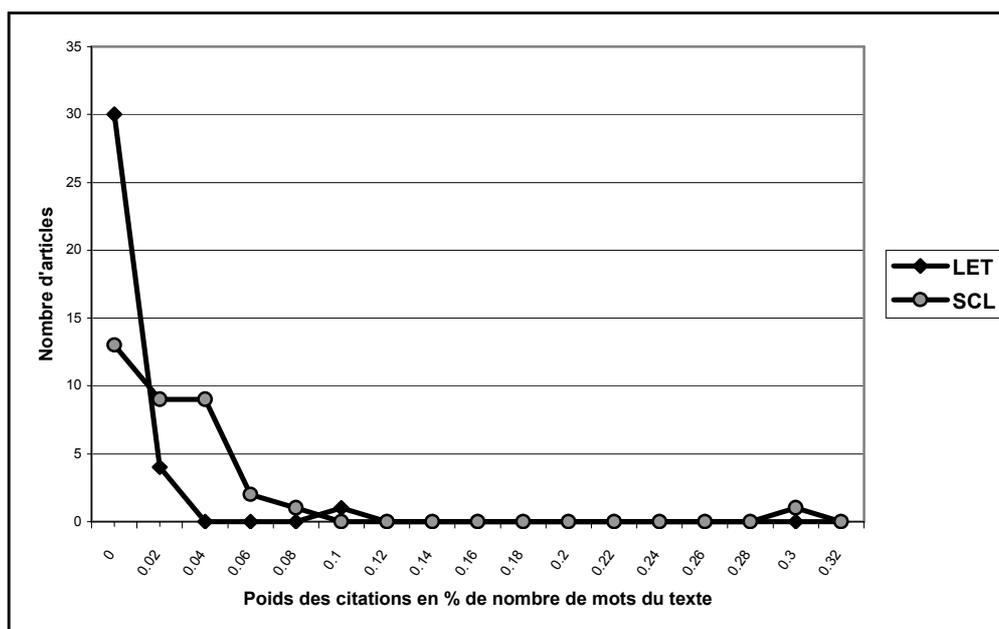


Figure 4.18. Distribution en fréquence du poids des citations : comparaison disciplinaire.

La différence disciplinaire dans l'usage des citations se vérifie statistiquement¹. Elle reste en elle-même difficilement interprétable, sauf à titre d'hypothèses sur le primat d'autres modes de référencement ; la différence pourrait aussi s'expliquer par une tendance en LET, ou du moins dans les articles du corpus, à préférer des études plus empiriques à des articles discutant des thèses et justifiant en cela davantage peut-être le recours à la pratique citationnelle.

Il convient à ce propos de souligner que la moyenne en SCL est largement influencée par une valeur très élevée de 28,9% ; sans que cela remette en question la différence disciplinaire observable dans les taux compris entre 2 et 5%, on remarquera que cette valeur a statut d'hapax, et résulte de la spécificité de l'étude proposée dans l'article : il s'agit en effet d'un examen des théories saussuriennes sur la base du *Cours de Linguistique générale*. Les citations ont en cela un statut ambivalent, car elles relèvent vraisemblablement plus du matériau de la recherche que de ses sources théoriques.

On voit au passage comment la difficulté posée à l'analyste dans cet article est résolue au niveau du genre, la mise en évidence de ses caractéristiques permettant de pointer la singularité de ce texte. C'est ce qu'a déjà montré F. Rastier à propos la classification automatique des textes, qui signale à partir de leurs caractéristiques morpho-syntaxiques les éventuelles erreurs de la classification établie *a priori* par l'analyste (Malrieu et Rastier, 2001).

Par ailleurs, la pratique citation fait également l'objet d'une différenciation au sein du corpus de LET en fonction du statut de l'auteur. Les auteurs doctorants recourent significativement plus aux citations que les non doctorants (avec une probabilité pour que cet effet ne soit pas le fruit du hasard à la limite acceptable de 5%) : le poids moyen des citations est de 1,4% dans les textes de doctorants contre 0,7% dans ceux des autres ; par ailleurs, le maximum évoqué *supra* de 8,9% est le fait d'un article de doctorant ; dans le groupe des non-doctorants, le poids maximum des citations n'est que de 1,7% du texte.

Cet usage plus important des citations dans les articles de doctorants de LET vérifie là encore l'analyse faite par F. Boch et F. Grossmann (2002) des pratiques néophytes et expertes de la citation. S'ils s'intéressent à un public d'étudiants non encore concerné par la publication d'un article, il est intéressant de voir que le constat d'une pratique citative marquée chez les néophytes par comparaison avec les experts qui ne l'utilisent que peu (Boch et Grossmann, *ibid.*, et Grossmann, 2002) reste vraie chez cette catégorie intermédiaire du « nouvel entrant dans le champ scientifique » (Bourdieu, 2001 : 95) qu'est le doctorant.

Alors que la spécificité des pratiques doctorantes pourrait être indicielle d'une évolution des normes de l'article en LET, le fait qu'elle soit ici attestée chez des étudiants qui ont à produire des mémoires montre qu'elle est à envisager plutôt en termes de spécificité statutaire dans l'acculturation au champ scientifique. Dans cette perspective, la citation serait à comprendre comme le résultat de l'intériorisation de normes prescriptives ; les manuels d'aide à la rédaction scientifique recommandent au scripteur de citer ses sources.

L'exigence est ambiguë puisque elle peut signaler simplement la nécessité du référencement, et non la nécessité de citations à proprement parler. De fait, la citation ne correspond manifestement pas aux normes d'usage, où ce serait plutôt une obligation plus large de renvoi à d'autres sources qui l'emporterait, mais il est possible que la norme qui veut que l'on se réfère à d'autres travaux soit interprétée chez les doctorants comme une nécessité de citer de passages de textes qu'une interprétation spécifique de la norme se manifeste ainsi chez les scripteurs doctorants. Il nous reviendra donc d'analyser si le renvoi aux sources de

¹ Mann Withney, $U = 296$, $p = 0,0002$.

manière plus générale est plus fréquent dans les textes de doctorants que chez les autres scripteurs¹, ou si la pratique citationnelle est en elle-même un facteur de différenciation des pratiques.

Pour l'heure, il reste à voir si la pratique de la citation est liée d'une quelconque manière à celle de l'exemple.

Un poids relatif des exemples par rapport à celui des citations ?

Les deux catégories des exemples et des citations permettent de penser que ces deux composantes pourraient être inversement proportionnelles, en raison de leur fonction dans l'économie du texte ; on pourrait ainsi faire l'hypothèse qu'une étude plus empirique tendrait à recourir davantage aux exemples et qu'une étude plus théorique mobiliserait moins d'exemples et davantage de citations. Les exemples et les citations constitueraient ainsi une entrée intéressante pour la mise en évidence de types d'articles ou de types d'études, quoique la distinction entre étude empirique et théorique demande une certaine prudence : l'empirisme n'exclut en rien la théorisation, et inversement, une entreprise théorique doit – ou du moins peut – s'appuyer sur des exemples.

Cette corrélation négative a été vérifiée dans l'étude de C. Poudat (2006a), qui conclut ainsi à cette possible existence d'un lien entre les composantes structurelles que sont les exemples et des citations et le type d'étude mené dans l'article. Dans notre corpus, on remarque que les articles les articles présentant le plus de citations sont aussi ceux qui contiennent le moins d'exemples, et qu'inversement ceux qui contiennent le plus d'exemples contiennent peu de citations, ce qui argue en faveur de la tendance mise en évidence dans les travaux de C. Poudat. Cependant, les deux séries mesures ne sont pas statistiquement corrélées, puisque le test de corrélation affiche un taux très faible de -10% en SCL, et un taux également faible et de surcroît positif de 12% en LET. Le taux ne permet pas de conclure non plus à une corrélation entre exemples et citations chez les doctorants et les autres scripteurs.

On voit ici poindre le problème de l'irréductible diversité des textes : alors même que le recours aux statistiques vise à se prémunir des effets liés au hasard de l'échantillonnage, la textualité a ceci de spécifique que les distributions de valeurs obtenues sur ses traits caractéristiques sont rarement gaussiennes, et que ce qui s'avère statistiquement significatif dans un corpus, donc *a priori* indépendant de ce dernier, peut ne pas l'être dans un corpus réuni à partir de critères relativement proches.

Cette différence de résultat peut donc s'expliquer par une différence, même faible, au niveau de ces critères constitutifs du corpus, en l'occurrence la présence de revues et numéros différents dans les deux corpus ; elle suggère ainsi que la corrélation négative entre exemples et citations est à valider en tant que tendance qui structure potentiellement mais pas systématiquement les pratiques d'écriture de l'article, et que c'est peut-être un fait plus spécifique à certaines revues ou à certains numéros que de contenir des contributions dont les usages des citations et des exemples sont inversement proportionnels.

On retiendra donc que la corrélation reste dépendante du corpus où elle est étudiée, et que la distinction entre un article et un type d'étude plus théorique et empirique reste très délicate ; fonder une telle distinction sur l'usage des exemples et des citations dans les textes représente une piste très intéressante comme le montre l'étude de C. Poudat (*ibid.*), mais pour mieux cerner cette distinction, il faudrait intégrer à l'analyse d'autres indices qui affinent les résultats et les systématisent.

¹ Cf. chapitre 7.

Les résultats obtenus quant aux poids des exemples et des citations entre les textes de doctorants et de non doctorants renforce l'idée que les exemples et citations ne constitueraient pas à elles seules une entrée suffisante pour l'analyse de types d'études : on pourrait penser que les doctorants produisent davantage d'études empiriques et, partant, recourent de manière plus abondante aux exemples ; or, la seconde au moins de ces deux intuitions est contredite par l'analyse, qui atteste au contraire d'un usage équivalent des exemples quel que soit le statut de l'auteur, et d'un usage plus important des citations dans les textes des doctorants de LET.

Aussi, comme on a vu que la pratique de citation reste assez faible dans l'article, on pourrait tenter d'y associer la prise en compte d'autres modes de référencement comme la reformulation et l'évocation (Boch et Grossmann, 2001 : 96, 2002) pour voir en quoi la pratique du référencement sous toutes ses formes peut être inversement proportionnelle à celle des exemples¹, ce qui correspondrait intuitivement à une distinction que l'on peut faire entre un article plus théorique qui ferait la synthèse de travaux, et un article plus empirique et très exemplifié. Il faudrait aussi approfondir l'analyse des exemples pour voir avec quels types d'exemples des corrélations peuvent se vérifier ou non. Plus largement, un tel objectif d'analyse de types d'articles et d'études gagnerait donc à se fonder sur une analyse multicritériée des textes et des fonctions attachées à leurs propriétés formelles.

4.5.2. Sources documentaires et empiriques : pistes pour l'analyse

A défaut de vérifier l'existence de deux pôles théorique et empirique identifiables sur la base d'un usage des citations d'un côté et des exemples de l'autre, nous avons cependant cherché à voir en quoi les exemples et les citations envisagés en tant que composantes structurelles pouvaient fournir des pistes pour l'analyse de la textualité et des types d'étude et d'orientations épistémologiques qu'elle peut servir.

A partir d'une entrée formelle, d'ordre sémiographique, relative à la manière dont les exemples et les citations sont introduits et référencés, nous avons commencé par dresser un inventaire de ce qui est pris comme exemple dans l'article, pour voir alors comment chaque discipline structure le champ constitué par ses sources documentaires et ses matériaux. Cet inventaire des matériaux observés dans 35 articles de LET et 35 articles de SCL est présenté en annexes 2.

Premières remarques formelles

Dans les articles de SCL, les citations sont détachées sur un critère de longueur. Quant aux exemples, ils sont généralement démarqués et numérotés, alors que cette numérotation est extrêmement rare en LET, ce qui rejoint ce que nous avons vu à propos des indices de structuration de bas niveau, et des modèles de textualité dans les deux disciplines.

Cet usage de la numérotation des exemples permet une distinction nette entre exemples et citations, y compris dans les cas où l'exemple correspond linguistiquement à une citation (e.g. citations extraites d'enregistrements de locuteurs, de textes littéraires ou journalistiques etc.). Cette distinction entre le matériau empirique de l'étude et les citations des travaux des pairs se manifeste jusqu'au niveau du référencement de ces éléments : on en a parlé à propos des bibliographies, celles-ci ne réunissent dans les articles de SCL que les

¹ En s'en tenant aux références aux auteurs lus pour exclure bien entendu les références des exemples dans le cas d'exemples attestés ou cités.

travaux des pairs, et peuvent s'accompagner d'une seconde bibliographie dédiée aux matériaux, qui peuvent également être référencés en notes ou dans le corps du texte.

Au contraire, en LET où les exemples correspondent très massivement à des citations d'extraits de textes littéraires, la distinction formelle entre les deux n'est pas marquée au niveau du référencement, qui se fait quasi-systématiquement en notes comme on l'a dit. Les exemples ne se limitent presque jamais à une fonction autonymique, puisqu'ils servent un propos qui concerne doublement l'expression et le contenu du texte littéraire.

Sur la question du détachement de ces composantes, en termes de disposition dans le texte, ou de leur intégration au texte, les pratiques en LET sont variables ; si le critère de longueur joue assez largement, on remarque cependant dans un certain nombre de textes une intégration systématique de ces composantes, qui concourt au modèle de textualité mis en évidence *supra*, où le texte se présente comme peu apte à se scinder en composantes spécifiques. La pratique du détachement des exemples et des citations peut également permettre de citer des vers, le critère de longueur demeurant alors en usage dans les autres cas. Un autre critère enfin semble pouvoir intervenir, qui tient au statut du matériau. Il n'est pas attesté en SCL et semble pouvoir servir la mise en évidence de types d'études en LET.

Dans les articles de LET : l'intertexte comme matériau et comme objet

Le traitement distinct des exemples (intégrés ou détachés) peut se baser en LET sur une distinction entre matériau principal et matériau secondaire ; cette distinction des matériaux se vérifie cependant aussi dans d'autres articles où tous les exemples font l'objet du même traitement graphique, et semble en cela propre à certains types d'études de la discipline.

Cette distinction des matériaux se manifeste dans l'article suivant, où le premier extrait cité, de Dante, ne sert manifestement pas la même fonction que les extraits de *Astrée* d'Honoré d'Urfé qui constituent l'objet de cet article :

Dans un passage célèbre de la *Vita nova*, Dante raconte comment, alors qu'il s'éloigne de la dame de ses pensées et qu'il souffre de la séparation, Amour apparaît « en [son] imagination » et lui adresse un conseil visant à protéger son aimée par un sentiment feint pour une autre dame :

E dette queste parole, disparve questa mia imaginazione tutta subitamente per la grandissima parte che mi parve che Amore mi desse di sé ; e, quasi cambiato ne la vista mia cavalcai quel giorno pensando molto e accompagnato da molti sospiri.

[...] **Ce passage présente un aperçu saisissant sur le sens de l'imagination médiévale, son pouvoir de consolation et sa force dans la saisie de l'expérience humaine** — en l'occurrence, il s'agit de la séparation amoureuse et de l'absence. Parce qu'il montre que ce pouvoir se confond avec celui d'une image, **il constitue un point de départ particulièrement opératoire pour une réflexion sur l'imagination pastorale**. C'est en effet précisément cette coïncidence entre imagination et image qui est au cœur de la réévaluation qui affecte le discours sur l'imagination à la fin de la Renaissance et au XVII^e siècle. [LIT053]

L'extrait cité ne peut être mis au rang d'une citation de travaux de pair sur l'objet étudié ; il sert aux dires de l'auteur de « point de départ » pour une réflexion sur cet objet (l'imagination pastorale dans *Astrée*), ou pour le dire autrement de mise en perspective du matériau principal avec d'autres textes, qui placent le premier dans la problématique de l'imagination médiévale dont propose de traiter l'auteur.

De la même manière, en LET009, le « plaidoyer pour le membre » tiré des *Essais* de Montaigne constitue le matériau principal, dont l'auteur s'emploie à étudier, entre autres, la rhétorique, en montrant comment celle-ci est annoncée dans d'autres textes de Montaigne.

Ces derniers sont intégrés au texte, de même que des extraits de textes d'autres auteurs, susceptibles d'éclairer les influences qui se manifestent dans cet essai. Ces autres textes du même auteur et ces textes d'autres auteurs servent en quelque sorte de matériau secondaire, qui se trouve intégré au texte tandis que le matériau principal se présente sous la forme de paragraphes séparés.

En LET042, une disposition spatiale spécifique des exemples atteste de cette mise en perspective du matériau principal avec d'autres textes : par l'utilisation de colonnes, l'auteur rend compte des échos que font d'autres textes à l'*Ovide Moralisé*, en reprenant la fable de Céyx et Alcyone :

Iris ne restera pas longtemps auprès du dieu du Sommeil, craignant d'être gagnée par la torpeur qui règne en ces lieux. Là encore, *La Fontaine amoureuse* fait en partie écho à l'*Ovide moralisé* :

Yris n'a pas attendu qu'il adjourne,	La messagiere atant <i>s'en torne</i>
Eins se depart et sans congié <i>s'en tourne</i> ,	Qui ja se sent endeble <i>et morne</i>
Car volentiers ilec pas ne sejourne.	Pour le someil qui leenz iere.
Ce fist li lieus,	Si s'en est repairié ariere
Qui la tenoit mate, endormie <i>et mourne</i> .	Par l'arc où elle estoir venue.
(<i>Fontaine</i> , v. 619-623)	(<i>OM</i> , v. 3510-3514)

[LIT042]

Le titre de l'article « Entre amour et politique : métamorphoses ovidiennes à la fin du Moyen Âge. La fable de Céyx et Alcyoné, de l'*Ovide moralisé* à Christine de Pizan et Alain Chartier » montre que l'*Ovide Moralisé* sert de point de départ, mais que ce statut de matériau principal tient pour beaucoup à la thématique du numéro où paraît la contribution. Cet article suggère que la distinction entre matériau principal et autres textes est délicate dans l'ensemble des cas où les différents textes n'obéissent pas à une hiérarchie quant à leur statut dans l'étude.

Sur la base du statut du matériau, couplé à la prise en compte de l'objectif annoncé de l'article, il semble ainsi possible de distinguer des types d'études en LET.

(1) Les articles qui concernent un texte

(2) Les articles qui concernent un auteur et l'un ou plusieurs de ses textes, catégorie très représentée dans notre corpus comme en atteste l'inventaire des matériaux en annexes 2. Ce type d'étude peut constituer l'objet même de certaines revues ; c'est le cas dans notre corpus avec le *Bulletin de la Société des Amis de Montaigne* et la *Revue de Montesquieu*, et c'est ce et qui peut expliquer la prédominance de ce types d'étude dans les articles analysés.

(3) Les articles qui explorent une question transversale à travers différents textes ; on peut alors spécifier cette dernière catégories, selon que la question est plutôt d'ordre linguistique (par exemple, l'énigme (LET028), les glossolalies (LET094) ou d'ordre historique (par exemple les expéditions de Jomard en Egypte (LET107)). Une dernière catégorie concernerait ainsi des faits textuels dans leur dimension historique (typiquement, en littérature comparée, les mythes comme par exemple Don Juan dans l'article LET103). Ces subdivisions restent périlleuses : on pourrait en effet évoquer aussi des études sur des faits historiques tels qu'ils sont traités dans les textes, comme c'est le cas dans l'article LET101, qui porte sur la nouvelle coloniale. La distinction entre les trois types d'études, quoique discutable elle aussi, semble cependant relativement opératoire dans le corpus, comme le montre l'inventaire du type de matériaux (cf. en annexes 2.).

Dans le troisième type d'études, la distinction entre matériau principal et secondaire n'est pas valide ; elle risque plus largement d'entraver l'analyse de l'objet même, car elle peut s'appliquer qu'aux études qui focalisent sur un texte (ou un ensemble de textes) plutôt que sur

sa mise en perspective avec d'autres. Or, c'est bien cette mise en perspective entre textes qui est massivement affichée comme objet même des études de LET.

Ainsi la catégorie intermédiaire des « autres textes », pour être opératoire dans certains cas, montre que c'est la question des sources primaires et secondaires qui se pose en LET, plutôt que celle du partage entre exemple et citation ; les sources sont toutes documentaires, et les sources primaires peuvent endosser une fonction critique qui se rapproche alors de celle de la citation des travaux des pairs ; c'est le cas lorsque l'auteur étudié se livre lui-même à des commentaires réflexifs sur son œuvre, celle d'un autre, ou sur la littérature de manière plus générale¹.

Finalement, dans la manière dont exemples et citations se présentent dans le texte, s'affiche l'idée que les articles de LET concernent un intertexte plutôt qu'un texte ; ce sont les relations entre textes qui font l'objet des thèses défendues par l'auteur et de ses investigations empiriques. Même lorsque l'étude se centre sur un texte principal, c'est bien sa mise en perspective avec un contexte historique, littéraire, idéologique, défini à l'aide de divers textes qui caractérise l'objet de l'article. Les textes littéraires, en tant que matériaux, peuvent fournir une fonction critique, historique, idéologique, ou un cadre littéraire d'ordre intertextuel, autrement dit un ensemble de fonctions qui couvrent un continuum entre sources primaire et secondaire.

Dans les articles de SCL : partage strict entre exemple et citation et diversité des exemples

Dans les articles de SCL, la distinction sémiographique entre les exemples et les citations, si elle n'est pas exclue en LET, se systématisent et concourt à un modèle de textualité qui manifeste très nettement un partage épistémologique entre ce qui relève de la documentation théorique et du matériau empirique.

Des types d'étude peuvent être mis en évidence sur la base des références des exemples ; ainsi la typologie proposée par C. Poudat (2006a) permet-elle de distinguer des exemples attestés, fabriqués et cités selon leur mode de référencement : les premiers sont associés à une référence exhaustive dans le cas de sources livresque, ou de références au lieu et temps de l'acte d'énonciation dans le cas d'interactions orales ; les seconds peuvent être introduits en tant qu'exemples fabriqués, ou se voir attribuer cette valeur par défaut en cas d'absence de référence. Les exemples cités à un autre auteur peuvent faire l'objet d'un double référencement qui permet de dire s'ils ont été fabriqués par cet auteur, ou attestés et cités par lui ; dans le cas d'une référence unique qui signale l'emprunt uniquement, une ambiguïté demeure sur la question de savoir si l'exemple est, originellement, attesté ou fabriqué, ambiguïté qui justifie cette troisième catégorie.

C. Poudat propose ensuite de distinguer au sein des exemples attestés 4 types d'exemples en fonction du type de discours dont ils relèvent, et que signalent les références : exemples littéraires, journalistiques, pédagogiques et exemples attestés « autres » (oral conversationnel, etc.).

Appliquée à notre corpus, la typologie montre un usage massif des exemples attestés, qui représentent 75% des exemples ; les exemples attestés « autres » sont les plus fréquents (la moitié), suivis des exemples littéraires (un tiers) et journalistiques ; la distribution des types d'exemples n'est pas soumise à un effet de l'auteur. Elle dépend par contre étroitement

¹ Cette fonction critique remplie par les « autres textes » pourrait expliquer, entre autres pistes interprétatives, la présence moins importante de citations observée dans les articles de LET par rapport à ceux de SCL. Mais ce constat demande prioritairement on l'a dit des investigations du côté des autres modes de références aux travaux (reformulation, évocation, etc.) dans les articles de Lettres.

des revues et numéros¹, ce qui permet d'étayer l'hypothèse explicative du poids des revues sur la différence observée entre le corpus de C. Poudat et le nôtre concernant l'usage relatif des exemples et des citations.

Cette typologie des exemples présente ainsi l'avantage de mettre en évidence la diversité des éléments qui prennent le statut d'exemple, et représente une entrée intéressante pour rendre compte de tendances épistémologiques dans la discipline, comme le suggère dans notre corpus l'usage massif des exemples attestés.

Les limites de la typologie se manifestent dans le poids pris par les exemples attestés qui ne sont ni littéraires, ni journalistiques, ni pédagogiques, et demanderaient donc à faire l'objet peut-être de nouvelles distinctions. Par ailleurs, les exemples littéraires et journalistiques sont bien représentés, mais la catégorie des exemples pédagogiques est assez peu productive ; cela ne remet pas pour autant en cause sa pertinence, bien qu'elle masque des différences importantes quant au statut linguistique de ce qui est pris comme exemple : les catégories des exemples journalistiques et littéraires renvoient à des discours et à des champs de pratiques bien identifiés, mais les exemples pédagogiques peuvent quant à eux regrouper des documents pédagogiques (manuels, instructions officielles), des interactions en situation pédagogique, ou encore des productions d'apprenants et des interviews d'enseignants et/ou d'apprenants à propos de leurs pratiques respectives d'enseignement et d'apprentissage.

Ainsi, certains exemples pédagogiques peuvent avoir un statut d'autonome et être étudiés en tant que phénomènes linguistiques (les productions des apprenants, la textualité des manuels, etc.), alors que pour les interviews, il peut davantage s'agir d'étudier leur contenu et les représentations sociales qu'ils véhiculent que les pratiques linguistiques.

L'entrée par les exemples demanderait à tenir compte conjointement de l'objectif annoncé de l'étude pour mettre en évidence des types d'études en SCL ; ainsi, les articles SCL048 et SCL053 concernent tous deux des textes littéraires dont des extraits sont cités pour exemples ; dans le premier cas, ces extraits cités sont analysés en tant qu'échantillons d'usages linguistiques en ancien français, tandis que l'analyse d'une nouvelle de Claude Simon dans le second vise à décrire les propriétés littéraires du texte sur la base de l'usage qu'il fait des ponctuations notamment.

L'étude des exemples et des citations montre que ces composantes structurelles ne font pas l'objet d'un même traitement dans les deux disciplines. En accord avec les tendances mises au jour quant aux modèles de textualité en usage, on remarque que les articles de SCL se caractérisent contrairement à ceux de LET par un partage sémiographiquement marqué entre sources documentaires et matériaux. Cette analyse suggère que ces deux composantes structurelles que sont les exemples et les citations ne vont pas de soi, mais résultent de pratiques d'écriture et d'épistémologies spécifiques ; elles se vérifient en SCL où la distinction renvoie à celle des sources théoriques et empiriques, alors qu'en LET, c'est vraisemblablement plutôt un partage entre sources documentaires primaires et secondaires qui prévaudrait, à l'instar d'une épistémologie propre à l'histoire notamment, et l'objet intertexte est assumé en tant que tel, jusque dans la disposition spatiale des extraits de textes cités dans certains articles.

¹ Si certaines revues semblent attester d'une préférence pour certains types d'exemples (les exemples attestés pour les *Cahiers du Français contemporain*, *Langage et Société*, *Faits de Langue* et les *Cahiers de pragmatique*, les exemples fabriqués pour les *Travaux de l'Institut de Phonétique de Strasbourg*), le choix mixte observable dans d'autres (exemples fabriqués et littéraires dans les *Recherches linguistiques de Vincennes* et dans *Langages*) suggère que le type d'exemples est largement conditionné par les numéros retenus dans le corpus.

L'homogénéité du traitement des sources en LET s'accompagne d'une homogénéité des matériaux, alors qu'en SCL c'est la diversité qui prime. Celle-ci montre l'intérêt qu'il y a à dépasser une conception restrictive de l'exemple en tant qu'énoncé autonymique fabriqué. Ces exemples méritent bien entendu toute l'attention des analyses, comme le montre la très intéressante proposition de C. Poudat (2006a) : elle montre comment il est possible de mettre au jour une doxa de l'exemple en linguistique, doxa fondé sur un langage ordinaire où certains emblèmes forts se manifestent¹, et qui suggèrent aussi une proximité entre les exemples fabriqués et les exemples attestés. Cela étant, il reste que de nombreux « exemples » de SCL ne se limitent pas à une fonction de validation d'un principe linguistique par une occurrence² langagière prise comme échantillon.

Pistes pour l'analyse de la notion d'exemple

Face à ce constat de la diversité des exemples et de l'irréductibilité du modèle de l'échantillon, nous proposons d'indiquer à partir du corpus de SCL deux pistes pour une réflexion sur l'exemple et l'exemplification, et la notion de preuve qui les caractérise.

Dans l'étude des styles que nous développerons *infra*, nous verrons que l'usage du syntagme « par exemple » chez l'un des auteurs du corpus est à ce point massif qu'il constitue un trait discriminant de ses textes. Or, un examen plus attentif peut laisser perplexe : l'auteur n'utilise pas, ou que très peu, d'exemples au sens de d'échantillons du langage démarqués par la disposition spatiale ou les guillemets. Cela étant, rien ne permet d'avancer que l'on utilise « par exemple » avec de tels échantillons, qui sont généralement introduits par une pratique de numérotation comme nous l'avons dit, et par ailleurs, si ce que l'auteur qualifie d'exemples ne renvoie pas à de tels échantillons, ce sont toutefois des phénomènes langagiers et plus précisément textuels qui sont concernés :

(1) **par exemple**, au sein du discours littéraire, à l'époque classique, le champ générique du théâtre se divisait en farce, comédie, comédie héroïque et tragédie. [A15-088 corpus styles]

(2) Or, un genre définit précisément un rapport normé entre signifiant et signifié au palier textuel : par **exemple**, dans le genre de l'article scientifique, le premier paragraphe, sur le plan du signifiant, correspond ordinairement une introduction, sur le plan du signifié ; dans le genre de la nouvelle, il s'agit le plus souvent d'une description. [A15-088 corpus styles]

(3) C'est aussi le moyen de sortir des apories théoriques suscitées par la philosophie du langage : par **exemple**, l'analyse de corpus reste le seul moyen éprouvé pour relativiser la polysémie et contrôler l'ambiguïté ; ou encore, pour déterminer les valeurs des formes grammaticales (par **exemple**, le futur n'a pas les mêmes valeurs dans le discours juridique que dans le roman). [A15-089 corpus styles]

Les exemples ci-dessus remplissent bien une fonction d'exemplification des principes et concepts introduits par l'auteur ; ils se démarquent cependant d'un modèle structural en linguistique car ils rendent compte de phénomènes langagiers sans toutefois en actualiser des occurrences. Fonctionnellement, ils sont en cela proches de citations ou du moins de références aux travaux de l'auteur ou de pairs dans la mesure où ils correspondent à des

¹ Pierre et Marie font ainsi partie, entre autres, des personnages récurrents de ce monde des exemples, de même que leurs activités respectives (*la voiture, la cuisine*) (Poudat, 2006), à l'image du fameux *chat sur le paillason* (à l'origine, *a cat on the mat*, qui a fait couler beaucoup d'encre dans les champs linguistiques, logiques et philosophiques anglo-saxons).

² C'est sur cette notion d'occurrence, opposée à celle de type, que P. Ouellet (1989) introduit sa réflexion sur le « statut cognitif et [la] portée argumentative de l'exemplification dans les sciences du langage ».

résultats de recherche sur les phénomènes langagiers, mais formellement, aucune référence n'est faite à un tel travail de recherche.

Envisagés comme composantes structurelles, les exemples ne seraient donc qu'une partie du procédé textuel de l'exemplification, qui demanderait à distinguer, comme le suggère P. Ouellet (1989 : 102) l'exemplification au moyen de faits empiriques, et celle qui passe par des faits scientifiques historiquement admis. La validation des propos apportée par l'exemplification se rapproche ainsi de la fonction de preuve apportée par la citation, et montre l'intérêt défendu par P. Ouellet de s'interroger sur le lien entre le cas exemplifiant et la proposition exemplifiée, et la fonction objectivante et problématisante¹ plutôt que simplement illustrative (i.e. légitimante et typifiante) de l'exemple (op.cité : 100-104)

Si le procédé d'exemplification peut se fonder non pas sur un exemple au sens strict d'échantillon brut, mais sur un phénomène textuel tel qu'il a été mis en évidence par l'analyse, les pratiques de l'exemple dans l'article rappellent que tout exemple est fabriqué, comme nous le montrerons à partir d'un autre extrait de notre corpus.

Prenons comme point de départ le balisage logique du document ; celui-ci pose la question de savoir comment déterminer ce qui relève de l'exemple, et ce qui relève des mots de l'auteur qui l'introduit et le commente. A l'instar de ce que nous avons évoqué à propos du péritexte *supra*, la numérisation du document peut ainsi servir d'heuristique dans l'analyse philologique des documents, y compris de support plus traditionnel.

Nous avons dit que les exemples pouvaient être intégrés à la syntaxe même du texte qui les introduit et les commente, mais le cas inverse se présente aussi, comme dans cet extrait d'article de SCL sur les pratiques langagières d'adolescents en milieu urbain :

1. K aimez-vous: les fricasses (**silence de 5 s., sourire crispé de I, recherche d'un contact oculaire avec E**)
2. I was ist das?! (qu'est-ce que c'est)
3. K ça c'est une question / aimez-vous les fricasses?! (**débit rapide**)
4. I no sé lo que es (**je ne sais pas ce que c'est, geste céphalique**)
5. K hein?!
6. I j(e) te répons aussi compréhensible que toi tu me poses la question j(e) comprends pas / tu comprends pas ma réponse (**sourire et ton un peu supérieurs, ou pour le moins satisfaits**)
7. K AIMEZ-vous les fricasses (**plus lent et articulé ; le r de fricasse semble moins postérieur**)
8. I ah:: les FLIcasses (moue) / ah les policiers (**rires des ados**).

[SCL009]

Les éléments mis entre parenthèses viennent compléter la transcription verbale des enregistrements des adolescents : l'auteur de l'article apporte des commentaires sur la réalisation phonétique et prosodique et sur les éléments mimo-gestuels. La difficulté de la notion d'exemple se manifeste : les commentaires servent à rendre compte dans un texte écrit de l'occurrence langagière prise comme exemple des pratiques adolescentes, et la question se pose donc de savoir si on doit restreindre l'exemple à la transcription fidèle des seuls propos, ou si on doit l'étendre à ces descriptions métalinguistiques, dans la mesure où elles participent de la transcription.

A l'encontre de l'idée que la preuve par l'exemple tient à son statut d'échantillon factuel, emblématique du phénomène qu'il entend illustrer, on voit ici comment l'exemple se présente en tant qu'échantillon des données dont traite l'auteur par l'entremise d'un travail métalinguistique. Le mythe, sans doute assez prégnant en SCL, de l'exemple attesté,

¹ P. Ouellet (1989 : 100) montre ainsi comment l'exemple, d'occurrence, peut passer au statut d'exemple type jusqu'à devenir un titre de problème pour un programme de recherche ; il en donne pour exemple (sic) un ouvrage de Lakoff qui prend pour titre un exemple de catégorisation dans une langue australienne à partir duquel l'auteur développe le problème de la typicité.

occurrence brute du langage qui a en cela statut de preuve mériterait que l'on s'interroge plus avant sur ce travail de l'analyste par lequel l'exemple est produit en tant que tel, pour servir sa fonction de validation. Pour rester limités, ces quelques constats sur l'exemple et l'exemplification suggèrent à titre programmatique que les textes peuvent fournir des pistes intéressantes pour leur analyse et pour s'interroger plus largement sur l'empiricité d'une discipline et sur le rôle de l'exemple dans la formulation des problèmes et la construction des objets ; une telle entrée permettrait ainsi de dépasser une conception positiviste de l'exemple, ou, pour le moins, de montrer comment celle-ci se positionne par rapport à d'autres épistémologies.

4.6. Macro-analyse du genre : bilan

L'analyse de la structure de l'article et de sa pluri-sémiotité atteste d'une forte diversité des pratiques d'écriture. Des modèles de textualité se profilent, qui peuvent être envisagés sur un continuum entre deux pôles. Le premier se caractérise par une exploitation corrélée du péri-texte (les annexes notamment), des indices de structuration et de guidage du lecteur (titres, numérotation, sous-parties, indices internes au paragraphe, détachement et numérotation systématique des exemples), et des formalismes de visualisation. L'autre pôle peut se définir négativement par l'absence de ces éléments.

Le premier exploite les possibilités structurelles et sémiotiques du texte, et se présente comme un assemblage composite, en termes d'espaces textuels investis et affichés comme tels (l'introduction, l'exemple, etc.), et par le recours à des éléments symboliques et graphiques réputés servir la production scientifique et sa communication. Au contraire, la textualité dans le second tend à s'en s'affranchir et se prête ainsi mieux que l'autre à une analyse restreinte aux mots ; recentrée sur le corps et les notes, elle se présente de manière plus homogène et unifiée.

Ces modèles rendent tangible un rapport différent au lecteur selon que les principes « reader-friendly » de guidage sont à l'œuvre dans le texte, ou que davantage de coopération est demandé à ce dernier, quand les instructions relatives à la structuration des contenus se limitent aux divisions en paragraphes et au matériau verbal. Avec les exemples et les citations, ces modèles de texte signalent également des images différentes de la recherche, et des tendances épistémologiques hétérogènes, y compris au sein d'une même discipline.

L'analyse textuelle se présente ainsi comme une entrée intéressante pour une analyse épistémologique, mais appelle aussi à déterminer le degré d'ostension ou de fonctionnalité des procédés rhétoriques, et à s'interroger sur un doxa de la science telle qu'elle se constitue dans certaines pratiques propres au discours scientifique et à sa circulation dans d'autres champs.

Notre macro-analyse du genre permet de conclure à des identités disciplinaires fortes, fondées sur des modèles de texte privilégiés, et, partant, sur un rapport spécifique au lecteur et à l'objet. Le point sur lequel les pratiques sont manifestement le plus contrastées concerne la gestion des sources avec deux tendances diamétralement opposées dans l'usage des notes (LET) et des bibliographies (SCL) et dans le traitement unificateur (LET) ou différenciateur (SCL) des sources théoriques et empiriques.

Les articles de LET assument un modèle contraire aux sciences dures et expérimentales, en revendiquant l'intertexte comme objet et en tendant vers le second pôle d'une textualité plus dissertative ; tandis que les LET affichent une création savante plutôt que scientifique, prioritairement fondée sur les mots et propice à la recherche d'effets stylistiques, on décèle en SCL la tentation d'un modèle dominant en science, séduit par la langue formelle et par la

distinction entre le matériau comme fait brut et son analyse¹. Cela étant, l'ensemble des traits textuels soumis à l'analyse fait état d'une diversité importante des pratiques dans le corpus de SCL, tandis que celui de LET, tout attestant d'une certaine variabilité se présente par contraste comme plus homogène, d'un double point de vue textuel et épistémique.

Le statut de l'auteur est quant à lui peu influent : les textes des doctorants se distinguent par une propension à des textes plus courts, au titrage plus systématique des introductions, et à l'exploitation plus importante des bibliographies et des annexes. Si l'on peut y voir le signe d'une évolution des pratiques d'écriture de l'article dans ce sens, le caractère marginal de ce facteur « statut de l'auteur » corrobore la validité de l'hypothèse des genres, en tant que normes textuelles résultant d'une régulation des usages au niveau éditorial.

L'effet de la revue et du numéro dont relève l'article, à relativiser au vu de la diversité observable, reste cependant important dans la structuration des articles et dans les types d'exemples. La question se pose alors de savoir si les revues qui bénéficient d'une notoriété plus grande, davantage lues, exercent un rôle plus important que d'autres dans la définition des normes du genre², mais aussi peut-être d'épistémologies dominantes dans la discipline.

Au final, nous retiendrons donc l'intérêt tout particulier d'une analyse du genre menée au niveau de la structure des textes, et qui permet de déterminer des modèles de textes sur la base de faisceaux de critères. Ces modèles ouvrent des pistes pour l'analyse de la textualité au sein d'un genre déterminé, en montrant notamment comment elle s'organise en différents espaces textuels. Ils peuvent être mis au compte de procédés d'explicitation maximale, avec des procédés qui visent à rendre la structure plus lisible, ou à l'inverse signaler un rapport au lecteur tendanciellement non didactique, et un surdestinataire (M. Bakhtine, 1984) tel qu'il est jugé apte à mettre en œuvre des inférences, avec en cela un possible effet de connivence. A titre de pistes pour des recherches ultérieures, ils signalent aussi avec la question des exemples l'intérêt d'une analyse de la construction linguistique du savoir, et du statut de la preuve en tant que caractéristique disciplinairement marquée.

¹ S. Shapin et S. Shafer, dans un essai consacré aux sciences expérimentales (1985) voient cette distinction comme une caractéristique du discours scientifique en général résultante d'une « technologie littéraire ».

² Nous en avons parlé au début de ce chapitre à propos du volume des textes, les revues les plus notoires se situant nettement dans une tendance centripète du genre.

Chapitre 5.

Éléments pour une caractérisation linguistique du genre.

Après avoir procédé à une macro-analyse du genre en traitant de ses composantes structurelles, nous poursuivons ici l'étude de l'article en prenant comme entrée le palier inférieur du texte (le palier méso-textuel¹ dans la terminologie proposée par F. Rastier, 2001). A partir des modèles de textualité que nous avons ébauchés, l'objectif est d'affiner la caractérisation du genre, et la comparaison des pratiques de chaque groupe disciplinaire (LET/SCL) et de chaque groupe d'auteurs (Doctorants/Non-doctorants).

Trois volets successifs sont abordés :

- dans la partie 2.1, les aspects morpho-syntaxiques du texte, et plus précisément : le découpage du texte en phrases et la structure phrastique, les parties du discours, les temps verbaux et les marques de personne et de non-personne ;
- dans la partie 2.2, les ponctuations fréquentes ou rares et en particulier des emplois spécifiques de certaines ponctuations à valeur énonciative (points de suspension, d'exclamation, d'interrogation et guillemets) ;
- dans la partie 2.3, les aspects lexicaux et plus précisément un lexique par hypothèse transdisciplinaire, que nous appelons ainsi pour rendre compte du fait qu'il est spécifique à l'activité de recherche dont il sert à désigner les composantes et les procédures.

Les variables ainsi retenues en raison de la pertinence qu'elles suggèrent pour l'étude de l'article font l'objet d'une analyse quantitative relayée par des observations plus qualitatives des énoncés en contexte ; par ailleurs, ce sont ces mêmes variables qui sont reprises dans le chapitre suivant, consacré à l'étude des usages stylistiques du genre (i.e. les usages que chaque auteur en fait) : cette étude appelait un dispositif méthodologique spécifique justifiant son traitement dans une partie séparée.

Les entrées, formelles, ont trait à la phraséologie et à l'énonciation représentée. Nous proposons de mettre à jour des caractéristiques centrales du genre (i.e. les traits qui le caractérisent le mieux) et des caractéristiques marginales (i.e. rares), et de les confronter aux caractéristiques attendues et supposées du discours scientifique. La comparaison des textes de SCL et de LET, et des doctorants et non-doctorants fait apparaître des caractéristiques

¹ Palier méso-textuel qui correspond schématiquement comme nous l'avons dit au palier intermédiaire entre le palier macro-sémantique du texte et le palier micro-sémantique du mot.

marquées, qui distinguent les groupes, et montre que les textes se positionnent diversement par rapport aux canons de scientificité.

Il s'agit également de questionner la fonctionnalité des usages et de voir, comme avec les composantes structurelles, ce que les aspects morpho-syntaxiques, la ponctuation et le lexique transdisciplinaire disent du rapport au lecteur et à l'objet. C'est sous cet angle des effets que nous cherchons à analyser comment les usages varient (en fonction des groupes et au sein de chacun d'eux) et comment ils se spécifient pour servir certains enjeux propres de l'article.

5.1. Les aspects morpho-syntaxiques

L'étude des aspects morpho-syntaxiques du genre nous a fait retenir, comme entrées qui semblent pertinentes pour le caractériser et mettre en évidence des usages contrastés, d'une part la longueur et la structure de la phrase, d'autre part ses constituants, et enfin les temps verbaux et les marques de personne. Ces aspects hétérogènes du niveau méso-textuel ont partie liée avec la structuration du texte (cf. les analyses sur la phrase), son objet (cf. les analyses sur les temps verbaux), son rapport au lecteur (cf. la syntaxe de la phrase et les temps verbaux) et l'inscription de son auteur (cf. les marques de personne et de non-personne).

A propos de la phrase dans le genre de l'article, nous commencerons par traiter de la structuration du texte en unités phrastiques, à partir des deux variables que sont le nombre de phrases par texte et le nombre de mots par phrase. Nous nous intéressons alors à la phrase en elle-même, avec les propositions qui structurent sa syntaxe et mettrons en évidence à travers ces deux aspects des caractéristiques marquées, en termes d'usages disciplinaires et d'usages liés au statut de l'auteur.

5.1.1. Le découpage du texte en phrases.

Le nombre de phrases dans le corps de l'article¹ est extrêmement variable : si le corpus pris dans son ensemble atteste d'une moyenne de 301 phrases par texte, le nombre de phrases se situe entre 49 et en LET, et entre 97 et 1053 en SCL.

Ces différences dépendent cependant de la taille des textes. Il convient donc d'analyser le nombre de phrases en fréquence relative (ratio sur le nombre de mots du texte). La distribution des valeurs ainsi obtenues fait état d'une différence disciplinaire significative² ; elle se vérifie également au niveau du nombre moyen de mots par phrase³ et c'est ce dont rend compte le graphique suivant :

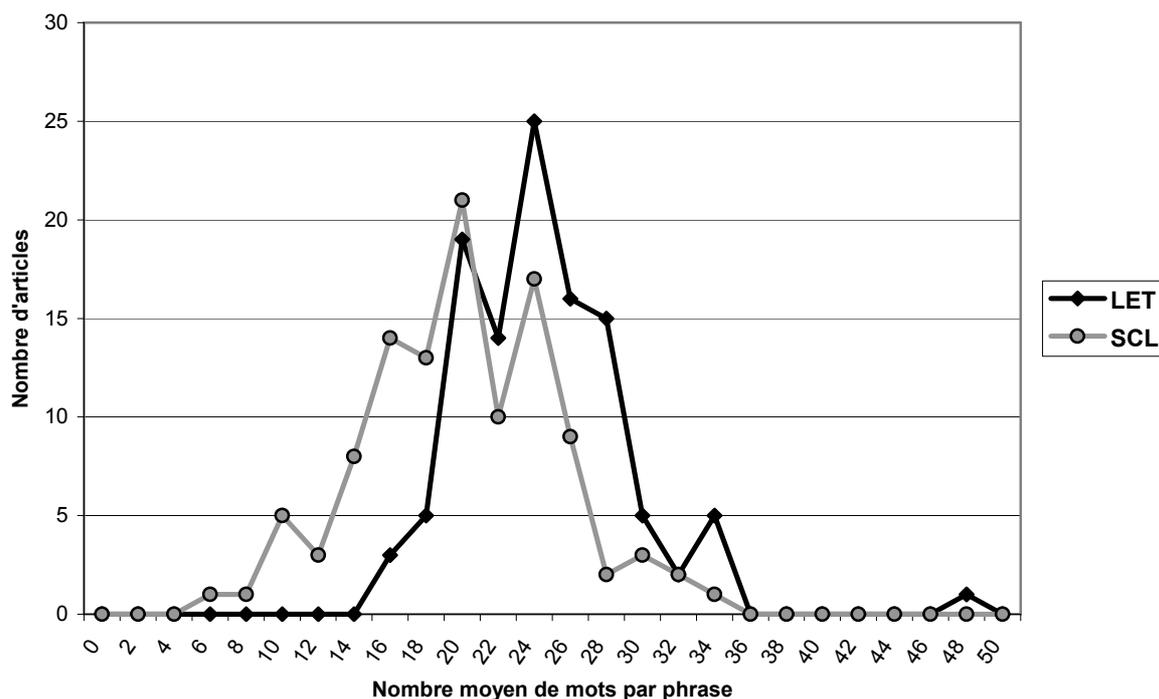


Figure 5.1. Distribution en fréquence du nombre moyen de mots par phrase : comparaison disciplinaire.

La différence de distribution en LET et en SCL sur ces deux variables du nombre de phrases (en fréquence relative) et du nombre moyen de mots par phrase montre deux tendances.

D'une part, les textes de SCL se caractérisent par un découpage plus important du texte en phrases, et, corrélativement, par des phrases plus courtes en nombre de mots. En moyenne, la phrase dans l'article de SCL représente 19 mots, contre 23,5 en LET. Cela rejoint ce que nous avons dit de la structuration du texte, et en particulier des paragraphes, également plus nombreux en SCL qu'en LET.

D'autre part, comme le suggère le graphique ci-dessus pour le nombre moyen de mots par phrase, les valeurs sont plus homogènes en LET qu'en SCL, ce qui confirme là encore la première analyse des modèles de texte dans les deux disciplines. Bien qu'un article en LET ait valeur d'hapax avec un nombre maximum de 46,1 mots par phrase, qui dépasse le maximum

¹ L'étude est menée sur le corps de l'article (i.e. l'article à l'exclusion des notes, annexes et bibliographies), mais nous évoquons cependant les données obtenues pour le corps de l'article sans les exemples ni les citations. Ces données sont celles fournies par le logiciel d'analyse *Cordial*.

² *Mann Whitney*, $U=3180$, $p<0,0001$.

³ *Anova*, $F=43,703$, $p<0,0001$.

de 34 mots par phrase enregistré en SCL, la déviation standard reste plus faible en LET. La phrase en LET n'outrepasse pas la limite inférieure de 15,3 mots en moyenne tandis que 18 articles de SCL présentent un modèle de texte aux phrases très courtes, avec un nombre moyen de mots compris entre 5 et 15.

On pourrait voir dans cette tendance l'effet des exemples de SCL, peut-être plus courts que ceux de LET qui consistent en des citations de textes littéraires comme nous l'avons vu en traitant de la différence de statut des exemples dans les deux disciplines. Cependant, l'effet significatif de la discipline sur le découpage du texte en phrases se vérifie aussi bien dans le corps du texte que dans le corps sans les exemples et citations¹. La taille des exemples et citations doit toutefois jouer, mais dans les deux disciplines, où le nombre moyen de mots par phrase est légèrement supérieur une fois les exemples et citations supprimés du texte.

Caractéristique des articles de SCL, la tendance à des textes davantage découpés en phrases se renverse en revanche dans les notes : si là aussi la discipline influe significativement sur les valeurs, c'est en LET cette fois que l'on observe un nombre de phrases plus important et des phrases plus courtes². Cette inversion de tendance suggère une différence de fonction des notes et peut être comprise en référence à ce que nous avons dit de l'usage différent des notes dans les deux disciplines : la fonction de référencement bibliographique qu'elles servent en LET donne lieu à des phrases plus courtes que pour d'autres fonctions, d'autant que, par souci d'économie dans les notes, des abréviations des noms d'ouvrage sont utilisées en LET comme nous l'avons dit.

Cela étant, la différence entre les deux disciplines au niveau de la phrase dans les notes tient aussi à une variabilité plus grande entre les articles de SCL qu'entre les articles de LET. Comme la référencement des sources en notes (référence exhaustive, ou sous la forme d'un indicateur de référence qui renvoie à son tour à la bibliographie)³ est rare mais possible en SCL, certains articles de SCL se rapprocheraient à ce niveau du profil de LET. Plus largement, cette variabilité s'inscrit dans la continuité des résultats mis en évidence *supra*, d'une variabilité des modèles de textes caractéristique des articles de SCL par rapport à ceux de LET.

Cette hypothèse disciplinaire sur les notes est corroborée par l'absence d'effet du statut de l'auteur sur la phrase dans cet espace textuel⁴, alors qu'il joue dans le corps du texte. En effet, la propension observée en SCL à davantage découper le corps du texte en phrases est vraie aussi des doctorants par rapport aux autres scripteurs⁵. Dans le corps de l'article, le nombre moyen de mots par phrase est de 19,3 dans les articles de doctorants contre 21,6 chez les autres. C'est de cette différence dont rend compte le graphique ci-après, où est représentée la distribution en fréquence du nombre moyen de mots par article dans les textes des deux groupes de scripteurs.

¹ *Anova*, $F=25,2$, $p<0,0001$.

² Pour le nombre de phrases en fréquence relative, *Anova*, $F=8,074$, $p=0,0049$, et pour le nombre moyen de mots par phrase, *Anova*, $F=14,168$, $p=0,0002$.

³ Cf. les différentes stratégies de renvoi aux sources.

⁴ La note se présente en cela comme un espace relativement figé du genre, dont la variabilité tient principalement à l'exploitation de cette fonction spécifique de référencement bibliographique.

⁵ Pour le nombre de phrase en fréquence relative, dans le corps de l'article *Anova*, $F=7,309$, $p=0,0074$ (effet significatif, mais non significatif dans les notes). Pour le nombre moyen de mots par phrase, dans le corps de l'article *Anova*, $F=5,749$, $p=0,0173$ (effet significatif), dans les notes *Anova*, $F=0,205$, $p=0,65$ (effet non significatif).

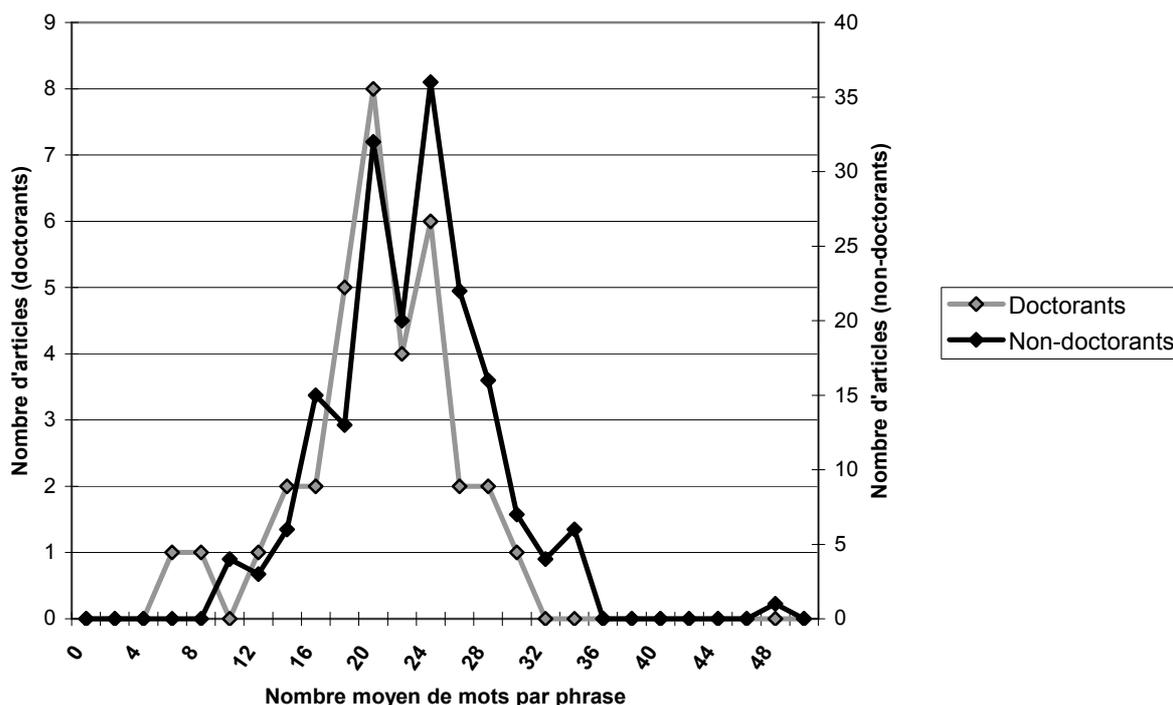


Figure 5.2. Distribution en fréquence du nombre moyen de mots par phrase : comparaison des deux groupes d'auteurs.

La tendance centrale signale ainsi des phrases plus longues chez les non-doctorants que chez les doctorants, et les valeurs minimales et maximales corroborent ce partage entre les deux groupes : c'est dans un article de non-doctorant (LET087) que l'on trouve le nombre moyen de mots par phrases le plus élevé (avec 46,1 mots) et dans un article de doctorant (SCL075) que l'on compte le plus de phrases et corrélativement le nombre moyen de mots par phrase le plus faible (avec 5,35 mots).

Là encore, les poids des exemples doit jouer : ainsi cet article présentant les phrases les plus courtes comporte-t-il de nombreux énoncés très brefs, utilisés pour l'étude de la réalisation du [R] français par des étudiants brésiliens. Mais la significativité de la variable « auteur » se vérifie dans le corps de l'article, de même qu'une fois soustraits les exemples et les citations¹. Au final, on obtient donc schématiquement un continuum des modèles de textes en termes de découpage croissant des phrases qui va des articles de doctorants à ceux des non-doctorants et des articles de SCL à ceux de LET : les articles de SCL produits par des doctorants sont ceux dont les textes présentent les phrases les plus courtes, et au pôle inverse les textes avec les phrases les plus longues sont ceux de LET produits par des non-doctorants.

La question du découpage du texte en phrases montre donc une tendance au découpage plus important de l'article en phrases en SCL par rapport aux LET, qui confirme ce que nous avons dit des modèles de textualité dans les deux disciplines ; elle corrobore aussi l'idée signalée avec ces modèles que la force centripète du genre est plus marquée en LET qu'en

¹ La différence liée au statut de l'auteur dans le corps de l'article sans les exemples et les citations donne, pour le nombre de phrases (en fréquence relative) *Anova*, $F=7,872$, $p=0,0055$, pour le nombre moyen de mots par phrase *Anova*, $F=8,089$, $p=0,0049$.

SCL où l'on aurait donc des articles proches du modèle en LET, et d'autres plus singuliers. Le fait que cette tendance à des phrases plus courtes soit également caractéristique des textes de doctorants par rapport à ceux des autres scripteurs suggère une possible évolution du genre vers ce découpage plus important du texte en phrases. De manière plus nette encore qu'avec les modèles de textualité ébauchés au chapitre précédent (chapitre 4), il semble pouvoir indiquer chez les doctorants, comme on l'a vu pour les SCL, des textes plus didactiques, obéissant à des principes réputés facilitateurs pour le lecteur.

5.1.2. La syntaxe de la phrase

Si la phrase présente un profil différent dans les disciplines et en fonction du statut de l'auteur au niveau du nombre de mots qu'elle contient, c'est vrai aussi de sa structure syntaxique et en particulier de l'usage des subordonnées.

La phrase la plus employée dans le genre de l'article est sans conteste la phrase indépendante, à hauteur de 85% pour le corpus pris dans sa globalité. Ce constat se vérifie également une fois les exemples et citations soustraits du corps du texte, et à plus forte proportion dans les notes, avec un taux de 92,8% de phrases indépendantes, et corrélativement, un taux plus faible de phrases avec subordonnée.

Le graphique suivant montre le taux moyen de propositions subordonnées dans le corps de l'article, le corps sans exemples ni citations, et les notes, pour le corpus pris dans son ensemble (« Total ») et dans chaque groupe disciplinaire (LET/SCL) et d'auteur (D/Nd).

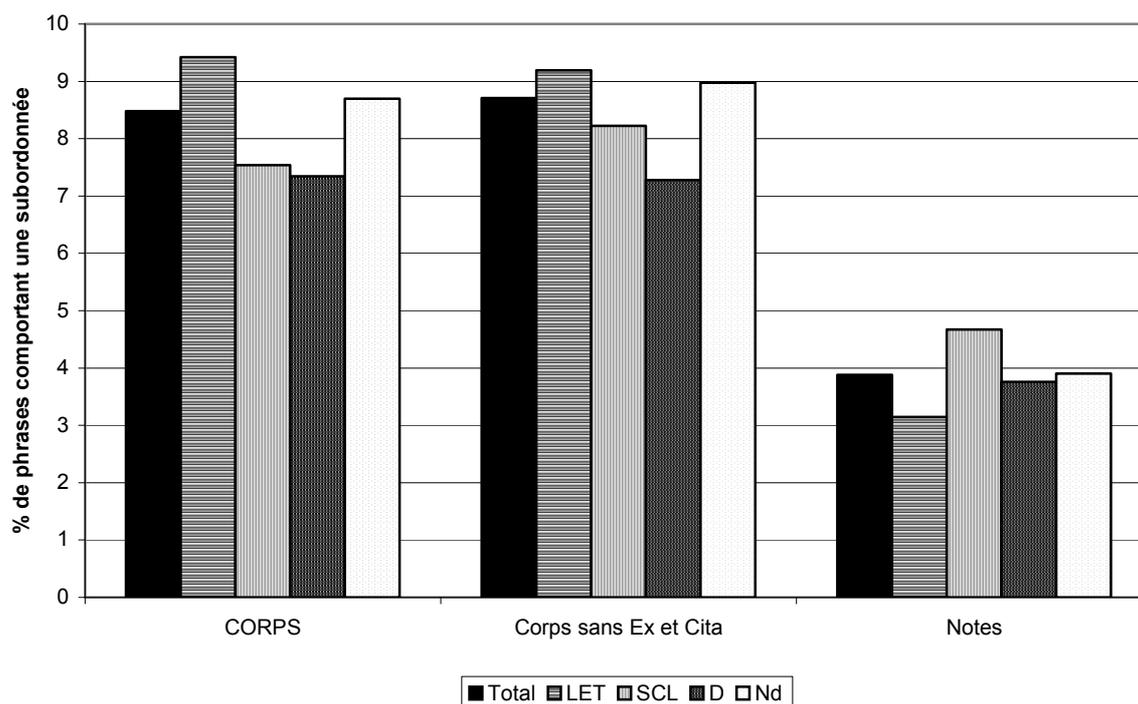


Figure 5.3. Taux moyen de phrases comportant une subordonnée dans le corps, le corps sans les exemples ni les citations, et les notes : comparaison entre les groupes.

Le taux nettement plus faible de subordonnées dans les notes signale de nouveau la spécificité de cet espace textuel et rejoint ce que nous avons dit de sa fonction de référencement bibliographique ; une différence disciplinaire se manifeste en effet, dans la proportion plus

faible de subordonnées dans les notes en LET par rapport aux SCL¹. La fonction spécifique des notes en LET par rapport à celles de SCL expliquerait cette tendance à un usage moindre des subordonnées dans les notes de LET, alors que, comme pour la longueur des phrases, cette tendance en notes est à l'opposé de ce que l'on observe dans le corps du texte.

Dans le corps du texte, avec ou sans exemples et citations, le recours aux subordonnées est en effet caractéristique des articles de LET par rapport à ceux de SCL². Ainsi, en moyenne dans le corps du texte, 9,4% des phrases comportent une subordonnée en LET, contre 7,5% en SCL. La tendance centrale de l'usage des subordonnées est donc plus élevée en LET, et si les indices de dispersion sont proches, on voit que la différence se joue non pas tant au niveau des valeurs minimales (1,4% de phrases avec subordonnées en SCL et 1,5% en LET) qu'au niveau des valeurs élevées, puisque le maximum atteint en SCL est de 13,8% de phrases avec subordonnées, alors qu'en LET il s'élève à 20,4%.

Plus largement, l'ensemble des variables relatives aux types de propositions convergent à mettre en évidence une phrase syntaxiquement plus complexe en LET : de manière significative, on note parmi l'ensemble des propositions une proportion plus importante en LET de relatives (avec une moyenne de 2,7% contre 2,1% en SCL), de subordonnées d'un autre type (avec une moyenne de 11,1% en LET contre 9,7% en SCL), et de coordonnées, ces dernières restant toutefois marginales dans le genre (avec une moyenne de 0,085% en LET et de 0,035% en SCL)³. Dans le même sens, la proportion de propositions comportant un objet direct et indirect est significativement plus importante en LET qu'en SCL (avec une moyenne de 36,11 en LET contre 30,2 en SCL pour les propositions avec objet direct, et 34,4% en LET contre 27,6% en SCL pour les propositions avec objet indirect)⁴.

Comme pour les disciplines, les données sur le volume de la phrase rejoignent celles sur sa syntaxe pour les deux groupes de scripteurs doctorants et non-doctorants. La subordination caractérise négativement les textes des doctorants par rapport à ceux des autres scripteurs (avec un taux moyen de 7,4% de phrases contenant au moins une subordonnée chez les doctorants contre 8,7% chez les autres) : les textes des doctorants sont donc marqués par des phrases plus courtes et par un usage moindre de subordonnées⁵, ce qui se vérifie aussi dans les textes des doctorants de SCL en particulier⁶ (en moyenne, 5,9% de phrase avec subordonnée contre 9,4% dans les textes de non-doctorants de SCL). Quoique de manière moins frappante qu'entre les deux disciplines, les autres variables (coordonnées, objets directs et indirects) confirment une distinction liée au statut de l'auteur entre des textes tendant vers des phrases courtes, et des textes tendant vers une syntaxe complexe⁷.

Ces distinctions quant aux composantes de la phrase et au niveau notamment de la longueur et de l'importance prise par les subordonnées se retrouvent avec les mots-outils et les mots signifiants, qui fournissent ainsi un critère supplémentaire dans les profils de textes. Ces deux catégories des mots outils et des mots signifiants s'expliquent par un critère fonctionnel qui appelle à dissocier ce qui relève du lexique (noms, verbes, adjectifs et

¹ *Anova*, $F=6,992$, $p=0,0088$.

² La différence disciplinaire donne, dans le corps de l'article *Anova*, $F=17,38$, $p<0,0001$, et dans le corps de l'article sans les exemples ni les citations, *Anova*, $F=6,99$, $p=0,0088$.

³ La différence disciplinaire donne, pour les propositions relatives, *Anova*, $F=8,273$, $p=0,0044$, pour les autres subordonnées *Anova*, $F=7,621$, $p=0,0063$, et pour les propositions coordonnées *Anova*, $F=6,55$, $p=0,0112$.

⁴ Pour les propositions avec objet direct, *Anova*, $F=60,7$, $p<0,0001$, et pour les propositions avec objet indirect, *Anova*, $F=75,5$, $p<0,0001$.

⁵ *Anova*, $F=4,521$, $p=0,035$.

⁶ *Anova*, $F=8,63$, $p=0,004$.

⁷ Cf. en annexes pour le détail des résultats.

adverbes) et ce qui relève des relations grammaticales entre mots et groupes de mots dans la structure (inter-)phrastique (prépositions, conjonctions, pronoms), ou de l'actualisation des syntagmes verbaux et nominaux (déterminants). Le partage – établi ici par le logiciel d'analyse, sur la base de la reconnaissance formelle des unités à partir de son dictionnaire – reste bien entendu discutable : pour n'en citer qu'un exemple, le verbe *être* peut être mis en tant que verbe au rang des mots signifiants, mais joue un rôle grammatical plutôt que lexical dès lors qu'il sert d'auxiliaire.

Le taux moyen des mots-outils et des mots signifiants est très proche dans l'article, quoique les seconds soient sensiblement plus importants : en moyenne sur l'ensemble du corpus, les mots-outils représentent 49,8% des mots et leur proportion peut varier entre 31,8% et 55,6%. Les mots signifiants représentent en moyenne 50,2% des mots et leur proportion peut quant à elle être comprise entre 44,3% et 68,2%. Le rapport entre mots outils et mots signifiants peut donc être compris entre un peu moins de 0,5 et 1,25¹.

Ce constat se vérifie une fois les exemples et les citations ôtés du texte ; ceux-ci ne parasiteraient donc pas outre mesure l'analyse à ce niveau : les distributions attestent dans leurs tendances centrales d'un rapport entre mots-outil et mots signifiants proche de 1, et compris entre 0,66 et 1,25. Se manifeste ainsi la valeur d'hapax du texte pour lequel, avec les exemples et les citations, la proportion des mots-outils par rapport aux mots signifiants n'atteint pas 0,5 : cette proportion retrouve une configuration normale (i.e. un rapport avoisinant 1) une fois extraits les exemples fréquents de cet article consacré à l'adverbe.

L'usage des mots-outils et des mots signifiants fait l'objet de différences significatives, d'une part entre les articles des deux disciplines, d'autre part entre les articles des deux groupes d'auteurs (doctorants et non-doctorants). Difficilement interprétable en lui-même, ce constat prend cependant du sens en termes de profils de textes, une fois rattaché au faisceau de critères déjà décrits à propos de la phrase.

Concernant les usages disciplinaires, on observe que les articles de SCL recourent significativement plus aux mots signifiants et, inversement, moins aux mots-outils que ceux de LET².

Concernant les usages liés au statut de l'auteur, ce sont les textes de doctorants qui présentent significativement plus de mots signifiants³ ; le recours plus important aux mots-outils caractérise donc à l'inverse les textes des non-doctorants.

Une dernière différence intervient qui renforce les deux tendances adverses mises en évidence entre un usage plus important des mots-outils en LET d'une part et chez les non-doctorants d'autre part, puisque les non-doctorants de LET recourent significativement plus que les doctorants de LET aux mots-outils⁴.

Il est donc possible de caractériser les groupes de manière contrastée selon que le rapport entre mots-outils et mots-signifiants tend à l'avantage des premiers ou des seconds, comme le résume le schéma ci-après.

¹ I.e. entre 1 mot-outil pour 2 mots signifiants et 2,5 mots-outils pour 2 mots signifiants.

² *Anova*, $F=35,311$, $p<0,0001$.

³ *Anova*, $F=6,799$, $p=0,0097$.

⁴ *Anova*, $F=5,206$, $p=0,0245$.

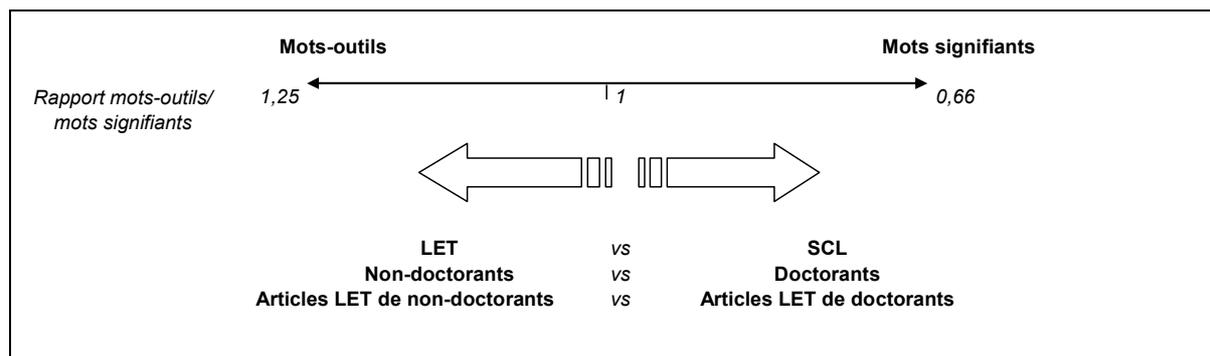


Figure 5.4. Des tendances contrastées dans l'usage des mots-outils et des mots significants : bilan.

Insistons cependant sur la relative homogénéité des valeurs au sein du corpus, et sur la proximité des distributions, qui présentent un profil tendanciellement gaussien ; en dépit de caractéristiques marquées dans l'opposition des groupes, et principalement au niveau de la différence disciplinaire, l'unité du genre se manifeste ainsi dans ces résultats.

Cependant, ces tendances contrastées font écho à ce que nous avons dit de la longueur des phrases et du type de propositions qu'elles comprennent. Elles corroborent les précédentes analyses : les mots-outils caractérisent les articles de LET et de non-doctorants, qui font un usage plus important de propositions subordonnées et de coordonnées, et conséquemment des éléments grammaticaux nécessaires pour les introduire. Inversement, les phrases plus courtes opéreraient une restriction sur ces unités grammaticales, au profit de l'importance donnée dans aux composantes lexicales¹, et déterminent le profil des articles de SCL par rapport à ceux de LET, des doctorants par rapport à ceux des non-doctorants et des doctorants de LET par rapport à ceux des non-doctorants de LET. Comme pour la structure des textes, des profils nets de la phrase se manifestent, et une véritable spécificité des textes de non-doctorants de LET.

Les différents indices relatifs à la structure de la phrase rejoignent les résultats mis en évidence à propos du découpage du texte en unités phrastiques. Ils signalent une tendance à la concision et à la simplicité syntaxique d'une phrase principalement centrée sur ses composantes lexicales dans les articles de SCL par rapport à ceux de LET, et dans ceux des doctorants par rapport à ceux des non-doctorants. S'il faut insister là encore sur la variabilité observable à l'intérieur de ces groupes, ces résultats suggèrent néanmoins un rapport différent du texte au lecteur avec d'un côté une écriture qui le guiderait davantage par le choix de la concision, et de l'autre, une écriture qui exploite davantage les possibilités de la subordination et de la coordination dans la phrase.

Les deux entrées prises en compte dans l'étude de la phrase attestent ainsi d'une continuité avec les traits décrits au palier supérieur du texte à propos de leur structure. Des caractéristiques marquées se manifestent dans des modèles scripturaux spécifiques à chaque groupe, mais il importe d'insister sur l'extrême variabilité des pratiques textuelles, qui engage à envisager en termes de continuum plutôt que de modèles opposés les caractéristiques

¹ Celles-ci font également l'objet de différences inter-groupales, sur la base de l'usage contrasté des adjectifs *versus* des adverbes et des substantifs *versus* des verbes, mais comme nous l'avons expliqué au chapitre 3., de tels résultats sont écartés de notre analyse car ils offrent trop peu de prise sur les textes ; les catégories linguistiques sont en cause : le phénomène de la dérivation lexicale d'une part et d'autre part l'hétérogénéité de la catégorie des adverbes notamment entravent l'analyse de la fonctionnalité dans les textes de telles entrées linguistiques.

groupales ainsi signalées. C'est dans cette même perspective que nous poursuivrons notre étude linguistique du genre, avec les temps verbaux et les marques de personne.

5.1.3. Les temps verbaux

Notre approche morpho-syntaxique du genre s'est intéressée jusqu'ici à une analyse quantitative des traits généraux relatifs à la phrase et ses constituants. Nous allons à présent nous centrer sur les temps verbaux, avant d'aborder les marques de personne et de non-personne.

Avec les temps verbaux, il sera de nouveau question de caractéristiques centrales du genre, conformes aux caractéristiques attendues du discours scientifique, et de caractéristiques groupales contrastées. L'analyse quantitative sera relayée par un examen qualitatif en contexte : il s'agit de sonder quelques emplois récurrents des « tiroirs verbaux »¹ dont la comparaison des textes des différents groupes indique le potentiel discriminant dans la description de l'article et de mettre ainsi au jour certaines séquences et procédés textuels qui concernent doublement le rapport du texte à son lecteur et à son objet.

L'étude des temps verbaux est menée sur le texte sans les exemples et citations, afin d'éviter de parasiter l'analyse par les caractéristiques de ces derniers.

Le premier constat que l'on peut faire est conforme aux caractéristiques attendues du genre de l'article : c'est le présent de l'indicatif qui domine largement le corpus, en représentant 80,6% des verbes conjugués. Si cette proportion varie considérablement selon les textes, puisqu'elle peut aller de 7,5% à 94,7% des verbes conjugués, la prédominance du présent apparaît toutefois comme une caractéristique centrale relativement stable du genre dans la mesure où les indices de dispersion sont assez peu élevés et où les variables indépendantes n'influent pas de manière significative sur la distribution des valeurs.

Les autres temps verbaux attestent en revanche de différences importantes entre les groupes. C'est ce que suggère le graphique suivant pour les temps principaux à l'exclusion du présent, au seul niveau des taux moyens de chaque temps par rapport à l'ensemble des verbes conjugués et ce pour l'ensemble du corpus (« Total ») et pour chacun des groupes considérés : groupes disciplinaires (LET et SCL), et groupes liés au statut de l'auteur, « doctorants » (D) et « non-doctorants » (Nd).

¹ L'expression, que l'on doit à J. Damourette et E. Pichon (1911-1940) vise à rendre compte du fait que les temps grammaticalement marqués dans la morphologie des verbes n'ont pas nécessairement ni exclusivement une valeur temporelle.

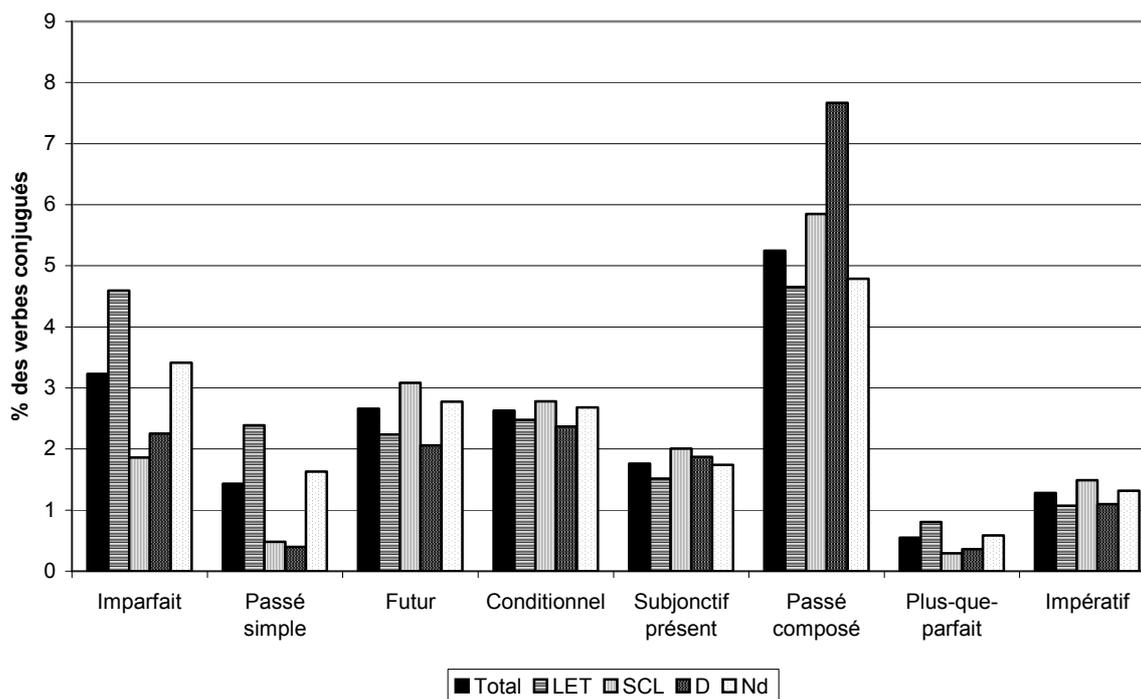


Figure 5.5. Taux moyen des principaux temps verbaux à l'exclusion du présent : comparaison entre les groupes.

Les temps verbaux non représentés sur ce graphique peuvent être considérés comme des variables extrêmement marginales du genre de l'article, avec un taux moyen dans le corpus qui est au plus de 0,28% des verbes conjugués pour le conditionnel passé, et, moindre encore pour le futur antérieur, le passé antérieur, le subjonctif imparfait, passé et plus-que-parfait. Cela ne signifie pas pour autant qu'ils ne fassent pas l'objet de différences disciplinaires, nous y reviendrons. Les différences statistiques de la distribution des temps verbaux dans les textes des deux disciplines concernent en effet 10 des 15 temps identifiés par l'analyseur, ce qui suggère l'intérêt d'une telle entrée dans l'étude d'usages contrastés du genre.

Parmi les temps verbaux qui font état d'une homogénéité entre les groupes disciplinaires et les groupes d'auteur, notons avec l'usage du présent celui du conditionnel présent. Il s'agirait donc là aussi d'une caractéristique stable du genre, qui en dépit de son caractère marginal en termes quantitatifs met en évidence l'intérêt de ce temps verbal dans les textes scientifiques, en particulier pour le rôle modal qu'il joue dans l'assertion en vertu de sa dimension polyphonique¹.

Comme nous l'avons dit, la grande majorité des autres temps verbaux se distribue de manière significativement différente dans les deux disciplines. Quelques-unes de ces différences affectent également le partage entre les textes des doctorants et les textes des autres scripteurs. Elles concernent, principalement, le passé simple, l'imparfait, le passé composé, le futur et l'impératif. Nous traiterons du futur et de l'impératif conjointement, après nous être intéressé aux temps du passé.

¹ Pour les contributions les plus récentes sur les problèmes et les enjeux polyphoniques du conditionnel en français, cf. notamment, avec les travaux de Kronning (2005), le numéro de *Recherches linguistiques* qui lui est consacré (Dendale et Tasmowski, 2001).

Les temps du passé

En réunissant les principaux temps de l'article à l'exception de l'indicatif présent, le graphique proposé ci-dessus montre, en creux, l'importance prise par ce dernier, qui représente comme nous l'avons dit 4/5^{ème} des verbes conjugués. Les quelques rares textes affichant un taux très faible de verbes au présent sont en cela suffisamment marqués pour mériter qu'on les regarde de plus près. On remarque ainsi qu'il s'agit de textes qui font usage de temps du passé et qui comportent une dimension historique dans leur objet. En atteste l'exemple ci-dessous, tiré de l'article LET080, qui compte 7,52% de verbes au présent, soit une valeur très faible qui s'explique par des taux forts de 51,8% de verbes à l'imparfait et de 27,9% au passé simple.

Qui **était** donc **chargé** de la littérature dans *La Petite République* ? Pour la période Millerand, entre 1893 et 1897, les choses **firent** claires : la critique des livres **était tenue** par Georges Renard (1847-1930), normalien, secrétaire de Rossel pendant la Commune, exilé en Suisse où il **continuait** à vivre après avoir obtenu un poste à l'université de Lausanne, aidé, à partir du début de 1895, par son épouse, Louise, de dix ans sa cadette. Dès lors tous deux **alternèrent** leurs articles, publiés sous le titre "Politique et Littérature" le lundi (numéro daté du mardi), selon une tradition déjà bien établie (les *Lundis* de Sainte-Beuve...). [LET080]

Le récit d'événements passés qui caractérise ce passage et l'ensemble de l'article trouve d'autres marques dans les repères chronologiques (datation et adverbies circonstanciels comme « *dès lors* », qui servent à inscrire les procès dans le passé et à les situer les uns par rapport aux autres) et dans l'abondance de noms propres. On en trouve trace, entre autres, dans un autre article du même numéro de la revue dix-neuviémiste *Romantisme*, numéro consacré à « la littérature fin de siècle au crible de la presse quotidienne ».

Le présent de vérité générale, qui contribue à la portée universalisante du discours scientifique n'est bien entendu pas exclu de ces textes historiques. Dans le second texte de LET que nous venons d'évoquer, on voit ainsi comment, dans l'introduction, le repérage dans le passé des faits relatifs à la *Revue des Deux Mondes* cède la place au présent pour la présentation d'une problématique fondée sur l'histoire de cette revue :

Lieu d'une mémoire sans doute, **la Revue des Deux Mondes fut** aussi, avec quelques autres, lieu d'expérimentation et surtout lieu de réaction devant les innovations du XIX^e siècle. La continuité qui caractérise celle qui fut, et de loin, la plus importante des revues françaises, incarnée par quelques hommes – François Buloz puis son fils Charles, enfin Ferdinand Brunetière – , **permet d'en retracer l'histoire** dans le temps long, **tout en l'interrogeant dans quelques moments cruciaux, au prisme desquels s'éclairent des logiques structurales voire quasi-ontologiques du périodique**. Ainsi, questionner la *Revue des Deux Mondes* au temps de l'entrée dans un « régime médiatique » offre des modalités efficaces d'appréciation des réactions de la culture lettrée devant la montée en puissance de la presse quotidienne en même temps que de son insertion dans le monde de l'imprimé et plus largement dans le régime culturel qui détermine la seconde moitié du siècle (...)

La Revue des Deux Mondes regarda la presse avec une inquiétude tenant à la fois de l'anxiété devant la montée du fait démocratique dans la société française et de la solidité de sa propre position dans l'espace public. **À cet égard, elle eut à défendre ce qui constituait sans doute son magistère le plus fort : la critique littéraire, à la fois arme de combat et modalité essentielle de son expression publique**. [LET082]

L'objet de l'auteur se trouve ainsi circonscrit par ces glissements entre le récit historique de procès délimités par le passé simple comme faits importants du passé, et l'atemporalité du présent de définition. Pour être liées à des types d'étude spécifiques, ces formes de narration ne sont pas exclues en SCL, où l'on peut citer, par exemple, un article consacré à une histoire politique des réformes orthographiques (SCL011).

Cette dimension historicisante peut concerner l'ensemble de l'article ou une seule de ses parties ; c'est le cas par exemple dans SCL082, qui porte sur la langue portugaise du Brésil : pour traiter la question (« *Le 'Brésilien' existe-t-il ?* ») de savoir si le « brésilien » et le portugais peuvent être considérés comme deux langues, une partie du propos consiste en l'analyse de la langue portugaise du Brésil sous un prisme historique, et passe linguistiquement par l'usage de l'imparfait et du passé simple. Le taux de présent, qui se situe dans une tranche intermédiaire de 53% suggère cependant comme on le vérifie dans le texte que le mode narratif, pour être bien présent, n'est le fait que de cette partie spécifique de l'article. Inversement, des valeurs très faibles du présent de l'indicatif (autour de 20% pour le premier texte de LET cité, i.e. LET080, et de même pour SCL011) marquent l'importance de la narration dans des textes dont l'objet central est d'ordre historique.

La distribution des temps verbaux dans le corpus met en évidence l'existence d'une opposition entre l'usage du présent et celui de l'imparfait d'une part (avec une corrélation négative à hauteur de 85,1%), et celui du passé simple d'autre part (avec une corrélation négative acceptable comme telle de 77,5% dans l'ensemble du corpus, très nette en LET avec un taux de 84,5% mais non significative en SCL avec seulement 65,3%)¹. L'imparfait et le passé simple sont également corrélés, mais de manière positive, dans les deux disciplines et à hauteur de 79,8% en SCL et 86,7% en LET. Le graphique suivant rend compte de cette tendance marquée d'un usage inversement proportionnel du présent d'une part et du passé simple et de l'imparfait de l'autre ; il ne porte que sur les articles de LET où l'opposition entre présent et passé simple est plus marquée qu'en SCL.

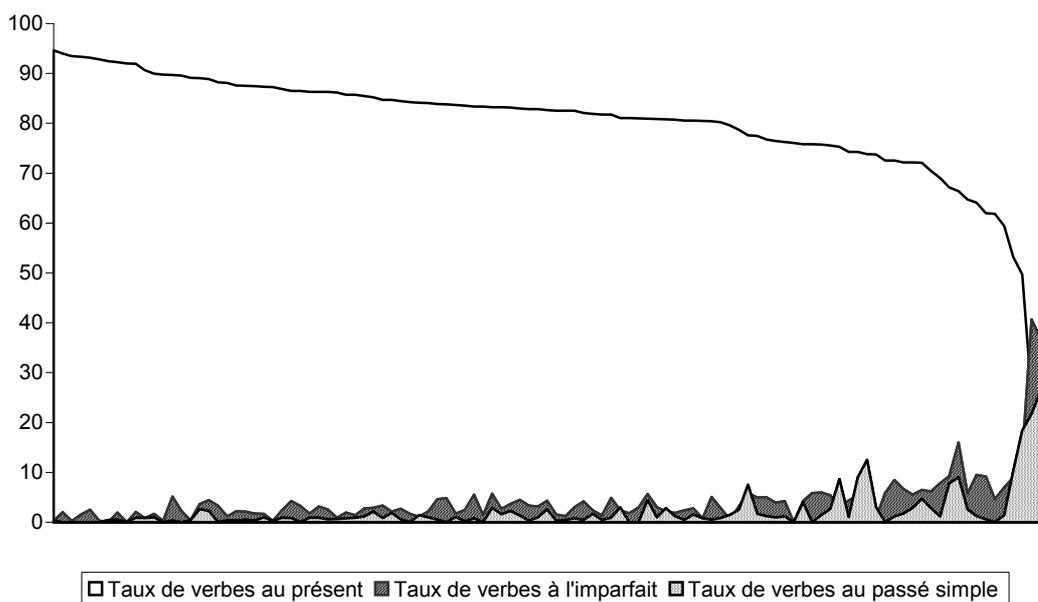


Figure 5.6. Une corrélation négative dans l'usage du présent, du passé simple et de l'imparfait en LET.

¹ Cf. la matrice de corrélation fournie en annexes. Rappelons qu'une corrélation peut aller de 0 à 1 (identité exacte de la distribution des deux variables) (i.e., en pourcentage, de 0 à 100%) et que nous considérons une corrélation comme telle à partir de 0,75 ou 75%.

La corrélation positive au contraire entre les deux temps du passé simple et de l'imparfait se conforme aux limites signalées de la distinction benvenistienne entre discours et récit¹ dans l'analyse des textes. Croisée aux autres résultats, elle indique un mode narratif (Weinrich, 1973, Adam, 1999) qui se caractérise par un usage conjoint de divers temps du passé. Ainsi, en SCL où contrairement à l'opposition présent/imparfait, l'opposition présent/passé simple n'est pas significative, la corrélation entre passé simple et imparfait s'associe à des corrélations positives avec le plus-que-parfait, vérifiées également en LET.

Le caractère marqué de cette dimension narrative dans le genre de l'article rejoint les résultats obtenus dans l'étude sur les styles d'auteur menée avec C. Poudat (cf. chapitre 6), et dans l'entreprise de profilage et de classification des articles de linguistique réalisée par cette dernière (Poudat, 2006a) : les textes comprenant de tels modes narratifs fédèrent une classe spécifique de traits caractéristiques du genre, qui tiennent à l'usage conjoint de certains temps verbaux, de certains types de connecteurs et de certains numéraux de datation ; ils présentent également une spécificité potentielle au niveau de la progression thématique comme nous le verrons dans l'étude des styles (chapitre 6).

Le mode narratif mis en évidence par l'analyse des temps verbaux permet ainsi de dépasser une première approche intuitive du genre, qui pour se fonder sur une caractéristique effectivement centrale, l'emploi massif du présent, reste trop unifiante. En particulier, il est intéressant de constater comment le récit d'événements passés peut impliquer l'intrusion d'un *je* narratif relativement étonnante par rapport aux traits supposés de l'écrit scientifique.

C'est le cas d'un article précédemment évoqué, et consacré à « *Une politique de la variante : brève histoire du rapport sur les Rectifications de l'orthographe* » (SCL011). On y relève, avec seulement 23,9% de verbes au présent, un taux de 26,7% verbes à l'imparfait et un taux moindre de 16,1% de verbes au passé simple, ainsi que l'importance du plus-que-parfait et du passé composé. Ce texte présente une mise en scène singulière du « je » de l'auteur, en tant que « je » historique, objectivé par le récit, et qui trouve sa place dans l'article en raison du rôle politique de l'auteur dans la réforme dont il traite.

Ainsi, dès mai 1988, **Michel Rocard m'avait chargé de lui proposer, puis de mettre en place**, une politique de la langue française qui ne soit plus l'expression de l'éternelle crispation anglophobe sur une ligne maginot de "défense" de la langue française, mais, à l'inverse, une politique dynamique qui prenne en compte la nécessité – notamment pour promouvoir internationalement le français, d'agir sur la langue elle-même, de la vitaliser, la moderniser, la faire bouger. **Je songeais** au jugement de Proust : "La seule manière de défendre la langue française, comme l'armée dans l'affaire Dreyfus, c'est de l'attaquer."

Mon premier mouvement fut d'évoquer la mise en application de la circulaire Fabius sur la féminisation des titres et noms de métiers. Signée aux derniers jours du ministère Fabius le 6 mars 1986, elle avait été immédiatement enterrée par son successeur, Jacques Chirac, auquel succédait Rocard. Il s'avéra rapidement qu'il n'y avait plus aucun groupe de pression pour la réactiver (...) [SCL011].

Ces occurrences d'un *je* narratif par lesquelles l'auteur se présente comme témoin de l'histoire se manifestent dans un autre article sous une forme un peu différente, dans le sens où elle servent la mise en scène du chercheur en tant que tel, et non plus dans les implications politiques de sa fonction comme c'était le cas avec l'extrait précédent :

¹ E. Benveniste (1966, 237-250) distingue deux plans d'énonciation sur la base de l'usage des temps verbaux mais également des marques de personne et de non-personne. Les énoncés ancrés dans la situation d'énonciation (« discours ») interdiraient l'usage du passé simple caractéristique des énoncés se présentant comme coupés de la situation d'énonciation (« récit »). L'imparfait quant à lui pose problème car, s'il est analysé par E. Benveniste comme s'opposant au passé simple, l'auteur souligne lui-même qu'il peut être attesté dans l'un et l'autre des plans d'énonciation, ce qui montre donc que le partage des formes linguistiques entre les deux plans d'énonciation n'est pas systématique.

Raoul de la Grasserie **a longtemps été pour moi** un mystère. C'est dans un article de Konrad Koerner consacré à l'histoire de la sociolinguistique que **j'ai pour la première fois découvert** son nom (...). Il y avait donc là une œuvre apparemment importante et dont **je ne savais rien**.

J'en ai un jour parlé à Patrick Mathieu, qui achevait à cette époque sous ma direction une thèse consacrée à l'argot, et c'est lui qui a trouvé à la Bibliothèque Nationale deux des textes dont je traiterai ci-dessous, me photocopiant un article (« La psychologie de l'argot ») et une partie d'un ouvrage (Etude scientifique sur l'argot et le parler populaire) dont il me résumait le reste, tandis que, tout récemment, Konrad Koerner retrouvait et m'envoyait l'article qu'il avait cité sur la « sociologie linguistique » [A3-01)]¹.

La mise en scène de l'auteur en tant que sujet historique, dans des passages narratifs du texte, pourrait ainsi permettre d'envisager des liens avec l'historiographie ou encore avec l'écriture de recherche en anthropologie, seule discipline peut-être à avoir thématisé², en raison de son objet propre, la question de la narration et celle de la mise en scène dans le texte de l'auteur en tant que sujet expérientiel et témoin du terrain pris comme objet d'étude.

Pour en revenir au mode narratif de manière plus générale, cette caractéristique marquée du genre de l'article et attestée en LET comme en SCL se révèle cependant discriminante entre les deux disciplines. Ainsi les articles de LET affichent-ils une prédilection pour la narration en recourant systématiquement plus que ceux de SCL aux temps du passé : ils utilisent davantage de passé simple et d'imparfait (avec un taux moyen de 4,6% de verbes à l'imparfait en LET contre 1,8% en SCL, et de 2,4% de verbes au passé simple contre 0,5% en SCL)³, et, partant, de plus-que-parfait, de conditionnel passé et enfin de subjonctif passé et imparfait alors que le subjonctif présent est plus fréquent en SCL.

Il est possible de voir dans cet usage disciplinairement marqué le fait de tendances épistémologiques. Le recours à un mode narratif se prêterait à certains types d'études plutôt qu'à d'autres ; la dimension historique de l'objet, possible dans les deux disciplines, caractériserait cependant mieux les articles de LET que ceux de SCL. Un tel constat fait écho à ce que nous avons mis en évidence à propos des types de matériaux de l'article dans le chapitre 1. Il peut être mis en perspective aussi avec les textes structuralistes réputés fondateurs de la linguistique contemporaine, qui définissent prioritairement son objet en synchronie.

Cependant, le caractère rétrospectif d'une étude est à distinguer de son traitement historicisant, puisque l'on peut fort bien, et on le voit en LET comme en SCL, revenir sur des événements passés en usant du présent ou encore, nous allons en parler, en recourant au passé composé. Si l'on peut voir dans l'usage d'un mode narratif une tendance épistémologique, c'est donc aussi dans le sens où assumer un traitement narratif de l'histoire est peut-être plus marginal en SCL qu'en LET. Du reste, un tel mode n'implique pas nécessairement le passé simple.

Il nous faut à présent souligner que si le passé simple caractérise positivement les textes de LET par rapport à ceux de SCL, il s'avère également discriminant dans la comparaison des deux groupes de scripteurs doctorants et non-doctorants. Les textes des doctorants pris dans leur ensemble recourent significativement moins que les autres scripteurs

¹ Cet extrait ne relève pas du corpus principal étudié ici mais du corpus « styles » (cf. la partie consacrée à l'étude des styles d'auteur) : c'est le texte numéroté 1 de l'auteur A3.

² En témoigne en particulier le numéro 58 de la revue *Communications* consacré à *L'écriture des Sciences de l'homme* (Perrot et Soudière, 1994), et en particulier les trois articles sur l'écriture en anthropologie de G. Balandier, de F. Affergan et de M. Kilani, ainsi que l'étude de M. Amorim (1996), *Dialogisme et altérité dans les sciences humaines*.

³ La différence disciplinaire donne, pour l'imparfait, *Mann Whitney*, $U=2908$, $p<0,0001$ et pour le passé simple, *Mann Whitney*, $U=2817$, $p<0,0001$.

au passé simple, avec un taux moyen de 0,4% contre 1,6% dans les textes des autres scripteurs, et, plus parlant encore, un maximum faible de 2,6% contre 27,9% dans les textes des autres scripteurs¹.

Ce constat se vérifie également au sein des articles de LET², les textes des non-doctorants de LET se présentant donc comme ceux qui font l'usage le plus abondant du passé simple et d'un traitement narratif des faits historiques. On pourrait voir dans ces résultats l'indice d'une possible évolution du genre de l'article vers un usage moindre du passé simple. Une telle hypothèse fait écho au recul signalé du passé simple dans le français contemporain de manière générale ; aussi son emploi marqué dans l'article pourrait-il signaler, au-delà du mode narratif, une spécificité stylistique, à mettre peut-être au compte de la littérarité caractéristique de ce temps verbal selon R. Barthes (1972 : 25). En cela, il peut attester d'un certain « goût de l'écriture »³ à l'œuvre dans les articles de LET.

Le potentiel discriminant du passé simple est à mettre en perspective avec les autres différences de distribution des temps verbaux dans le corpus. En particulier, s'il caractérise positivement les textes de LET par rapport à ceux de SCL, ces derniers se définissent positivement par l'usage plus abondant du passé composé, du futur et de l'impératif. Avant d'en venir à ces deux derniers, évoquons le passé composé qui se présente dans des emplois très variés dont nous ne fournirons qu'un aperçu rapide et partiel.

Non seulement les deux groupes disciplinaires s'opposent sur la base du passé composé, mais la différence de distribution concerne aussi, au sein des SCL, les textes des doctorants et des non-doctorants : les articles des doctorants recourent significativement plus que les autres au passé composé⁴ ; son usage s'opposerait ainsi au futur d'une part, qui caractérise positivement en SCL les textes de non-doctorants, et au passé simple, qui caractérise positivement les textes de l'ensemble des non-doctorants par rapport à ceux des doctorants.

Les usages du passé composé, variés on l'a dit, confirment que l'usage d'un mode narratif au passé simple n'est pas nécessairement requis par la dimension historique de l'étude. *A fortiori* dans les textes des doctorants, l'article en SCL tend à privilégier le passé composé au passé simple lorsqu'il s'agit de rendre compte de procès en tant que faits accomplis dans le passé ; on a donc affaire à des formes de récit au passé composé plutôt qu'au passé simple.

C'est ce que l'on observe par exemple dans cet extrait d'un article de doctorant qui traite de « *La transmission intergénérationnelle du valencien et son usage comme langue seconde* » et comporte une partie historique consacrée au processus de substitution linguistique dans le pays valencien. Celle-ci est menée principalement à l'aide du passé composé, dont le texte fait un usage abondant avec 19,8% des verbes conjugués concernés, et se caractérise aussi par une absence complète de passé simple.

Au pays valencien, la substitution linguistique (ou castillanisation) a été un processus lent et progressif et ce, dans trois directions principales, selon les différentes périodes de l'histoire. Au XVI^e siècle, l'extension du castillan a suivi une direction "horizontale" (en ce qu'elle se déplace à l'intérieur d'une même strate sociale, à savoir l'aristocratie et le haut clergé). La castillanisation a eu lieu également dans les affaires administratives et dans la littérature savante (Sanchis Guarner 1963).

¹ La différence liée au statut de l'auteur donne, pour le passé simple, *Mann Whitney*, $U=2543$, $p=0,044$.

² Avec un taux moyen de 0,45% dans les textes de doctorants contre 2,65% dans les textes de non-doctorants, et des résultats tels que la différence est significative (*Mann Whitney*, $U=277,5$, $p=0,001$).

³ L'expression est tirée de l'article de P. Sansot (1994) qui s'intéresse à l'écriture des sciences humaines et pose la question de savoir dans quelle mesure ce goût de l'écriture peut être envisagé comme une « dérive épistémologique ».

⁴ Avec un taux moyen de 9,3% contre 6% dans les textes des autres scripteurs et une différence significative (*Mann Whitney*, $U=636,5$, $p=0,035$).

La victoire des Bourbons, lors de la Guerre de Succession au tout début du XVIII^e siècle, a impliqué l'abolition du royaume de Valence, l'imposition des lois de Castille et la naissance de l'État-nation espagnol. Les nouvelles lois linguistiques, qui ont promulgué l'exclusivité du castillan, n'ont pas affecté toutefois le peuple valencien, qui a continué à parler sa propre langue. [SCL069]

Les formes de récit dans l'article peuvent bien entendu ne pas respecter une bipartition complète entre passé composé et passé simple : nombre d'articles présentent un usage de temps très hétérogène, qui montre la complexité de cette question du système verbal¹. Par ailleurs, le passé composé peut remplir des fonctions spécifiées par rapport à celle de la chronologie des événements passés. Parmi l'ensemble de ses emplois, il en est un qui pose de nouveau le problème de la mise en scène de l'auteur dans son texte², et qui concerne les procès liés à la démarche de recherche.

C'est ce qu'illustrent ces différents extraits d'un texte qui affiche le taux maximum dans le corpus de 37,3%³ de verbes au passé composé :

Nous avons choisi d'étudier la sonorité du \R\ français réalisé par les étudiants brésiliens connaissant leur difficulté à produire cette consonne. (...)

Une fois établi le corpus, nous avons sélectionné des étudiants qui parlaient aisément le français, avec une expérience d'apprentissage de 3 à 4 ans. Des 11 locuteurs sélectionnés, 5 sont du sexe masculin et 6 du sexe féminin. Ils ont entre 18 et 30 ans. (...)

Par rapport à l'énoncé 1, nous avons relevé trois types de manifestations de la sonorité de \R\ réalisé par les différents locuteurs : 6 cas où elle a été réalisée totalement sourde, 4 cas où elle a été réalisée totalement sonore et un cas où elle a été réalisée très assourdie, à l'exception de la fin de la durée (...)

Toutefois, en position finale de mot mais pas d'énoncé, l'occurrence de réalisation complètement sourde de la consonne dans un contexte sonore a été de 33,33 % et en contexte sourd de 51,52 %, pourcentages assez expressifs. Il est important de souligner qu'en position intervocalique, la désonorisation totale de la consonne a été significative (57,59 %). [SCL075]

Le passé composé alterne avec un présent de vérité générale, et, en caractérisant des procès d'ordre scientifique (*sélectionner*, *relever*, etc.) ou langagiers (*réaliser*), il sert la mise en scène de l'auteur lui-même dans la démarche scientifique et des sujets qui ont été soumis à l'expérimentation. Ce type d'emploi du passé composé est manifestement prisé pour rendre compte de la méthodologie de la recherche dans les articles de SCL.

Il semble principalement mobiliser sa valeur aspectuelle d'accompli du présent : il ne s'agit sans doute pas tant de situer les procès dans le passé que de mettre en évidence leur caractère accompli et leur résultante dans le présent. Cela étant, la dimension temporelle n'est pas écartée, et c'est vraisemblablement même cette notion d'antériorité qui paraît quelque peu surprenante en conclusion (les 4 dernières lignes citées), lorsque l'auteur fait le bilan sur ses résultats en les présentant tels qu'ils ont été produits (« *la désonorisation totale de la consonne a été significative* »), plutôt que d'énoncer cette significativité au présent, comme une propriété à valeur de loi.

Ce texte est produit par un néophyte (l'auteur annonce que l'étude présentée est issue d'un travail de maîtrise) et les résultats quantitatifs suggèrent que l'on peut voir aussi dans l'abondance de ces tours un effet spécifique du statut de l'auteur ; les non-doctorants

¹ C'est ainsi que les usages diversifiés de l'imparfait appellent des études spécifiques de ce tiroir verbal, dont les valeurs ne sont pas faciles à interpréter hors des séquences narratives où il s'associe au passé simple et/ou au passé composé, pour présenter les faits d'arrière plan par rapport aux repères temporels installés par ces deux temps verbaux.

² Comme nous l'avons vu *supra* avec la mise en scène d'un témoin historique dans des séquences narratives.

³ Rappels qu'en raison des difficultés que pose la reconnaissance automatique du passé composé, il est possible que du bruit explique pour partie ce taux élevé, en particulier au niveau de l'ambiguïté formelle entre passif et passé composé ; il reste que le texte comporte effectivement un nombre très important de verbes au passé composé.

opéreraient davantage alors pour un usage du présent dans ce type de tours, selon une possibilité attestée dans ce même article où la mise en scène de l'auteur et/ou des sujets dans le compte-rendu de l'expérimentation se fait parfois aussi au présent.

Le futur et l'impératif

Nous nous intéresserons pour finir aux deux tiroirs verbaux du futur et de l'impératif présent, qui s'avèrent, rappelons-le, caractériser positivement les articles de SCL par rapport à ceux de LET¹. Si l'impératif introduit une visée temporelle de futur, dans la mesure où sa valeur injonctive implique la réalisation à venir du procès, et si le futur peut lui-même se voir associer une valeur injonctive, nous les traitons ensemble en raison d'un emploi commun largement attesté dans l'article, et qui explique en grande partie sans doute le caractère marqué de ces deux traits dans les textes de SCL par rapport à ceux de LET.

Le futur et l'impératif s'inscrivent en effet dans la majorité de leurs emplois dans des formes de métadiscours, que l'on peut définir en référence aux travaux de J. Rey Debove (1978) comme une forme spécifique de métalangage où les autonymes désignent le texte lui-même et les opérations accomplies par l'auteur dans sa construction. En cela, ces usages métadiscursifs correspondent à des balises textuelles qui servent d'instructions au lecteur.

Le métadiscours a fait l'objet de nombreux travaux, et en particulier dans l'écrit de recherche dont il représente une caractéristique centrale comme le montrent, entre autres, A. Mauranen (1993), S. Thompson (2003), T. Dahl (2003), et K. Hyland (1998, 2005). Sans entrer dans une analyse détaillée de ce phénomène, nous proposons très modestement d'en signaler quelques usages à partir de cette entrée spécifique constituée par les temps verbaux.

Dans son étude du métadiscours dans l'article de recherche menée à partir du corpus KIAP, T. Dahl (2003 :121) signale l'extrême plasticité de la notion de métadiscours, qui dans son acception large de discours sur le discours peut aller jusqu'à inclure les connecteurs ou les marqueurs de l'attitude du scripteur. Le problème tient à ce qu'il n'est pas possible d'envisager un partage strict entre le métadiscours et les énoncés qui servent l'apport d'informations : les travaux de J. Authier-Revuz (1995) sur la connotation autonymique l'ont montré, un grand nombre de phénomènes linguistiques attestent d'une double fonction d'usage et de mention du signe. Aussi la restriction aux seuls items concernant le texte et/ou le guidage du lecteur est-elle en général défendue dans les travaux sur l'écrit de recherche (par exemple, Breivegga, Dahl et Fløttum, 2002, Mauranen, 1993, Dahl, 2003), même si elle ne va pas, là encore, sans poser de problèmes de frontières.

T. Dahl (op.cité : 122-125) propose ainsi d'identifier trois catégories principales de métadiscours : les éléments locationnels, qui réfèrent au texte ou à ses sections et intègrent en particulier de ce que D. Maingueneau (1994) traite en termes de deixis textuelle (des balises telles que *maintenant*, *ici* qui relèvent de l'ordre du texte et non de la situation d'énonciation), et d'autres indices de structuration (*dans cette partie*, *ensuite*, *supra*, etc.) ; les éléments rhétoriques, d'ordre lexicaux, qui concernent plus spécifiquement les verbes signalant un procès relatif aux fonctions rhétoriques du texte (*introduire*, *conclure*, *discuter*, *souligner*, etc.) ; enfin, les instructions explicites au lecteur, dont le traitement dans une catégorie spécifique se justifie selon T. Dahl (op.cité :124) en particulier par l'usage de l'impératif.

Comme le suggère l'auteur, il reste cependant difficile de faire le partage entre un métadiscours explicitement adressé au lecteur et un métadiscours qui servirait la construction du texte en termes de structuration et d'opérations rhétoriques de l'auteur ; l'ensemble des

¹ La différence disciplinaire donne, pour le futur, *Anova*, $F=8,649$, $p=0,036$, pour l'impératif *Anova*, $F=11,284$, $p=0,0009$.

balises textuelles servent d'instructions au lecteur en guidant ce dernier, qu'elles soient linguistiquement adressées au lecteur, qu'elles marquent la présence de l'auteur dans son texte, ou en désignent uniquement des parties sur le mode du délocuté.

La valeur temporelle du futur et de l'impératif dans la réalisation du procès en fait manifestement des tiroirs verbaux privilégiés de ce type de balises, en servant d'annonce à ce qui va suivre. C'est ainsi que l'on observe un usage important du futur dans les introductions, comme dans ce texte de SCL où l'on compte un taux important de plus de 10% de verbes au futur, avec dans cet extrait un nombre important de verbes rhétoriques et d'indicateurs locationnels selon les catégories de T. Dahl (op.cité) :

J'avancerai ci-après sous forme de postulat, **le raisonnement qui guidera ma démarche** :

— dans un discours encyclopédique **on avancera** qu'un N d'objet — quel que soit son statut : objet social ou objet physique — **désignera** un objet en tant qu'il est construit par des productions discursives savantes.

— **il s'ensuivra** qu'un N.ENTREE DE DICTIONNAIRE (soit N.ENTREE) aura pour référentiel non pas l'objet lui-même mais sa description.

— **on dira alors** qu'un tel objet **aura** le statut d'objet de connaissance, c'est-à-dire de représentation d'objet conçue dans un réseau d'énoncés dont la validité relève d'une épistémè. [SCL101]

On retrouve de telles balises qui servent un effet d'annonce dans le corps du texte également, en guise de transitions :

C'est cet aspect politique de l'eau qu'**illustrera**, en complément, ma seconde fable. [LET004]

Ce double emploi du futur métadiscursif dans les introductions et le corps du texte se vérifie avec l'impératif, et ce dans les deux disciplines :

Comme la réponse à cette question ne saurait être donnée d'avance, **partons d'abord en exploration**. [LET066]

Mentionnons, enfin et surtout, la mode française qui foisonne depuis l'an 2000 en lançant des constructions par préfixation de l'adjectif *electronic* tronqué en *e-*. *Libération* titre, le 29 août 2000 : « le cyber-libraire américain inaugure aujourd'hui 'Amazon.fr'. [SCL018]

L'usage des temps verbaux dans ces balises joue, on le voit, un rôle central dans la structuration des contenus, en servant la représentation de la dimension spatiale du texte (*comme nous l'avons dit/ comme nous le verrons*) (voir par exemple Maingueneau, 1994).

De telles balises textuelles peuvent être plutôt, ou à la fois, « writer- and reader-oriented » selon la distinction souvent reprise dans les travaux anglo-saxons sur les textes dits académiques entre ce qui relève du marquage du scripteur et ce qui relève de celui du lecteur. La distinction concerne les fonctions pragmatiques des traits textuels, et vise à rendre compte de la dimension interactionnelle du texte. Ce qui est « writer-oriented » a trait à la manière dont l'auteur promeut son approche, et ce qui est « reader-oriented » a trait aux formes de politesse, à la recherche d'un consensus ou d'une coopérativité du lecteur. Concernant plus spécifiquement le métadiscours, elle est explicitée dans (Hyland, 2005).

L'usage de l'impératif à la première personne du pluriel suggère que c'est en fonction du sémantisme du lexème lui-même que la différence pourrait se faire ; ainsi, « mentionner » dans l'extrait ci-dessus serait-il plutôt « writer-oriented », de même que l'ensemble des verbes qui impliquent la mise en scène d'un locuteur (*signaler, dire, rappeler, étudier, etc.*) ; mais les formes relativement fréquentes de *voir* ou de *noter* sont ambiguës de ce point de vue, comme c'est le cas aussi par exemple dans « À titre d'illustration préliminaire, lisons ce premier

passage, tiré d'une pièce de J. Giraudoux » (SCL093), où deux figures d'un locuteur et d'un interlocuteur sont conjointement mises en scène par le sémantisme même du verbe.

Au-delà de ces seules balises, la 1^{ère} personne du pluriel de l'impératif manifeste un usage du *nous* qui ne résiste pas à la réduction supposée d'un *nous* académique très prisé dans l'écrit de recherche français, où il renvoie à l'auteur¹. Si ce *nous* académique, qui exclut nettement le lecteur est employé dans ces balises (« *Nous renvoyons donc le lecteur au dictionnaire ou au lexique publié dans nos trois articles déjà cités* », SCL085), il semble qu'elles se caractérisent fondamentalement par une implication mutuelle de l'auteur et du lecteur. C'est ce que l'on voit dans cet extrait où l'impératif intervient après une série d'interrogations oratoires²,

Pourquoi étudier le paradigme *burlesque* dans le *Dictionnaire* de Furetière ? Quel intérêt, à côté de paradigmes sérieux, comme *noblesse*, *commerce* ou *étranger* ? Ne s'agit-il pas tout simplement d'une catégorie esthétique à la fois limitée et de rang hiérarchique bien inférieur, qui ne saurait constituer qu'une entrée fort étroite dans une entreprise si signifiante à bien d'autres égards ? Comme la réponse à cette question ne saurait être donnée d'avance, **partons d'abord en exploration.** [LET066]

et c'est ce que suggère également l'usage fréquent du « on » dans ce type de balises (« comme on le verra », par ex. dans SCL007, « on peut ajouter à cette liste », LET066). On obtient ainsi des formes de coénonciation, par lesquelles l'auteur et le lecteur tendent énonciativement à ne faire qu'un en se trouvant impliqués conjointement dans la réalisation des procès :

Reste à regarder de plus près ce que signifie “ne peut pas se dire” : il s'agit d'un usage particulier du langage dont **il faudrait montrer** les conditions d'existence sociale. [SCL063-notes]

Si les balises méta-textuelles peuvent donc se limiter pour certaines à des marqueurs de structuration locationnels, elles intéressent aussi une description polyphonique du texte en se caractérisant souvent par une double implication – plus ou moins explicite – de la figure de l'auteur et du lecteur ; si elles peuvent en cela être considérées comme relevant des principes « reader-friendly » qui assurent le guidage du lecteur dans des textes plus didactiques, elles se présentent aussi comme des formes de routines scripturales, lexicalement figées (*on l'a dit, notons*, etc.), plus ou moins fonctionnelles et manifestement susceptibles d'être le fait de prédilections idiomatiques.

L'usage de ces balises méta-textuelles subsume la totalité ou presque des emplois de l'impératif, alors qu'il n'en va pas de même pour le futur dont deux autres emplois importants se manifestent dans le corpus. D'une part, le futur peut être employé avec une valeur prospective dans la formulation d'hypothèses³, mais le statut d'hypothèse n'est pas

¹ La polyphonie de ces formes d'impératif se manifeste dans l'extrait suivant ; il explicite la dimension argumentative de la proposition et renvoie de manière ambiguë au couple auteur-lecteur d'une part et à l'ensemble des français d'autre part : il s'associe à un *nous* à valeur générique, qui renvoie à un collectif cependant défini (en France) en impliquant ainsi le lecteur et l'auteur déjà mis en scène dans le procès de « reconnaissance » : « *Les cris d'animaux peuvent paraître peu nombreux, mais ils correspondent pourtant à ceux des principaux animaux (domestiques) rencontrés en France. En revanche, reconnaissons que nous n'imitons pas couramment le cri de la girafe, de la hyène, de l'éléphant, du singe, de l'autruche ou du chameau, alors que les onomatopées correspondantes existent dans les pays où vivent ces animaux* ». [SCL085]. Nous approfondirons la question de la polyphonie du *nous* au chapitre 7.

² Nous revenons sur ce type d'interrogations dans la partie consacrée aux ponctuations (5.2.).

³ Par exemple : « Dans le cas où on admettrait hypothétiquement, ici, un joncteur *et* « recatégorisé », cette supposition impliquera la fonction du tonème d'un côté, mais aussi, d'un autre côté, la part de prise d'initiative du locuteur dans la continuité ou dans la discontinuité dialogale. En effet, le joncteur *et*, n'opérant quant à lui qu'une rupture infime entre *p* et *q*, **inscrira** *q* davantage dans la continuité de ce qui se dit. » [SCL093]

nécessairement explicite et l'usage du futur sert alors des assertions qui se présentent comme tournées vers l'avenir, et qui présentent un caractère modal plus ou moins marqué de vérité générale, de conjecture, ou d'affirmation atténuée.

C'est pourquoi **on admettra mal** une co-occurrence des deux homophones, cacophonie mise à part, comme suit : « *Eh ! et ton père, comment va-t-il ?* », mais plutôt la co-présence d'une locution comme *eh bien* et de *et* dans une séquence cotextuelle restreinte. [SCL093]

Un dernier emploi assez important du futur se manifeste dans des séquences narratives, où sa valeur temporelle est mobilisée pour le repérage de la chronologie des événements :

Cette rêverie hydraulique préfigure le traité de Ramelli, *Le diverse et artificieuse machine*, **qui paraîtra à Paris en 1588**. (...) *Les forces mouvantes*, **comme dira plus tard Salomon de Caus**, constituent un concept opératoire aussi bien pour la physique que pour la métaphysique. "C'est ma métaphysique, c'est ma physique", **répéteront les Essais**. [LET004]

Par ailleurs, en 1987, j'avais suggéré à Michel Rocard d'exprimer son intérêt pour un aménagement modéré de l'orthographe dans son livre *Le coeur à l'ouvrage*. **Nommé Premier ministre en mai 1988, il appellera à son cabinet un linguiste (moi-même en l'occurrence), auquel il confiera à la fois la tutelle des organismes chargés de la langue française (...)** [SCL011]

Le futur est une caractéristique marquée de l'article au niveau de l'opposition SCL et LET, comme l'impératif : les articles de SCL, on l'a dit, y recourent significativement plus avec un taux moyen de 3,1% contre 2,2% dans les articles de LET. Par ailleurs, le futur est également le lieu d'une opposition au sein des articles de SCL entre ceux des doctorants et ceux des non-doctorants. Si dans le premier cas, les emplois communs de l'impératif et du futur peuvent expliquer la différence disciplinaire, la différence quant au statut de l'auteur est plus difficilement interprétable : ce sont les textes de SCL de non-doctorants qui se caractérisent par un usage abondant du futur par contraste avec les textes de doctorants avec un taux moyen de 2,4% dans les premiers et de 3,25% dans les seconds¹.

Or, nous avons jusqu'ici montré que les oppositions entre les textes de doctorants et de non-doctorants peuvent tenir à un recours plus net chez les doctorants aux principes dits « reader-friendly », qui assurent le guidage du lecteur dans le texte. Les autres emplois du futur expliquent donc peut-être mieux que les emplois métadiscursifs cette caractéristique positive des textes des non-doctorants de SCL par opposition à ceux des doctorants. Ce serait donc plutôt des futurs servant la prospection d'une part, et la mise en intrigue d'événements historiques d'autre part qui feraient la spécificité des textes des non-doctorants ; mais il reste possible aussi que les doctorants recourent moins à des balises méta-textuelles très « impliquantes » quant au marquage de soi et du lecteur, alors même que leurs textes mobilisent de nombreux autres procédés de guidage du lecteur.

Dans le même sens, l'examen des emplois des temps verbaux pour les deux groupes de LET et de SCL atteste d'une certaine unité des pratiques. S'il existe des tendances adverses structurant la diversité au sein du corpus, le seul fait que peu de temps verbaux soit corrélés montre qu'elles sont à l'œuvre de manière complexe dans chacun des groupes, et signale ce faisant aussi la complexité des variables linguistiques étudiées.

Nous résumerons pour finir ces tendances adverses sous la forme d'un schéma.

¹ La différence quant au statut de l'auteur donne, pour le futur, *Mann Whitney* $U=674,5$ et $p=0,028$.

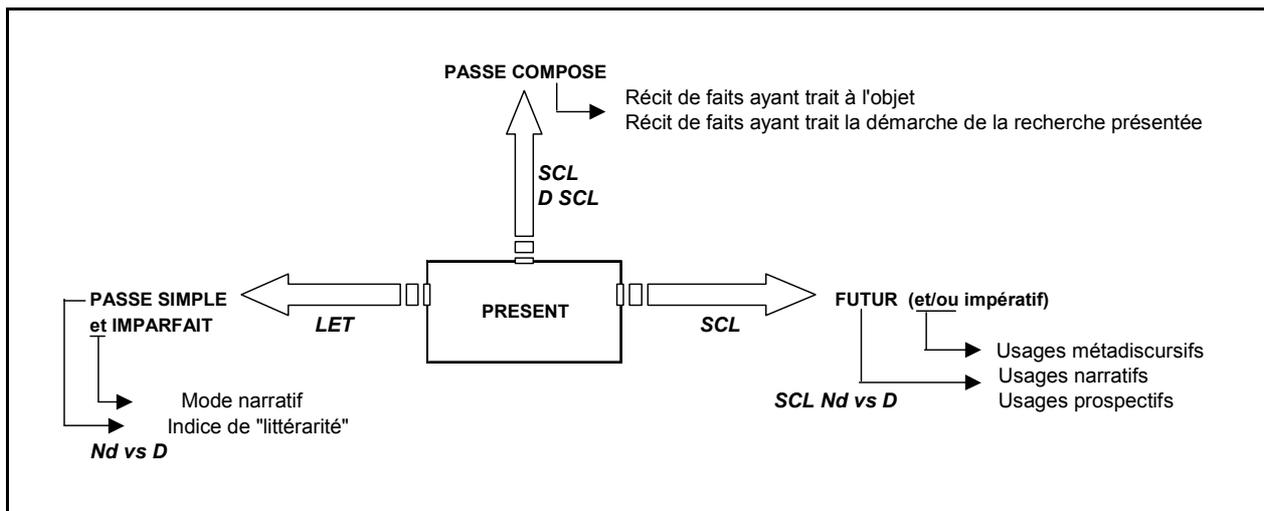


Figure 5.7. Des tendances contrastées dans l'usage des temps verbaux : bilan

Le système des temps verbaux en jeu dans l'article, pour confirmer la caractéristique attendue de la prédominance du présent de vérité générale, fait état du caractère discriminant de certains traits textuels (en tant que tendances adverses représentées dans le schéma au moyen des flèches pleines) ; l'usage du passé simple et de l'imparfait d'une part, du passé composé d'autre part et du futur et de l'impératif enfin se présentent comme des facteurs de différenciation des textes, qui jouent différemment selon les groupes (en italique sur le schéma) et en fait leur spécificité.

Il convient toutefois d'insister sur l'irréductible diversité des textes tout autant que sur le caractère partagé entre les groupes de nombre d'emplois marqués (le mode narratif au passé simple ou au passé composé, l'exposé au passé composé des faits relatifs à la démarche de la recherche présentée, les balises méta-textuelles au futur et à l'impératif, etc.) ; ainsi se manifeste une forme d'unité dans les possibilités diverses offertes par le genre, notamment en termes de séquences textuelles (Adam, 1999) signalées par ces emplois verbaux, et au niveau de leur fonctionnalité.

5.1.4. Les personnes linguistiques

Au terme de ce parcours des aspects morpho-syntaxiques du genre, et avant de nous intéresser à ses ponctuations et à son lexique, nous nous centrerons ici sur les personnes linguistiques, telles qu'elles se manifestent dans l'usage des pronoms personnels, des déterminants possessifs et des marques flexionnelles des verbes.

Il ne s'agit ici que de fournir un premier aperçu quantitatif sur la distribution dans le corpus de ces marques, qui intéressent cependant de près une approche énonciative du genre. Aussi reviendrons-nous sur leurs emplois dans le chapitre consacré à la subjectivité et la polyphonie du genre (chapitre 7) : il conviendra en effet de s'intéresser de plus près au fonctionnement de la prise en charge énonciative dans l'article, et les résultats indiqués ici nous serviront donc d'entrée à compléter par une analyse des occurrences en contexte menée sur un corpus plus restreint.

L'analyse est menée sur les textes une fois soustraits les exemples et les citations, en raison du fort risque de bruit induit par les spécificités linguistiques de ces composantes (cf. chapitre 4 pour un explicatif sur les risques de bruit liés aux exemples et aux citations). Précisons les 3 groupes de variables conjointement étudiées :

- les pronoms personnels sujets, et le poids des pronoms de 1^{ère} et 2^{ème} personne, au singulier et au pluriel : *je*¹, *tu*, *nous*, *vous* ;
- les déterminants possessifs, et le poids de chacune des catégories de 1^{ère}, 2^{ème} et non-personne au singulier et au pluriel : *mon/ma/mes*, *ton/ta/tes*, *son/sa/ses*, *notre/nos*, *votre/vos*, *leur/leurs* ;
- les personnes grammaticales, telles qu'elles sont marquées dans la morphologie verbale et le poids de chacune des catégories des verbes à la 1^{ère}, 2^{ème} et 3^{ème} (ou non-) personne au singulier et au pluriel.

Ce faisant, nous laissons de côté les pronoms possessifs (*le mien*, etc.), car cette variable est extrêmement rare et son caractère marginal se présente comme un des traits figés du genre.

Deux axes prioritaires d'analyse vont guider ce premier aperçu : d'une part, il s'agit de voir dans quelle mesure le genre se conforme aux caractéristiques attendues du discours scientifique au niveau des marques de personne ; d'autre part, il s'agit de mettre en évidence des caractéristiques marquées de l'usage des personnes par la comparaison entre les deux disciplines et en fonction du statut de l'auteur.

Le premier constat que l'on peut faire n'a rien de surprenant ; au contraire, il se situe dans la conformité des résultats avec les caractéristiques supposées du discours scientifique : la troisième personne grammaticale, c'est-à-dire, d'un point de vue énonciatif, la non-personne selon les propositions de Benveniste (1966) domine largement comme on le voit au niveau des verbes (cf. le graphique *infra*).

C'est plus exactement le groupe sujet à la troisième personne du singulier qui se présente comme une caractéristique syntaxique centrale du genre, avec un taux moyen de 77% de verbes concernés parmi l'ensemble des verbes conjugués. Viennent ensuite les verbes à la troisième personne du pluriel qui représentent en moyenne dans l'ensemble du corpus 19,2% des verbes. Les marques au niveau du verbe de ce que E. Benveniste traite en termes de véritables personnes sont donc très faibles : 3% des verbes en moyenne sont à la première personne du pluriel, 0,9% à la première personne du singulier. La seconde personne se manifeste également, avec 0,7% de verbes pour le singulier, et 0,05% pour le pluriel.

Le graphique suivant rend compte des taux moyens de verbes pour chaque catégorie de personnes, dans l'ensemble du corpus (« Total ») mais également pour chaque groupe disciplinaire (LET/SCL) et en fonction du statut de l'auteur (Doctorants / Non-doctorants).

¹ Et la forme *j'*.

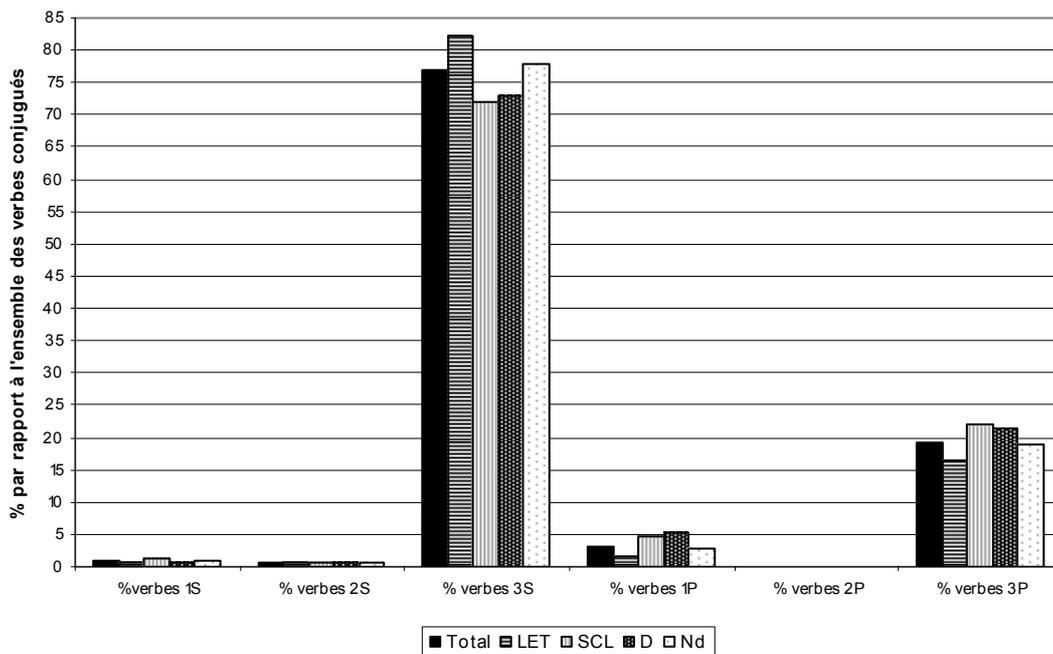


Figure 5.8. Taux moyen des personnes verbales : comparaison entre les groupes.

Les données sur les verbes représentées dans ce graphique appellent deux remarques. La première est que, de manière prévisible, les marques de personne énonciatives (i.e. les 1ères et 2èmes personnes grammaticales) sont marginales, mais qu'elles sont cependant attestées, y compris pour la seconde personne. Un tel constat oblige d'une part à dépasser l'image stéréotypée et non fondée d'un texte scientifique qui interdirait toute marque personnelle ; d'autre part, avec la présence étonnante de marques de 2^{ème} personne du singulier à hauteur de 0,7% des verbes, il signale le bruit produit par l'analyse automatique : même une fois les passages correspondant à des exemples et des citations ôtés, les textes en gardent certaines traces (cf. dans le chapitre 4, l'analyse des exemples et des citations).

La seconde remarque se base sur le constat d'une caractéristique centrale du genre dans l'usage de la 3^{ème} personne. Les proportions observées contredisent l'idée d'un texte absolument vide de toute marque personnelle mais se conforment aux caractéristiques attendues d'un usage central de la non-personne *versus* un usage marginal de marques de personne. Cependant le rapport entre d'une part la 1^{ère} et 2^{ème}, d'autre part la 3^{ème} personne est loin de constituer une caractéristique figée. Comme le suggère le graphique en ce qui concerne les seuls taux moyens, des différences de distribution se manifestent entre les groupes, qui se confirment avec les pronoms et dans une moindre mesure les déterminants possessifs, ces derniers étant de manière générale très marginaux dans le genre.

La distribution des emplois des personnels et des possessifs se présente en effet comme significativement distincte d'une part entre les groupes disciplinaires (LET/SCL) d'autre part entre les groupes fondés sur le statut de l'auteur (Doctorants/Non doctorants).

Les pronoms personnels de 1^{ère} personne sont particulièrement discriminants puisqu'ils marquent des spécificités dans l'ensemble des groupes : les textes de SCL recourent significativement plus que ceux de LET aux pronoms de 1^{ère} personne du singulier (*je*) et du pluriel (*nous*), et corrélativement au marquage verbal correspondant¹. Ainsi les

¹ La différence disciplinaire donne, concernant la première personne du singulier : pour le pronom *je*, *Anova*, $F=12,465$, $p=0,0005$, et pour les verbes à la 1^{ère} personne du singulier, *Anova*, $F=10,372$, $p=0,0015$; concernant la première personne du pluriel : pour le pronom *nous*, *Mann Whitney*, $U=3614$, $p<0,0001$, et pour les verbes à la 1^{ère} personne du pluriel, *Mann Whitney*, $U=2972,5$, $p<0,0001$.

verbes à la première personne du singulier représentent-ils en moyenne 1,2% des verbes dans les articles de SCL contre 0,6% dans ceux de LET, et les verbes à la première personne du pluriel 4,6% des verbes dans les articles de SCL contre 1,6% dans ceux de LET.

Les textes de doctorants se caractérisent par un usage plus abondant de *nous* que ceux des non-doctorants, et même si rien ne permet cependant de dire que les non-doctorants préfèrent quant à eux le *je*, cette tendance se vérifie dans les formes verbales¹. Les textes des doctorants affichent une moyenne de 5,2% de verbes à la première personne du pluriel, ceux des non-doctorants une moyenne de 2,7%. Au sein de chaque discipline, cet usage plus abondant de *nous* dans les textes de doctorants est également marqué. Plus précisément, les articles de SCL produits par des doctorants comprennent significativement plus de verbes à la 1^{ère} personne du pluriel que ceux des non-doctorants² (7% des verbes en moyenne contre 3,9%); en LET, c'est l'usage du pronom *nous* qui distingue significativement les deux groupes d'auteur³, avec un taux moyen de 7,3% de *nous* dans les textes de doctorants contre 3,9% dans ceux des non-doctorants (par rapport à l'ensemble des pronoms personnels sujet).

C'est de ces différences dans l'usage de *je* et *nous* dont rend compte le graphique suivant, pour les seuls pronoms; chaque point identifie un article à l'aide de deux coordonnées : en abscisse son taux de *je*, et en ordonnées son taux de *nous* par rapport à l'ensemble des pronoms personnels sujets.

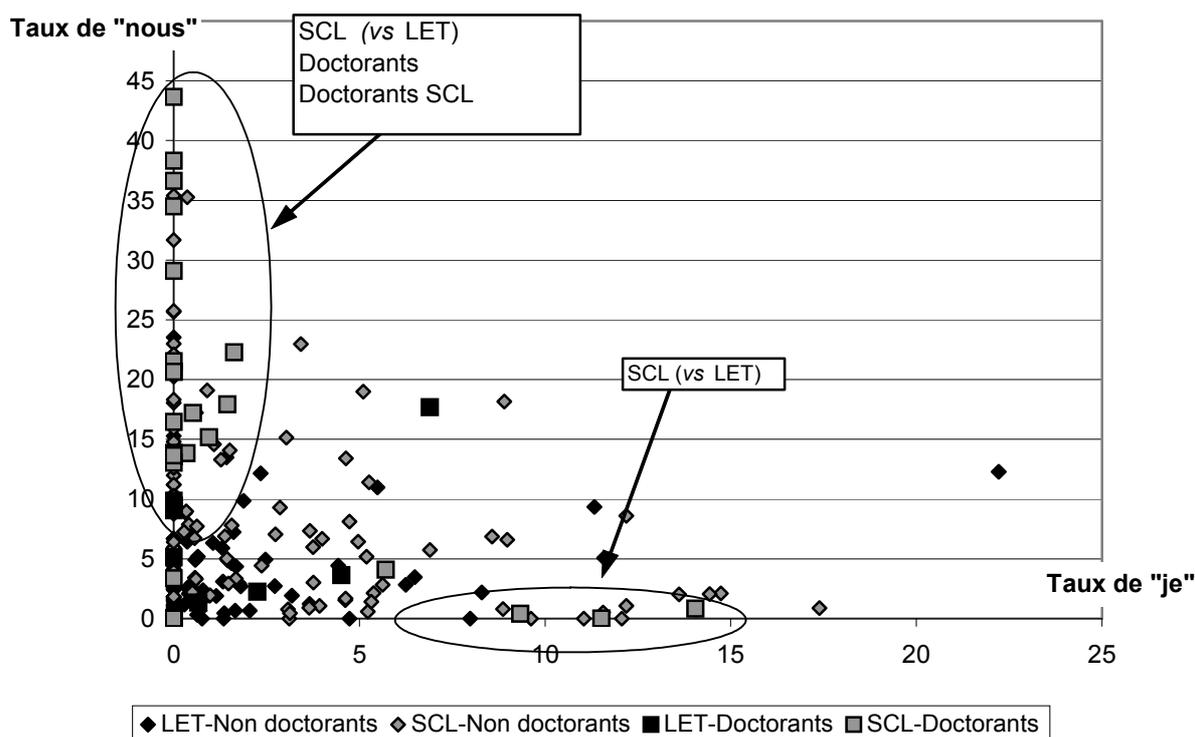


Figure 5.9. Le taux de je et de nous dans chaque article : comparaison entre les disciplines et les groupes d'auteurs.

¹ La différence liée au statut de l'auteur donne, concernant la première personne du pluriel : pour le pronom *nous*, *Anova*, $F=27,215$, $p<0,0001$ et pour les verbes à la première personne du pluriel, *Anova*, $F=16,018$, $p<0,0001$.

² En SCL, la différence liée au statut de l'auteur donne, pour les verbes à la 1^{ère} personne du pluriel, *Mann Whitney*, $U=621$, $p=0,0095$.

³ En LET, la différence liée au statut de l'auteur donne, pour le pronom *nous*, *Mann Whitney*, $U=405,5$, $p=0,037$.

Les zones encadrées sur le graphique signalent ces usages contrastés de la 1^{ère} personne en fonction de la discipline et/ou du statut de l'auteur. Globalement, la première personne du pluriel est plus importante que celle du singulier ce qui se conforme à l'idée qu'en contexte français, l'usage du *nous* de type académique est préféré à celui du *je* en tant que marque du scripteur dans son texte (Cf. par ex. Vassileva, 1998)¹. Il apparaît également qu'un certain nombre de textes ne contient aucune marque personnelle ou n'en contient que très peu, ce qui montre, en creux, l'importance prise par le mode délocuté.

Par ailleurs, on voit que certains textes se caractérisent par un recours mixte au *je* et au *nous*, ce qui est plus étonnant au regard des caractéristiques attendues du genre. On peut y voir un effet de bruit déjà signalé à propos de la présence de marques de 2^{nde} personne identifiées par l'analyseur mais ce constat appelle de toute évidence à spécifier les emplois de ces pronoms par une étude plus qualitative, que nous mènerons au chapitre 7. Comme l'indique le graphique, cet usage mixte serait toutefois davantage le fait des textes de non-doctorants que de ceux des doctorants, qui se situent plutôt le long de l'un ou l'autre des axes.

Le graphique signale enfin la présence de textes à valeur d'hapax tant leur taux de l'une, l'autre ou les deux marques est important ; c'est le cas par exemple du texte LET011, qui affiche un taux de 22,2% de *je* et de 12,3% de *nous* par rapport à l'ensemble des pronoms personnels sujets. L'examen de ce texte suggère une part de bruit dû à de courtes citations intégrées au texte, mais s'explique aussi par la présence faible de sujets pronominaux, et par un usage du *je* que l'on pourra qualifier de *je* philosophique (cf. chapitre 7), attestant une fois encore des limites d'une seule étude quantitative des marques dites personnelles et de la nécessité d'un examen qualitatif des occurrences.

Si l'usage plus faible de marques personnelles (*je*, *nous* ou les deux) dénote l'importance prise par la troisième personne, une différence peu prédictible se manifeste entre le singulier et le pluriel : les articles de LET affichent en effet un usage marqué de la troisième personne du singulier au niveau des verbes, et cette tendance affecte également les déterminants possessifs (*son*, *sa*, *ses*). Inversement, les articles de SCL ont pour spécificité un usage plus important de la troisième personne du pluriel².

Cette spécificité liée au nombre des syntagmes en fonction sujet est assez étonnante, mais se retrouve aussi au niveau de la différence entre doctorants et non-doctorants : les non-doctorants recourent significativement plus que les doctorants aux verbes à la 3^{ème} personne du singulier, et les doctorants plus à la 3^{ème} personne du pluriel, les deux disciplines confondues, et également au sein des SCL³.

L'opposition entre 3^{ème} personne du singulier et du pluriel se vérifie dans une corrélation négative forte entre les deux séries de données⁴. Cependant ces résultats n'offrent que peu de prise sur les textes et ne permettent pas de dire si l'usage plus abondant de la 3^{ème} personne du singulier tient à celui du *on*, du *il* impersonnel, d'infinitives ou encore de substantifs en fonction sujet, et de spécifier ainsi les implications en termes de prise en charge énonciative, et, secondairement aussi, en termes de choix lexicaux et de thématiques.

¹ Sauf à postuler qu'il renvoie à un « nous » d'équipe, ce qui est certes tout à fait possible et effectivement attesté, mais reste très marginal dans la mesure où les textes retenus sont signés par un seul auteur.

²² La différence disciplinaire donne, pour les verbes à la 3^{ème} personne du singulier, *Anova*, $F=137,544$, $p<0,0001$, pour les déterminants possessifs à la 3^{ème} personne du singulier, *Anova*, $F=6,647$, $p=0,0106$ et pour les verbes à la 3^{ème} personne du pluriel, *Anova*, $F=46,731$, $p<0,0001$.

³ La différence liée au statut de l'auteur donne, pour les verbes à la 3^{ème} personne du singulier, *Anova*, $F=11,98$, $p=0,0006$ les deux disciplines confondues, et *Anova*, $F=11,264$, $p=0,0011$ dans le corpus de SCL ; pour les verbes à la 3^{ème} personne du pluriel, *Anova*, $F=4,18$, $p=0,0421$ les deux disciplines confondues, et *Anova*, $F=5,265$, $p=0,0237$ dans le corpus de SCL.

⁴ Corrélation négative de 0,823 dans l'ensemble du corpus (test réalisé avec Statview, non consigné en annexes).

En guise de bilan, commençons par résumer les tendances adverses déterminant au sein du corpus l'usage des marques personnelles et des marques de troisième ou non-personne. La non-personne caractérise les articles de LET par opposition à l'usage du *je* et du *nous* dans les textes de SCL, et les articles des non-doctorants par opposition à l'usage du *nous* dans ceux des doctorants, les deux disciplines confondues comme au sein des LET et des SCL. Enfin, une distinction plus difficilement interprétable se manifeste dans l'usage de la 3^{ème} personne du singulier caractéristique des articles de LET et des articles des non-doctorants par rapport à ceux de SCL, des doctorants et des doctorants de SCL.

Ce sondage quantitatif des marques de personne et de non-personne rend tangible l'intérêt d'une observation fine des occurrences en contexte. Une étude quantitative ne permet en rien à elle seule de conclure à une prise en charge plus personnelle en SCL qu'en LET, plus académique dans les textes des doctorants que dans ceux des non-doctorants etc., en raison du bruit produit par l'analyse automatique (cf. la présence étonnante de marques de seconde personne), et de l'hétérogénéité des usages du seul *je* : l'entrée formelle ne garantit pas que l'on ait affaire à une marque transparente du scripteur. Aussi la prise en charge énonciative requiert-elle – et c'est ce à quoi nous procéderons au chapitre 7 – une étude plus approfondie de ces marques, qui tiennent compte également de la spécificité du *on*, de la distinction entre un *il* anaphorique et un *il* non référentiel, dit impersonnel, et enfin de la complexité du jeu énonciatif par lequel la non-personne et les marques personnelles se combinent au service d'une prise en charge doublement objectivante et auctoriale.

Avec la partie suivante, consacrée après les aspects morpho-syntaxiques du texte à l'analyse de leurs ponctuations, nous verrons comment un examen en contexte de certains de ces signes peut fournir des pistes intéressantes à la caractérisation énonciative du genre de l'article en signalant des formes attendues ou plus surprenantes de manifestation de l'auteur et d'implication dans le texte de son lecteur.

5.2. Les ponctuations

Nous avons étudié la phrase en tant que composante du texte et unité syntaxique, à partir du nombre de mots et de propositions qu'elle comprend, et en nous intéressant à certaines de ses composantes morpho-syntaxiques. La caractérisation du genre au niveau méso-textuel pose aussi la question des ponctuations ; celles-ci intéressent d'une part la structuration de la phrase et du contenu propositionnel en vertu de leur fonction syntagmatique (Anis, 1988 :122), notamment pour les marques de pause (virgules, points-virgules), et permettent par ailleurs d'aborder certains aspects énonciatifs du texte, avec le rôle de ponctuations dans le signalement d'une hiérarchisation discursive (les parenthèses), les types d'assertion ou modalités énonciatives (assertion, interrogation, ou exclamation) et les ruptures dans le statut du signe et sa prise en charge modale introduites par les guillemets.

Nous commencerons par une analyse quantitative des signes de ponctuation dans leur ensemble, pour en venir à une observation plus fine de caractéristiques groupales marquées, et d'emplois rares ou fréquents des signes à valeur énonciative ; ces aspects nous semblent fournir une entrée intéressante à une analyse du genre qui montre en quoi ses traits sont ou non homologues à des caractéristiques attendues et qui envisage la fonctionnalité de ces signes au regard des enjeux de l'article.

5.2.1. Nombre de ponctuations dans le texte

Le nombre de ponctuations, en valeurs absolues, est en moyenne de 869,72 signes dans le corps de l'article. Il varie en LET entre 145 et 1770, avec une moyenne de 768,8, et en SCL entre 336 et 2928 avec une moyenne de 970,6. Il convient alors de pondérer ces écarts par un ratio sur le nombre de mots du texte. Les fréquences relatives ainsi obtenues font état d'une part d'une différence disciplinaire dans le nombre de ponctuations, d'autre part d'une différence liée au statut de l'auteur.

Les articles de SCL recourent significativement plus aux ponctuations¹, et manifestent une plus grande variabilité que ceux de LET de ce point de vue. C'est ce que suggère le graphique suivant, qui représente la distribution en fréquence du nombre de ponctuations (en fréquence relative) dans les deux disciplines :

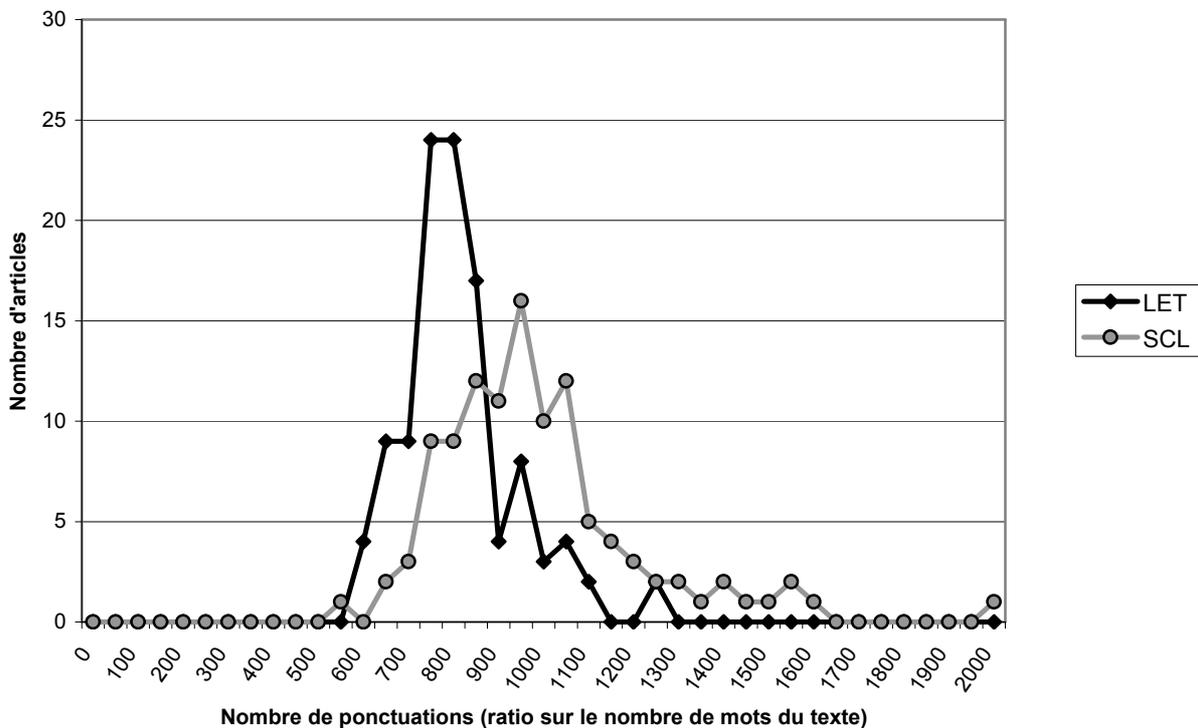


Figure 5.10. Distribution en fréquence du nombre de ponctuations par article (fréquence relative) : comparaison disciplinaire.

Par ailleurs, la pratique de la ponctuation caractérise également les textes des doctorants par rapport à ceux des autres scripteurs, les deux disciplines confondues, comme le montre, dans la figure suivante, la comparaison des distributions dans les deux groupes de scripteurs.

¹ Anova, $F=49,465$, $p<0,0001$.

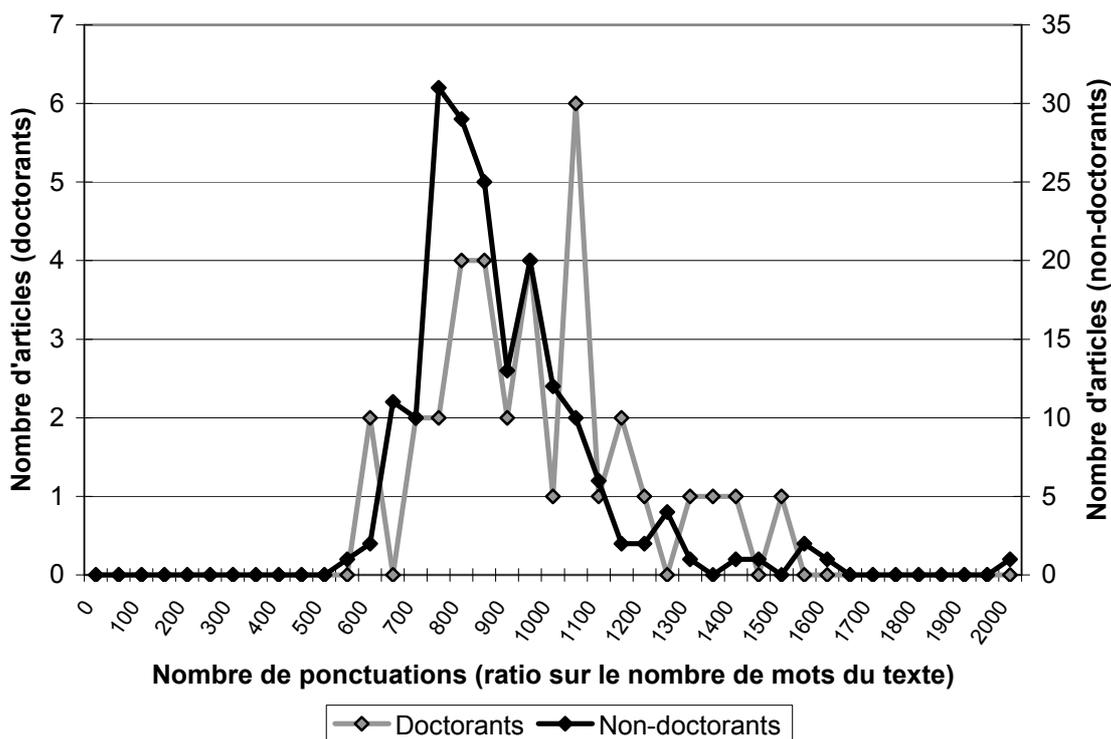


Figure 5.11. Distribution en fréquence du nombre de ponctuations par article (fréquence relative) : comparaison des deux groupes d'auteurs.

Les doctorants recourent significativement plus aux ponctuations que les autres scripteurs¹, quoique les tendances centrales et les indices de dispersion restent plus proches entre ces deux groupes qu'entre les deux disciplines.

On obtient donc un double partage dans le corpus en fonction de cet usage des ponctuations, qui distingue une pratique plus abondante en SCL qu'en LET, et chez les doctorants par rapport aux non-doctorants. Ce constat reste en lui-même difficilement interprétable. Il fait écho à ce que nous avons dit du découpage du texte en phrases en corroborant l'idée d'un découpage plus important dans les articles de SCL et dans ceux des doctorants ; cela étant, l'hétérogénéité des signes retenus comme signes de ponctuations oblige de toute évidence à traiter des différences entre groupes sous l'angle des types de ponctuation plutôt que de leur seul nombre.

5.2.2. Aperçu général du type de ponctuations en usage

Pour étudier les types de ponctuation dans le corps du texte, nous avons utilisé le corpus sans exemples ni citations, afin d'éviter de parasiter l'analyse par des caractéristiques qui seraient propres aux exemples en particulier.

Les types de ponctuations retenus par l'analyseur automatique sont les suivants : virgules, points-virgules, points, points de suspension, points d'interrogation, points d'exclamation, deux-points, parenthèses, tirets, crochets et accolades. Les guillemets ne sont pas pris en compte, et nous proposons pour clore l'analyse des ponctuations de mener une étude spécifique de ce signe dans un sous-corpus de 20 articles par discipline.

¹ Anova, $F= 4,014$, $p= 0,0464$.

Le poids des différentes ponctuations identifiées par l'analyseur est évalué sous la forme d'un pourcentage par rapport à l'ensemble des ponctuations. Cela présente l'avantage de neutraliser l'effet du nombre de mots du texte mais envisager ces signes comme système reste discutable, dans la mesure où leurs fonctions sont extrêmement hétérogènes ; ainsi, si l'on peut convenir que les points, les points d'exclamation et les points d'interrogation représentent des alternatives qui s'opposent entre elles¹, il n'en va pas de même pour les points et les parenthèses pour ne citer qu'un exemple.

En moyenne, le poids des différentes ponctuations se présente de la manière suivante dans l'ensemble du corpus (« Total »), dans les groupes disciplinaires (LET/SCL) et en fonction du statut de l'auteur (D/Nd) :

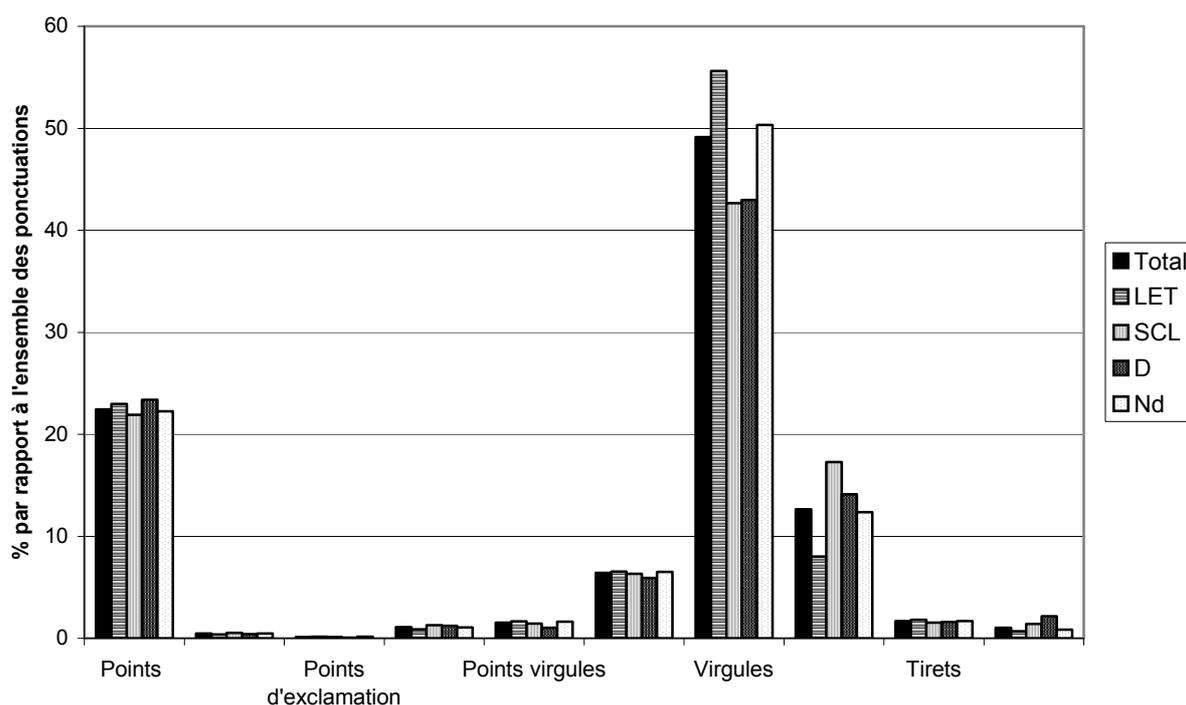


Figure 5.12. Taux moyen de chaque type de ponctuation : comparaison entre les groupes.

Le signe de ponctuation le plus représenté dans l'article est la virgule, avec un taux moyen de 49,1% par rapport à l'ensemble des ponctuations du corpus. Elle est suivie du point qui atteint un taux moyen de 22,4% et des parenthèses avec 12,6%. Viennent enfin les deux-points avec un taux moyen de 6,4%.

Certains signes de ponctuations se présentent au contraire comme des variables extrêmement marginales de l'article. Ainsi en est-il des tirets (1,7% en moyenne), des points-virgules (1,5%), des points d'interrogation (1,1%), des crochets et accolades (1%), des points de suspension (0,5%) et des points d'exclamation (0,1%).

Ces résultats ne sont pas surprenants, hormis peut-être pour la place prise par les parenthèses, l'importance quantitative de ce signe de ponctuation étant moins intuitive. Si l'usage des points et des virgules est élevé dans nombre de genres (Anis, 1988 : 125) le poids des parenthèses est davantage spécifique semble-t-il à celui de l'article ou du discours scientifique en général ; c'est ce à quoi conclut C. Poudat (2003) dans son étude de la

¹ Mais là aussi, les contraintes exercées par le genre sont de taille ; il n'est pas impossible en effet d'envisager dans certains genres la répétition d'un même signe (!!!), ou l'utilisation d'un signe double (?!), qui suggèrent que si système d'oppositions il y a, c'est au sein de genres qu'il demanderait à être étudié.

variation inter-générique entre les essais, les nouvelles, les textes de lois et les articles scientifiques : la parenthèse est une des différences significatives entre les quatre genres, qui caractérise positivement le genre de l'article scientifique comparativement aux autres en raison de sa forte présence. Cette première approche suggère l'intérêt d'une étude des genres pour l'analyse de la langue : c'est à travers les genres qu'on pourrait voir comment fonctionne le système linguistique des ponctuations de manière commune ou spécifique aux genres considérés.

Le nombre de ponctuations atteste d'un usage significativement plus important dans les articles de SCL par rapport à ceux de LET d'une part, et dans les articles des doctorants par rapport à ceux des non-doctorants d'autre part. Avec les types de ponctuations, le premier constat que l'on peut faire est que tous les types de ponctuations ne sont pas soumis à l'effet de la discipline et du statut de l'auteur, ce qui indique une certaine stabilité du genre à ce niveau. Par ailleurs, prises telles quelles, les différences quantitatives ne sont pas toujours aisément interprétables, de même qu'avec le nombre global de ponctuations.

Nous allons à présent approfondir l'analyse des ponctuations en pointant quelques emplois centraux, qui nous semblent indiquer l'importance de ces signes dans le genre, à la fois comme entrée pour caractériser ce dernier, et parce qu'ils suggèrent l'incidence que le genre peut avoir en retour sur l'étude de la langue. Nous nous intéresserons d'abord aux emplois des ponctuations relativement fréquentes, qui font état de différences inter-groupales, en fonction de la discipline et/ou du statut de l'auteur (les virgules et points-virgules, les deux-points, les parenthèses), pour mener ensuite une approche plus qualitative sur des ponctuations énonciatives, qui caractérisent le genre par leur rareté (points de suspension, points d'exclamation) ou leur usage modéré (points d'interrogation) d'une part, et sur les usages des guillemets d'autre part.

5.2.3. Des tendances contrastées dans l'usage des ponctuations, liées à la structure des textes et à des préférences idiomatiques.

Marqueurs de pause : virgule et points-virgules

Parmi les ponctuations pour lesquelles intervient un effet de la discipline et/ou du statut de l'auteur, les marques de pause sont les premières concernées, et notamment les virgules et les points-virgules. Les premières s'avèrent davantage employées en LET qu'en SCL, chez les non-doctorants que chez les doctorants, et chez les non-doctorants de SCL par rapport aux doctorants de SCL ; l'usage de la virgule peut ainsi s'expliquer comme une alternative au découpage du texte en unités phrastiques, qui caractérise, à l'inverse, les textes de SCL et des doctorants.

La virgule irait donc de pair avec des phrases et des syntagmes plus longs, comme le suggère l'extrait suivant, tiré de l'article LET087 où l'on note un taux maximal de 75,9% de virgules :

Les formes d'analyse déployées jusqu'ici, axées principalement sur une analytique des genres et sur des lectures synoptiques de journaux, ont notamment permis de mettre en cause les frontières établies rétrospectivement par la canonisation et l'institutionnalisation du discours savant entre la littérature « vraie », celle qui paraît dans les livres, et la littérature périodique, réputée vulgaire, éphémère et impure, à la fois par la constatation de la littérature généralisée de la presse, même celle dont la diffusion était la plus large et *a priori* la plus « journalistique », et de la porosité des frontières, qu'elles fussent

graphiques, stylistiques ou idéologiques, séparant les discours réputés informatifs et les discours réputés poétiques et surtout fictionnels. [LET087]

L'énumération, l'emploi de subordonnées, les connecteurs, les appositions et les incises, qui augmentent la taille des syntagmes et des phrases constituent les lieux privilégiés d'utilisation de la virgule.

Le point-virgule est également soumis à des différences importantes et plus particulièrement en fonction du statut de l'auteur. Comme la virgule, il est davantage prisé par les non-doctorants que par les doctorants, et s'avère donc là encore caractéristique d'une tendance inverse à celle du découpage d'unités phrastiques au moyen des points. Cependant, un examen des textes permet aussi de voir que le point-virgule représente une alternative à d'autres signes dans la notation du renvoi aux sources, comme le montre l'extrait suivant, tiré d'un article de non-doctorant de LET qui se caractérise par un usage maximum de plus de 9% de points-virgules :

Du reste, le rejet de la femme n'équivaut point pour eux au rejet de l'amour: «*ma haine aime*» (*ThL*; III; 436), lance fièrement Slagistri, tandis que Welf vante «*l'amour qu'a pour [lui] l'ombre*» (*LS2*, VIII; III; 317). [LET110]

Cet usage spécifique du point-virgule montre que le renvoi aux sources peut passer par des systèmes de notation largement idiomatiques.

Les deux-points

C'est aussi du côté de ces systèmes de notation du renvoi aux sources que l'on peut expliquer les différences constatées pour les deux-points, significativement plus utilisés dans les textes des non-doctorants de SCL que dans ceux des doctorants. Cependant, les deux-points attestent également d'un recours important à l'exemplification, comme c'est le cas dans l'article SCL110 qui en compte plus de 13% :

C'est sur une relation prédicative présupposée que porte ici l'adverbe. Le présupposé peut être reconstruit sous la forme : *Syntax-wise and tool-wise, something is planned.*

ou dans SCL021 qui atteint les 15% de deux-points :

En règle générale, les écrivains montrent une nette préférence - nous y reviendrons un peu plus loin - pour le tour équivalent employant une complétive : *entendre la nouvelle que quelqu'un...*, ou plus simplement : *entendre (dire) que quelqu'un...*

Les deux-points servent ici le marquage de la rupture syntactico-énonciative introduite par l'autonyme, fonction plus souvent remplie semble-t-il par les crochets¹ qui caractérisent positivement quant à eux les textes de SCL produits par des doctorants. Plus largement, cette fonction des deux-points semble pouvoir être rattachée à une autre fonction nettement représentée dans le corpus, à savoir le marquage d'un rapport logique entre deux

¹ Comme dans « Dans l'exemple (1) l'énoncé [" compatriote "] conserve sa SL habituelle mais à celle-ci s'ajoute une SI proche [Français de souche], cette inférence est validée par les points de suspension. L'exemple (2) avec l'énoncé [*sympathisé*] voit à sa SL s'ajouter la SI [entretenir des liens que la morale réprouve] ou [être la dupe de] » [SCL026].

propositions : l'accent est mis sur l'existence d'une relation sémantique entre les composantes plutôt que sur leur distinction au moyen d'une pause (comme c'est le cas avec la virgule), mais cette relation généralement explicative n'est pas davantage spécifiée.

Les parenthèses

Les parenthèses manifestent également des différences mais disciplinaires cette fois. Elle sont davantage attestées en SCL qu'en LET et un taux fort de parenthèses renvoie là aussi prioritairement semble-t-il à la pratique de l'exemplification, et à la numérotation dont les exemples font l'objet dans les articles de SCL contrairement à ceux de LET, comme nous l'avons dit en traitant des composantes structurelles du genre ; on le voit dans le premier extrait de l'article SCL108 dont les ponctuations sont à plus de 37% des parenthèses et dans le second, de SCL058, qui en compte près de 40% :

Un adjectif après le SN « un Wallon » se mettra moins naturellement au pluriel qu'au singulier, à moins que l'on ne sépare le SN introduit par *dont* du SN antécédent par une virgule, comme en (3d) et (3e) [SCL108]

Comme le suggère la représentation (1a), le présent article se propose de montrer, sur la base des langues Gbe, que l'interprétation spécifique découle de propriétés syntaxiques qui sont localisées dans la périphérie gauche nominale (PGN). Dans la section 2, nous montrons que les langues Gbe apportent la preuve d'une PGN éclatée comparable à la périphérie gauche de la proposition (Rizzi 1997). En d'autres termes, la PGN constitue un système articulé où les traits de définitude, de spécificité et de nombre, qui sont couramment associés à D, sont les propriétés de projections fonctionnelles distinctes (Knittel 1998; Aboh 1999, à paraître; Ihsane & Puskás 2001). Selon cette analyse, l'interprétation spécifique résulte du mouvement de la projection étendue du nom (*c'est-à-dire le système flexionnel nominal étiqueté ici EP*) dans SpecDP. [SCL058]

Le second extrait montre que les parenthèses ont évidemment partie liée aussi à la référencement des sources. Leur emploi plus important en SCL s'expliquerait ainsi par les stratégies de renvoi aux sources mises en évidence dans la macro-analyse de la structure des textes : dans le cadre d'une stratégie consistant à placer un indicateur de référence dans le corps du texte, les parenthèses rendent possible de telles insertions, que ne requiert pas une stratégie du type appel de notes et référence exhaustive en notes. Par ailleurs, le nombre de références doit jouer également sur le poids des parenthèses.

En vertu de leur rôle d'insertion de commentaires, les parenthèses servent à introduire des précisions tout en signalant ce statut spécifique en tant que tel. Ce rôle de hiérarchisation des contenus participe ainsi des formes de guidage du lecteur, ce qui fournit une autre explication de la différence disciplinaire dans leur usage, compte tenu de ce que nous avons dit des indices de structuration et des principes « reader-friendly » dans la macro-analyse du genre dans les deux disciplines.

Ce rôle d'explicitation, signalé comme tel à l'égard du lecteur, se manifeste dans l'énumération des 4 propriétés de l'unité lexicale dans cet extrait d'un article de SCL :

Les lexicologues n'ont pas unifié leur conception de l'UL. Sans entrer dans le débat nous rappellerons brièvement que celle-ci se caractérise par au moins quatre types de propriétés :

- phonologiques (elle est également justiciable d'une description graphématique) ;
- sémantiques (qui reflètent ses conditions d'applicabilité référentielle et d'usage pragmatique, ses connexions avec d'autres UL) ;
- syntaxiques (catégorisation grammaticale, sous-catégorisation contextuelle) ;
- morphologiques (unité simple *vs* construite). [SCL102]

Dans le même sens, on retrouve les parenthèses pour expliciter les abréviations utilisées,

« S'il est vrai que les traits topique (TopP), interrogatif (InterP) et spécifique (SpfP) sont une propriété de la périphérie gauche de la proposition » [SCL058]

ou plus largement à reformuler un syntagme ou une proposition comme on en voit ici un exemple :

Dans cette perspective deux questions centrales seront abordées : le couplage entre dénomination et grammaire (la dénomination est-elle une propriété du seul substantif ?) et l'hétérogénéité de la relation de dénomination (existe-t-il une ou plusieurs procédures de dénomination ?) [SCL102].

En cela, les parenthèses se présentent comme une entrée intéressante pour aborder la question de la conceptualisation à l'œuvre dans l'article, et les procédés de l'explicitation, réputée caractéristique du discours scientifique ; mais leur étude demande à tenir compte du fait qu'en dehors de quelques usages codifiés (le renvoi aux sources, l'explicitation des abréviations), l'exploitation plus ou moins importante de leur fonction dans la reformulation ou dans l'apport de commentaires relève semble-t-il d'usages largement idiomatiques, propres à certains auteurs plutôt que d'autres.

Les spécificités observées statistiquement dans l'usage des ponctuations mettent en évidence l'exploitation de leur rôle structurant (virgule, point-virgule), particulièrement dans le cas de phrases longues, ce qui expliquerait leur présence plus forte en LET et chez les non-doctorants des deux disciplines. Elles intègrent ainsi des caractéristiques structurelles du texte, comme le montre leur rôle dans l'exemplification et la numérotation des exemples, dans le renvoi aux sources, et dans les procédés d'explicitation et de guidage du lecteur. Cependant, l'exploitation des ponctuations se présente aussi comme une pratique idiomatique qui manifeste des choix stylistiques résultant de diverses alternatives (par exemple, dans l'insertion d'autonymes, ou dans les séparateurs d'indicateurs de référence).

L'intérêt d'une étude des signes de ponctuation et de leurs fonctions spécifiques au sein de genres donnés se manifeste à plusieurs niveaux : d'abord, l'interprétation des différences inter-groupales observées quantitativement se heurte à l'extrême hétérogénéité des fonctions qu'un même signe peut endosser. Elle appelle à aller voir de plus près quels usages sont faits de ces signes, pour les mettre en perspective avec d'autres aspects du texte.

D'autre part, comme ce sont plutôt des ponctuations plus ou moins rares (le point d'exclamation, le point d'interrogation, le point de suspension, le tiret) qui ne font pas l'objet d'usages différenciés, il est permis de penser que la stabilité du genre peut tenir à l'exclusion tendancielle de certaines marques, ou à un usage modéré, restreint à certaines fonctions. Si, d'un point de vue quantitatif, la rareté de ces signes se conforme donc à des caractéristiques attendues du genre, leur présence, même marginale, pourrait ainsi déplacer l'idée que l'on s'en fait. Il importe pour le moins qu'on s'intéresse aux emplois de ces signes, pour voir en quoi ils se conforment ou non aux caractéristiques attendues du discours scientifique et pour analyser leur fonctionnalité par rapport aux enjeux de l'article.

Une étude au sein des genres doit permettre de mieux cerner l'éventail des fonctions effectives, et de pointer celles qui prévalent. C'est ce que nous proposons d'approfondir maintenant, en abordant d'une part quelques emplois de ponctuations très rares et cependant attestées, les points de suspension et les points d'exclamation, d'autre part l'usage attendu mais riche du point d'interrogation, et enfin celui d'un signe délaissé par l'analyse automatique, mais important d'un point de vue quantitatif comme dans le rôle qu'il joue dans le texte, les guillemets.

5.2.4. Les points de suspension et le point d'exclamation : des emplois rares... et peut-être insoupçonnés !

Parmi les ponctuations rares qui, parce qu'elles sont cependant attestées, méritent qu'on s'intéresse à leurs usages, en particulier en termes de conformité aux caractéristiques supposées du discours scientifique, nous traiterons d'abord des points de suspension, puis des points d'exclamation. Nous laissons de côté les tirets, l'examen de leurs occurrences ne faisant que confirmer l'idée qu'ils entrent dans des « structures-listes », déjà bien décrites par J. Anis (1988) et dont nous avons montré l'intérêt dans la caractérisation du genre de l'article au niveau de la macro-analyse de la structuration des textes dans les deux disciplines.

Les points de suspension

Les points de suspension sont au nombre de 171 en SCL, et de 155 en LET. En SCL, ils ne sont le fait que de 38 articles (soit un peu plus d'un tiers du corpus), contre 49 en LET¹. Leurs fonctions principales confirment des caractéristiques supposées du genre : ils servent prioritairement de marque de coupure textuelle dans les citations, sous la forme (...) ou [...], et de signalement d'une fin d'énumération avec suggestion de prolongation, selon un usage concurrencé par l'abréviation « *etc* » ou « *etc.* ». Cette abréviation est attestée 105 fois au total dans le corpus de LET, et 362 fois en SCL, ce qui suggère avec les points de suspension que la pratique énumérative tendrait à être plus importante en SCL. On peut vraisemblablement y voir en partie un effet du type d'exemples, les articles de SCL affectionnant l'usage de liste de termes, comme dans cet extrait :

« Ceci est un vieux débat chez Larousse qui a toujours eu pour tradition de mettre le tréma sur le e durant ces vingt dernières années pour les adjectifs féminins **aiguë**, **suraiguë**, **contiguë**, **exiguë**... excepté dans la période allant de 1989 à 1993 où l'on avait admis les deux formes » (SCL017).

A la manière du rôle joué par les marques (...) ou [...] dans les citations, l'interruption de la chaîne graphique indiquée par les points de suspension est ainsi fréquemment exploitée en SCL pour servir la représentation linguistique, comme on le voit dans cet extrait :

Or, si cette stratégie peut être poursuivie par les formes indirectes, elle peut également l'être par l'ajout de *s'il vous plaît*, par le recours aux performatifs du type *je vous invite à...*, *je vous prie de...*, à des accompagnateurs qui atténuent l'impact de la requête (SCL066)

ou dans des modes de représentation linguistique qui associent les abréviations et des points de suspension, comme c'est le cas en SCL079 où les points de suspension indiquent la durée dans la production syllabique :

VVO...VVT (ms) : Vocalic Voiced Onset and Vocalic Voiced Termination
= début et fin vocalique voisée (SCL079)

Plus surprenant, un dernier type d'usage des points de suspension exploite le silence qu'ils introduisent et la suggestion de prolongation indépendamment de la pratique des listes, des citations et des représentations linguistiques. Les points de suspension correspondent à

¹ Mais la différence pour les signes à usage modéré n'est pas significative comme nous l'avons dit (voir en annexes 3. pour le détail des tests).

une ellipse, et ne disent rien d'autre que cette part de suggestivité relativement hermétique qu'ils introduisent :

Mon opinion (non ma doctrine...) est que ce livre doit quelque chose à cet espace et à la façon dont Montaigne, par intervalles, l'habitait. [LET021]

Montaigne, d'ailleurs, le dit : la "tanière" (latebra) où il s'installe pour s'y "cacher" (mais ce poste est aussi le meilleur pour surveiller la maison...) n'était auparavant qu'une "garderobe", autrement dit un débarras, "lieu [le] plus inutile de [sa] maison" (III, 3) (LET021)

L'usage des parenthèses signale au lecteur que leur contenu est à comprendre comme une précision apportée par l'auteur ; avec les points de suspension, le commentaire apporté fait fonctionner une forme de connivence avec le lecteur. De nombreux exemples peuvent également être donnés en SCL et dans une discipline comme dans l'autre, cet usage des points de suspension contredit l'explicitation maximale réputée caractéristique du discours scientifique. Ce poids de l'implicite et le rapport spécifique au lecteur qui en résulte se manifestent de manière très nette dans l'extrait suivant :

Ainsi presque tous les composés sur casse- sont soudés, ex. cassepied, mais pas casse-pipe !, sur chasse-également, ex. chasseneige, mais pas chasse-marée ! **Sans commentaires...** (SCL012)

où l'expression lexicalisée « sans commentaires... » s'associe à des points d'exclamation dans un extrait qui se charge ainsi d'une dimension polémique à l'égard de l'objet, et rend explicite le rôle de clin d'oeil au lecteur joué par les points de suspension. La prise en charge d'un point de vue polémique à propos des réformes de l'orthographe est à mettre au compte de l'auteur mais implique dans le texte un lecteur co-énonciateur supposé confirmer cette analyse qui dit se passer de commentaires.

L'allusion et la connivence avec le lecteur introduites par les points de suspension peuvent par ailleurs être exploitées dans la création d'un effet d'attente¹. Avec les points d'interrogation, ils représentent une entrée intéressante dans l'analyse de formes de mise en intrigue², car ils permettent de créer un effet d'attente, exploité en particulier dans les titres (« *Des connaissances... pour quoi faire ?* » SCL037) mais également dans le corps du texte :

Montaigne a probablement dû embrasser la petite Basilia lors de son passage à Bâle. Car qui était l'heureuse mère de Basilia, Mme Pérez, née Lopez? Rien moins que la petite cousine de ... Montaigne! (LET004)

En s'associant aux marqueurs interrogatif et exclamatif, les points de suspension participent donc de procédés expressifs et polyphoniques qui peuvent paraître tout à fait étonnants dans le genre de l'article. On peut y voir l'effet de certaines pratiques idiomatiques dans la mesure où l'emploi des points de suspension est loin de concerner tous les articles, et *a fortiori* les emplois qui produisent un effet de connivence avec le lecteur en impliquant ce dernier en tant que « co-énonciateur » du contenu asserté (Rabatel, 2005).

¹ Par exemple dans « *Il faut souligner, je crois, cette alliance qui n'allait pas de soi entre le pouvoir politique, s'impliquant directement au niveau du chef du gouvernement, les linguistes réformateurs, qui acceptaient d'assumer des rôles de responsabilité, et l'Académie française, Maurice Druon soutenant d'emblée le programme du Premier ministre. Jamais une "réforme de l'orthographique" ne s'était préparée en France sous d'aussi favorables auspices...* » [SCL011].

² Nous y revenons en traitant *infra* des points d'interrogation.

Il n'en reste pas moins que si le genre se caractérise *a priori* par l'absence de ces signes et de tels emplois, il les exploite cependant, même modérément ; s'il peut s'agir de tenir des propos tendanciellement polémiques, l'article manifeste aussi une recherche stylistique de mise en intrigue, des formes d'allusion et de connivence avec le lecteur, voire, peut-être, une forme de complaisance dans l'écriture, qui relativisent fortement les caractéristiques canoniques attribuées à l'écriture scientifique, et sa centration supposée exclusive et transparente sur l'objet étudié.

Le point d'exclamation

L'étude des points d'exclamation, dont on a eu l'occasion de voir quelques occurrences en traitant des points de suspension aboutit à des conclusions similaires. Une fois les exemples et les citations ôtés du texte, on remarque qu'il reste 101 occurrences en SCL et 118 en LET, dans un nombre de textes restreint dans les deux cas quoique plus important en LET avec 38 articles concernés contre 28 en SCL. Quelques occurrences s'avèrent liées au matériau étudié dans l'article, et d'autres sont comprises dans des titres d'ouvrages cités.

Le dernier type d'usage montre la présence inattendue¹ dans l'article d'une modalité d'énonciation exclamative, bien entendu simulée (Anis, 1988 : 131) ; le scripteur feint de s'étonner de son propre énoncé (ibid.), mettant ainsi en scène une réaction attendue du lecteur. On retrouve les points d'exclamation dans des formes de mise en intrigue évoquées *supra* à propos des points de suspension dans un article de LET (« *Rien moins que la petite cousine de ... Montaigne!* » LET004), où l'effet d'attente se clôt par cette marque expressive du point d'exclamation.

Le point d'exclamation intervient de manière plus spécifique en LET dans une mise en scène singulière du discours de l'auteur étudié, comme on en voit ci-dessous deux extraits :

La lecture des Essais nous obligerait à répondre par l'affirmative. **Mais quoi! répondrait l'auteur de l'essai "Des Cannibales", les voyages ne sont-ils pas faits pour rêver?** [LET004]

On perçoit incidemment l'audace du propos: le péché originel, dont l'indiscipline du membre, chez Augustin, est la marque et le signe persistant, a cessé d'être ressenti par Montaigne comme une malédiction. C'est au contraire un motif de joie et une bonne occasion de célébrer l'amour de la vie. **Vive la sexualité, qui garantit la pérennité de l'espèce!** Et qu'importe si elle met en péril l'unité du corps et si la volonté n'a sur elle aucune prise. [LET009]

Une caractéristique relativement surprenante dans l'écriture de l'article se manifeste ici : l'auteur de l'article fait parler l'auteur qu'il étudie sous la forme d'une pseudo-citation dans LET004, et de discours indirect libre dans LET009 ; ces procédés autorisent des assertions expressives puisqu'elles sont le fait du texte étudié. L'interprétation du matériau se présente comme énoncée par l'auteur étudié, et se confond donc énonciativement avec le matériau lui-même. Ainsi les ponctuations à valeur énonciative peuvent-elles servir d'entrée à l'analyse du statut donné au matériau et au traitement qui en est fait dans les disciplines, en complétant ce qui peut être montré au niveau structurel comme nous l'avons proposé.

De tels exemples sont relativement fréquents en LET, mais n'épuisent pas les occurrences des points d'exclamation. Ainsi, en SCL comme en LET, nombre de phrases exclamatives interviennent sous la forme de commentaires polyphoniques, qui se présentent souvent dans le cadre de propositions en incise (entre tirets ou entre parenthèses), et

¹ Et possiblement niée ou proscrite dans certains manuels d'aide à la rédaction scientifique.

fonctionnent comme instruction à « une réaction déclenchée par la représentation d'une situation » (Ducrot, 1984 : 186) :

La parenté des deux œuvres fait ressortir l'orientation satirique de la glose — **et l'exemple n'est pas isolé !** — dans l'Ovide moralisé. [LET042]

Il peut sembler assez surprenant que les sommaires de cette revue ne portent en apparence aucun indice d'une quelconque préoccupation pour la traduction de poésie allemande. Pourtant, dans l'édition de janvier-février 1837, l'article sur les "Traditions d'Allemagne" de Xavier Marmier contient la traduction en prose de la "Lorelay" de Brentano ; en janvier-février 1839, aux "Poésies" de Sainte-Beuve s'ajoutent un sonnet "traduit d'Uhland" et un poème "imité d'Uhland" (**noter la nuance !**) [LET102]

Nous étudierons dans un premier temps les éléments qui sont utilisés par les intervenants pour forger une image d'eux-mêmes, **un masque (en latin persona !)** et dans un deuxième temps comment la personne se met en scène et s'autodécrit dans la conversation. [SCL036]

Le rapport au lecteur est ici clairement concerné. Le point d'exclamation montre dans ces exemples comment l'incise, tout en se présentant comme une procédure d'explicitation, fonctionne tout à la fois sur le mode de l'implicite et de l'allusif. La réaction présupposée du lecteur se présente sur le mode d'une connivence feinte, *via* la mise en scène dans le texte d'une interlocution : la figure du lecteur ainsi impliquée dans le texte est celle d'un lecteur prenant une part active à la construction du sens.

Le point d'exclamation peut aller jusqu'à servir une forme d'humour qui marque là aussi une connivence avec le lecteur. Ainsi en est-il dans cet extrait de SCL036 qui traite du choix des avatars dans les forums de discussion et intègre un clin d'œil exclamatif qui se présente comme un jeu avec les normes du genre :

Rien ne s'oppose donc à choisir une photo de soi, mais le choix le plus fréquent se porte sur une photographie d'une personne du même sexe correspondant aux canons de beauté en vigueur, photo souvent suggestive ou pire (**on n'en montrera pas ici !**). [SCL036]

C'est le cas aussi dans cet article qui étudie la manière dont les linguistes se réfèrent aux travaux des pairs dans leurs textes. La dernière phrase de conclusion consiste en une adresse explicite aux lecteurs et prend la forme d'une injonction qui justifie le point d'exclamation :

Le discours théorique et scientifique contrôle peut-être mieux l'usage de ses sources que le discours ordinaire (qui d'ailleurs n'a cure de références), mais la traçabilité des sources est loin d'y être toujours assurée. C'est qu'un discours, même à prétention scientifique, reste soumis aux lois du langage ordinaire : la polysémie et l'ambiguïté en sont constitutives. **Cette remarque ne doit pas servir d'alibi à la paresse. Linguistes, encore un effort pour améliorer l'indexation de vos sources !** [SCL034]

Enfin, l'humour et la connivence avec le lecteur peuvent également se déceler dans des emplois des points de ponctuation à visée polémique. On en a vu un exemple dans l'utilisation conjointe des points de suspension et des points d'exclamation dans l'expression « sans commentaires » dans un article sur les réformes de l'orthographe, et on retrouve cette distance critique dans d'autres textes de SCL comme de LET.

Il peut s'agir de polémiquer sur l'objet lui-même :

ou encore, à l'opposé, **le vers 13, qui, saturé d'entités abstraites, fait plus penser à Racine qu'à un vers romantique... !** On peut s'étonner de tels choix d'écriture chez un écrivain qui a lui-même encouragé les poètes français à aller revisiter les auteurs du XVII^e siècle afin de renouveler l'expression poétique nationale. Sans doute Sainte-Beuve n'a-t-il pas perçu lui-même la simplicité immédiate du poème de Uhland [LET102]

et certaines thématiques s'y prêteraient ainsi plus que d'autres : choix stylistiques du matériau étudié en LET, enjeux idéologiques des sujets traités dans les deux disciplines, comme les réformes de l'orthographe évoqués *supra* en SCL ou les rapports des concours de l'éducation nationale dans l'extrait suivant,

Chaque rapport s'efforce d'identifier (**puisque'il faut faire un plan, nous dit-on, avec des titres soulignés et des parties en petit nombre, de préférence trois, mais ce n'est pas une obligation, bien sûr !**) ce qu'on pourrait appeler des « figures organisatrices », des figures trans-locales qui regroupent un certain nombre de figures locales. Il en est ainsi, par exemple, dans Capes LM 94 (...) [SCL047].

où le point d'exclamation, en tant que caractéristique *a priori* marginale du discours scientifique, s'associe avec d'autres marques polyphoniques (*puisque, dit-on, bien sûr*) dans une incise qui sert l'intention polémique de l'auteur de l'article vis-à-vis ici d'un discours scolaire normatif.

La dimension polémique signalée, entre autres moyens linguistiques, par le point d'exclamation peut également affecter non plus seulement l'objet étudié mais les paradigmes théoriques. Cet usage des points d'exclamation n'est cependant attesté que dans le corpus de SCL et de manière rare. Donnons-en deux exemples :

Comme le montre encore Labov, le paradoxe darwinien tient en effet en ceci que le changement linguistique moteur de l'évolution des langues ne possède aucun des caractères d'adaptation et de fonctionnalité exigé par la sélection naturelle, mais se révèle au contraire absolument dysfonctionnel, anormal, destructeur et fondamentalement inadapté. **Or, sans principe de sélection naturelle et sans critère de meilleure adaptation écologique, il n'y a plus de darwinisme ! Ainsi le comparatisme généalogique se trouve-t-il contesté par le paradigme théorique même auquel il prétendait s'adosser.** Ne reste alors qu'un ensemble de thèses de sens commun appuyées par un jeu de métaphores incontrôlées. C'est, on l'a vu, la critique de ces métaphores qui conduit la linguistique naissante à s'interroger sur son objet et à proposer d'autres approches des interactions entre langues que la simple filiation. [SCL039]

MR3 ne représente qu'une première approximation de ce qu'il faudrait avoir. Afin de pouvoir dériver des groupes de verbes plus complexes, tels que ceux qui sont exprimés dans (5 c,d), la métarègle devrait être généralisée récursivement et formulée en termes de variables de types plutôt qu'en termes d'instances spécifiques *i* (*i*), *v* comme il a été fait dans les exemples (5 a, b). **Je reste cependant tout à fait confiant dans le fait que les tenants de l'approche en prégroupes, qui ont déjà montré une grande inventivité pour rendre compte de la complexité des séquences de clitiques en français et en italien [1, 4], se feront un plaisir d'élaborer une solution pour résoudre cette petite énigme !** [SCL042]

Le point d'exclamation, en s'associant avec d'autres marques d'emphase (*même, tout à fait*) et avec des axiologiques (*prétendait, confiant, inventivité, petite, etc.*), se présente ainsi comme une entrée intéressante pour étudier des aspects *a priori* tendanciellement prohibés de l'écriture scientifique, ou supposés rares en vertu d'un idéal d'objectivité dans la science, en particulier l'ironie et la polémique. Cette dernière, pour concerner certains types d'objets, et ce dans les deux disciplines, semble également affleurer en SCL à l'égard de certains paradigmes théoriques voire de leurs représentants. Le second extrait cité, qui combine polémique et ironie à l'égard des « tenants de l'approche en prégroupes » paraît à cet égard tout à fait surprenant ; comme il s'agit d'une traduction de l'anglais, il est possible de voir dans sa singularité l'effet de pratiques culturelles du genre de l'article, et peut-être aussi de la dimension culturelle de l'humour.

5.2.5. *Le point d'interrogation : question rhétorique ou rhétorique de la découverte ?*

Le point d'interrogation est une variable moins marginale que le point d'exclamation et le point de suspension. L'étude de ses occurrences dans les textes montre d'abord que ces derniers intègrent des caractéristiques de leur matériau même une fois les exemples et les citations ôtés.

Le second emploi observé, quantitativement primordial, concerne les interrogatives directes, totales (qui portent sur l'ensemble du contenu propositionnel) ou partielles (qui portent sur un des constituants de la phrase). Ces interrogatives, syntaxiquement directes et participant d'un point de vue énonciatif à la mise en scène de l'intersubjectivité dans les textes constitueraient, dans un usage modéré, un trait relativement prédictible du genre de l'article ; il est signalé dans divers manuels relatifs à la rédaction scientifique, où il est associé en particulier à l'exigence de la problématisation dans les introductions.

Ainsi, dans l'ouvrage didactique de C. Schnedecker sur les textes théoriques, les formes d'interrogation sont traitées comme un des moyens linguistiques privilégiés de la problématisation dans les introductions, avec le lexique signalant la construction du « problème » et de la « solution » (2002 : 150). Le caractère rhétorique de ces questions n'est pas thématisé, l'accent étant mis sur cette fonction des interrogatives, indifféremment directes ou indirectes.

Or pour J. Anis (1988 : 131), le caractère différé de l'écrit a pour conséquence que le point d'interrogation endosse des valeurs dérivées, au détriment de l'acte interrogatif. Deux cas de figure se présentent alors selon lui : les questions affectives dans la correspondance privée et la question rhétorique, qui camoufle une assertion ou annonce un développement – ce qui se prêterait à son rôle de présentation de la problématique dans les introductions.

Le problème de partage entre questions rhétoriques (dites aussi interrogations oratoires ou encore figurées) et interrogation repose comme le souligne A. Borillo (1981 : 2) sur leurs implications pragmatiques. Dans l'analyse indiquée *supra* qu'en fait J. Anis, ou encore dans celle de V. Dahlet (2002 : 100) qui reprend dans son étude de la valeur énonciative des ponctuations ce partage, l'interrogation se définit, en creux, comme une demande d'information qui n'est pas vérifiée dans la question rhétorique.

Les propositions de P. Fontanier (1821, ed. 1968 : 368) incluent cependant d'autres valeurs que cet acte de requête dans l'interrogation : pour lui, on utilise l'interrogation en tant que figure de style (la question rhétorique) « non pas pour marquer un doute ou provoquer une réponse mais pour indiquer au contraire la plus grande persuasion et défier ceux à qui on parle de pouvoir nier ou même répondre ». S'intéressant pour sa part aux interrogations totales et à leur valeur argumentative, O. Ducrot (1981) indique ces deux propriétés pragmatiques de l'interrogation évoquées par Fontanier que sont l'obligation de répondre d'une part et l'expression de l'incertitude d'autre part, dont il précise qu'elle peut être réelle ou feinte et qu'elle détermine l'aspect argumentatif de l'interrogation.

O. Ducrot (*ibid.*) ajoute à ces deux propriétés l'acte d'assertion impliqué dans la structure polyphonique de l'interrogation : l'interrogation « Pierre est-il venu ? » présuppose un point de vue « Pierre est venu ». Il propose alors de voir une valeur argumentative de l'interrogation dans le fait que l'interrogation « est-ce que P ? » s'appuie sur l'évidence d'une réponse « non P » et qu'avec une double négation, l'interrogation « est-ce que non P ? » prend pour allant de soi la réponse « P » (1981, 83-84).

La question rhétorique, dont la définition repose traditionnellement sur cette idée d'absence d'interrogation véritable ou d'« assertion déguisée » comme le rappelle J. Anis (1988 : 131) se caractériserait ainsi par une valeur argumentative, qui dans l'analyse de O. Ducrot (1981) est cependant commune à toutes les interrogations totales ; syntaxiquement, la

question rhétorique se restreint aux interrogatives directes mais peut se présenter comme le souligne A. Borillo (1981) sous une forme positive ou négative.

Les points d'interrogation dans l'article produisent des questions nécessairement « simulées » (Anis, *ibid.*), en vertu du caractère différé de la communication écrite, mais aussi des contraintes exercées par ce genre monologal qui s'autonomise par rapport à la situation de communication interpersonnelle ; ces questions signalent une modalité d'énonciation¹ spécifique (Meunier, 1974) – qui concerne l'énonciation, mais l'énonciation représentée dans le texte devrait-on préciser – et semblent ainsi pouvoir être traitées en des termes polyphoniques : elles mettent en scène dans le texte une interlocution, avec un sujet locuteur à l'origine de l'acte simulé de requête et partant, son interlocuteur. Si elles sont rhétoriques, elles ne constituent peut-être pas pour autant un simple ornement mais peuvent au moins pour une partie d'entre elles être au service de la démarche scientifique telle qu'elle est en jeu dans l'article².

Nous proposons de décrire quelques emplois privilégiés et pour certaines surprenants de ces interrogatives directes. Le trait commun à l'ensemble des emplois relevés est que ces questions simulées servent une forme de « mise en intrigue de la découverte »³ ; il s'agit d'impliquer un locuteur fictif qui construit le raisonnement scientifique, et un interlocuteur fictif (ou locuteur fictif de la réponse présumée par la question) pour créer un effet d'attente.

Le corpus atteste du caractère privilégié de ce procédé pour la présentation de l'objet de l'article conformément aux analyses de C. Schnedecker (2002) :

Comment Montaigne envisage-t-il la problématique des énoncés pyrrhoniens et comment répond-il aux objections de l'aproxia et de l'autocontradiction ? [LET006]

De tels emplois servent à circonscrire l'objet et à annoncer comment il va être traité (et selon quel plan). Ils mobilisent ainsi l'obligation de réponse impliquée par l'interrogation selon l'analyse d'O. Ducrot (1981), car ils appellent un développement⁴. Le locuteur mis en scène par la question peut être rattaché à l'auteur se donnant un tel objet d'étude ; mais l'auteur est bien aussi celui que la question interpelle et à qui elle s'adresse, question qui serait donc en cela à mettre aussi au compte d'une figure du lecteur en attente du développement.

C'est ainsi que la question simulée se présente comme une interrogation effective entre des pseudo-locuteurs représentés dans le texte ; plus exactement, c'est une interlocution qui est impliquée, avec pour caractéristique première sa réciprocité, i.e. le fait que les rôles de locuteur et d'interlocuteur (en tant qu'acteurs empiriques de l'acte d'énonciation) sont partagés. Il s'avère alors difficile de rattacher une voix interne à l'énoncé avec une figure de l'auteur et du lecteur au niveau du texte, car elles sont en quelque sorte mises en commun dans l'interlocution que met en scène l'énoncé. C'est sans doute sur ce type de ressorts que fonctionne ce qu'on a appelé une mise en intrigue de la découverte.

¹ A. Meunier (1974) distingue les modalités d'énonciation des modalités d'énoncé, en soulignant que si les secondes portent sur le dit et rassemblent les moyens linguistiques par lesquels le locuteur manifeste une attitude par rapport au contenu de l'énoncé, les premières portent sur le dire et marquent l'attitude du sujet de l'énonciation dans sa relation à son interlocuteur.

² C'est cette idée de la rhétorique que défend F. Hallyn (2004) dans son étude de *La rhétorique de la science* en montrant son rôle dans la construction des savoirs scientifiques.

³ Cette jolie formule est d'E. Nonnon (2002) dans un article où elle note l'importance de cette mise en intrigue dans l'écrit de recherche et sa didactisation, plus précisément au niveau de la construction d'une problématique.

⁴ Annonce de développement évoquée aussi on l'a dit par J. Anis (1988) comme une fonction des questions rhétoriques à l'écrit.

La fonction de problématisation de la question caractérise son emploi fréquent dans les titres d'articles et de sections. L'annonce de développement impliquée par l'interrogative directe se combine à la valeur d'« assertion déguisée » réputée propre à la question rhétorique :

Eh, un allomorphe de et ? [SCL093]

1) Le pont : ailes d'ange ou tour de Babel ? [LET095]

Conformément à l'analyse d'O. Ducrot (ibid.), la question rhétorique repose sur la présentation d'un point de vue asserté sous une forme interrogative, qui signale une incertitude déterminant sa dimension argumentative : l'obligation contractuelle de réponse appelée par la question sert doublement un effet d'attente et l'annonce du développement, dont l'orientation argumentative est de défendre le point de vue asserté. Notons toutefois que dans le premier exemple, l'article ainsi titré ne vérifie pas la proposition faite par O. Ducrot de voir dans le point de vue asserté la négation de la proposition qui fait l'objet de l'interrogation totale, l'idée défendue étant que *eh* serait bien un allomorphe de *et* ; cela dit, l'effet d'attente joue sur la possibilité inscrite linguistiquement dans la question d'une réponse négative. Dans le second exemple, l'alternative signalée par « *ou* » a pour effet de mettre l'accent sur l'incertitude, et servirait au final à défendre une double assertion (« Le pont : ailes d'ange et tour de Babel »).

Le titre manifeste ainsi une transparence à l'égard de l'objet qu'il nomme (par ex. dans le premier cas, il ne s'agit en aucun cas de défendre que « *eh* n'est pas un allomorphe de *et* »), selon une codification vraisemblablement propre au genre de l'article, et par laquelle le titre et le texte de l'article ne pourraient entretenir un rapport qui correspondrait linguistiquement à la structure décrite par O. Ducrot de réponse négative présupposée dans la question¹.

Par ailleurs, l'effet d'attente, en tant qu'effet pragmatique privilégié dans les titres, se construit énonciativement ; il repose sur le dédoublement des points de vue et la mise en scène de locuteurs : un « répondant » est impliqué par la question, c'est lui qui prendrait en charge la négation du point de vue faisant l'objet de l'interrogation selon l'analyse d'O. Ducrot (ou, dans le cas de l'alternative avec *ou*, un des deux points de vue à l'exclusion du second). Un autre locuteur fictif prend en charge la question elle-même ; il se présente comme indexable à l'auteur, et le point de vue mis en question annonce une assertion. Mais l'emplacement en titre signale également un décalage entre l'assertion directement impliquée par la question (i.e. l'assertion négative du répondant) : l'assertion appelée en réponse à la question se situe au niveau du texte, sous la forme d'une discussion ; la textualité dans le genre de l'article induit comme contrainte que l'obligation de réponse appelée dans un titre ne peut être réduite à la réponse linguistiquement inscrite dans la question.

Ces formes d'assertions déguisées se rencontrent également dans le corps du texte dans des formes canoniques pourrait-on dire de questions rhétoriques. Elles se présentent sur le mode fictif d'une question que l'auteur s'auto-adresserait pour faire avancer son raisonnement. Conformément aux analyses de O. Ducrot, la configuration appelle à voir dans le contenu propositionnel de la question une assertion négative ; comparativement aux titres, l'emplacement de la question dans le corps du texte amoindrit l'effet d'attente : l'assertion est

¹ Cette possibilité de sens semble exclue par l'article lui-même, et, partant, au niveau de la production du titre et de son interprétation.

définitivement impliquée par la question et non par un développement à venir¹. La question rhétorique sert ainsi à en appeler au consensus par l'évidence de la réponse qu'elle fournit :

Le moment semble venu de tirer les conclusions promises : pour l'amendement d'autrui, Montaigne accuse son "impatience" à souffrir "une forme diverse à la sienne", fût-ce celle des sots. **Son idéal n'est-il pas l'ataraxie stoïcienne, pour qui "la plus grande fadaise [est] de s'émouvoir et piquer des fadaises du monde" (ibid.) ?** Toute passion, tout mouvement est sottise. La véritable intelligence consiste donc à la supporter, surtout quand on n'en est pas exempt soi-même.

Ce type d'assertion déguisée suggère une convergence de figures de l'auteur et du lecteur dans la prise en charge de ces points de vue interrogatifs (mimer le raisonnement scientifique) et des négations qu'elles présupposent comme réponse (l'idéal de Montaigne est bien l'ataraxie stoïcienne, c'est évident). On en trouve d'autres formes, en particulier dans ce que V. Dahlet (2002 : 103) propose d'appeler des « fictions de dialogue »², où la question peut se présenter plus nettement comme le fait d'un lecteur ainsi mis en scène dans le texte ; c'est ce que l'on voit dans l'extrait suivant, où à la dimension dialogique de la question s'associe une double dimension dialogique et dialogale de l'assertion suivante (la réponse) avec la marque d'oralité « eh bien » :

Dit encore autrement, *aval* et *moucheron* n'ont pas seulement pour sens *manger rapidement* + statut présupposé et *petite mouche* + statut présupposé.

Qu'y a-t-il en plus ? Eh bien précisément la marque sémantique que le référent présupposé est une chose, une catégorie et non seulement un assemblage ou une combinaison de traits. C'est cet élément que nous allons développer maintenant [SCL099]

Le dédoublement énonciatif et la hiérarchisation des points de vue (impliqués par l'interrogation et partant, la négation, ainsi que par *eh bien*) font de la question une forme de marquage de la structure du texte³, de mise en intrigue de la découverte, et de procédé argumentatif tendanciellement polémique.

Quoiqu'elles restent relativement marginales, on trouve d'autres exemples de ces fictions de dialogue :

Mais à qui sont ces oreilles ? Ce sont celles de Montaigne, bien sûr, qui peut encore entendre parler d'amour. Celles de Montaigne ? Non, celles des femmes à qui il prétend donner conseil, celles de la lectrice qu'il vient séduire et posséder par la langue amoureuse des Essais. [LET016]

L'implication du lecteur dans de tels procédés peut aller jusqu'à être tout à fait explicite (comme dans « *Que trouvez vous à y redire cher collègue ? L'affaire est entendue, et depuis longtemps. [...]* », SCL106) où la question est nettement attribuée à un lecteur fictif, et sert à devancer une réaction supposée du lecteur, se rapprochant en cela des effets du point

¹ Quelques exemples, moins fréquents, montrent cependant là aussi la difficulté qu'il y a à exclure les possibilités inverses (une question suivie d'une réponse, qui en outre ne correspond pas à l'analyse d'O. Ducrot, comme dans : « *N'avons-nous pas là une voie qui nous permette de rencontrer Montaigne et de discerner quelque tendance matérialiste dans sa pensée ? Nous ne le croyons pas, pour deux raisons (...)* » [LET003].

² La désignation est discutable, car toute interrogation à l'écrit mime un dialogue ; cependant, elle a ceci d'intéressant qu'elle permet de distinguer les cas où on a affaire à une interrogation isolée, dont la réponse est impliquée dans la question ou annoncée comme à venir dans le texte, de ceux où l'interrogation est directement suivie de ce qui se présente comme une réponse possible (on pourrait imaginer plus aisément dans ce cas l'usage des tirets signalant des formes de discours direct).

³ *Qu'y a-t-il en plus ? et eh bien* servent en quelque sorte la représentation dans le texte d'une fonction phatique caractéristique des situations dialogales, et c'est ce qui semble-t-il pourrait expliquer, entre autres, l'effet de structuration des contenus que produit cette amorce.

d'exclamation. Elle sert un enjeu de démarcation (l'auteur va annoncer que « *l'affaire* » n'est pas aussi « *entendue* » que l'on croit), et elle est suivie par une autre intervention au discours direct, où c'est la voix de l'auteur qui intervient, avec des adresses explicites au lecteur (« *c'est cet unanimité qui précisément fait problème. [...] Souvenez-vous [...]* »).

Sans doute largement idiomatiques, ces usages montrent que la valeur argumentative de la question rhétorique peut passer par des cheminements plus complexes que ce que l'on trouve dans la question rhétorique que l'on pourrait dire canonique¹, qui vérifie la proposition de Ducrot d'y voir une négation de la proposition servant de base à la question. Tout du moins le point de vue défendu n'est-il pas nécessairement impliqué dans la question rhétorique elle-même, qui se présente en revanche toujours dans son rôle dialogique au service du raisonnement.

C'est ce que l'on voit aussi dans cet extrait où les questions se présentent comme un artefact heuristique qui permet à l'auteur d'introduire des alternatives interprétatives à propos de l'exemple (dans le premier paragraphe) qu'il analyse (dans le second paragraphe) :

(2) Dans ce même XIII^e siècle, des manuels versifiés se consacreront entièrement à ces deux sujets : Synonymes et Homonymes (Bursill-Hall, 1988). Les deux plus répandus sont attribués au célèbre Jean de Garlande. Souvent utilisés dans les cours d'Université, ils sont accompagnés de commentaires (Hunt, 1991).

Si les deux sources signalées par le parenthésage peuvent aisément être rapportées aux énoncés qui les précèdent, **se pose toutefois la question de l'attribution énonciative du chaînon intermédiaire : quelle est la source de l'information selon laquelle "les deux plus répandus sont attribués au célèbre Jean de Garlande" ? Bursill-Hall 1988 ? Hunt, 1991 ? Ou bien l'auteur de l'article lui-même ?** L'interruption de la chaîne d'indexation dans un discours fortement indexé fragilise les chaînons non indexés. [SCL034]

La première question posée par l'auteur (« *quelle est la source de l'information ?* ») est pour ainsi dire une véritable interrogation, annoncée d'ailleurs comme telle (« *se pose la question de* »). Les questions suivantes (« *Bursill-Hall 1988 ? Hunt, 1991 ?* », etc.) sont autant d'alternatives interprétatives grâce auxquelles l'auteur montre l'ambiguïté de l'attribution énonciative dans le passage qu'il analyse : sa première question n'a pas de réponse, car c'est l'indécision qui l'emporte face à ce passage analysé.

Si l'on se fie à l'analyse de Ducrot, on considérera que pour les interrogations totales (« *Bursill-Hall 1988 ? Hunt, 1991 ?* », etc.), les points de vue présentés sous forme interrogative trouvent une réponse négative. Cela étant, la valeur argumentative des questions ne se limite pas à la réponse négative qu'elles présupposent. La subtilité du procédé tient au fait que c'est à l'indécidabilité que mènent ces questions : le point de vue défendu par l'auteur ne se situe pas tant dans ces réponses négatives que dans le fait qu'aucune piste interprétative sur l'attribution énonciative du passage analysé ne se dégage, car toutes ces pistes peuvent aussi bien être niées que supposées valides.

Cet aperçu de quelques emplois du point d'interrogation dans l'article montre la complexité de la forme interrogative directe en tant que phénomène dialogique et argumentatif. L'analyse de la valeur argumentative en termes de négation (Ducrot, 1981) présupposée par la question fonctionne bien au niveau de la structure linguistique de la question prise isolément, et dans la plupart de ses emplois dans les textes. Mais elle peut être contredite quand une réponse suit la question, et dans le cas particulier des titres où la dimension structurelle du texte introduit une

¹ Peut-être plus répandue, dans le discours journalistique par exemple.

contrainte forte sur l'interprétation, qui consiste à annoncer que le texte est la réponse, apportée sous forme d'une discussion.

La question rhétorique que l'on a appelée canonique consisterait en une question isolée, interro-positive ou négative, qui impliquerait en elle-même sa réponse, à savoir la négation de la proposition qui fait l'objet de l'interrogation totale. Dans les autres emplois plus complexes de « fiction de dialogue » ou de succession de questions mimant une fonction heuristique, la négation impliquée par la forme interrogative est exploitée dans un jeu de dédoublement énonciatif et de hiérarchisation des points de vue.

Le dédoublement énonciatif à la base des questions est à l'origine d'un effet d'incertitude feinte et d'attente qui suggère leur intérêt pour la mise en intrigue de la découverte qu'elles rendent ainsi possible, en association éventuellement avec d'autres procédés linguistiques¹. Outre l'effet rythmique ainsi produit², la question se présente dans le texte scientifique comme un outil heuristique, par lequel sont mises en scène en tant que questions linguistiques des questions d'ordre épistémique, comme le suggère leur usage privilégié dans la construction de l'objet avec l'annonce de la problématique en introduction et son ouverture en conclusion.

Le point de vue finalement asserté par l'auteur à travers l'usage des questions se construit ainsi selon une procédure de « surénonciation » (Rabatel, 2005) qui sert l'argumentation, mais celle-ci se présente sous une forme de dialogue maïeutique³ de sorte que la surénonciation procède aussi de formes de coénonciation. Ainsi, les questions impliquent le lecteur dans la construction du sens *via* des figures plurielles qui sont celles des locuteurs à l'origine de la question, de la réponse qu'elle présuppose et de la réponse énoncée ; dans la représentation que donne de lui le texte, au moyen de ces instructions dialogiques, le lecteur se présente finalement comme un interprète qui se questionne, en des termes épistémiques et cognitifs, avec le scripteur.

L'étude des questions dans un genre particulier semble finalement permettre de mieux distinguer ce qui se joue au niveau de la structure linguistique de la question elle-même et de son usage textuel (selon qu'elle est en titre ou non, qu'elle entre dans des séries de questions propres aux fictions de dialogue ou non). Il est ainsi possible de repérer un usage canonique de la question rhétorique et d'autres emplois propres à une rhétorique de la découverte. Se manifestent ainsi des usages singuliers, sans doute largement idiomatiques, qui déplacent pour certains les caractéristiques attendues du genre, et du moins s'en jouent.

Avec les guillemets, dernier signe dont traitera notre étude des ponctuations dans l'article, nous retrouverons, sinon cette idée de jeu avec les normes du genre qui affleure dans l'usage des points d'exclamations, l'intérêt manifeste des faits énonciatifs et plus spécifiquement polyphoniques pour l'étude du genre, et, en retour, l'intérêt du genre pour l'étude du fonctionnement de ces faits linguistiques.

¹ Par exemple le futur dans LET004 : « C'est aux bains de Lucques justement que Montaigne apprend par courrier que ses concitoyens l'ont élu maire de Bordeaux. Voici donc que la pratique du pouvoir va suivre la théorie politique: Montaigne sera-t-il ce magistrat décidé à respecter la "franchise" et la "liberté" de ses administrés comme l'invitaient à le faire tant son expérience de l'Allemagne que le livre de son ami La Boétie? ».

² Une telle fonction ne va pas sans rappeler l'aspect ornemental des procédés rhétoriques, tout en signalant l'intérêt de tels ornements dans la lecture du texte.

³ Ce procédé est exploité dans un article de P. Siblot paru dans les *Cahiers de Praxématique*, entièrement bâti sur le modèle du dialogue socratique à partir des interventions de deux locuteurs introduites, à l'aide de tirets, sous la forme de discours direct. Le procédé est attesté mais de manière plus locale dans un article du même auteur compris dans le corpus (SCL106), ce qui suggère que les prédilections stylistiques, i.e. propres à un auteur, jouent pour beaucoup sur l'emploi de ces marques.

5.2.6. Les guillemets : l'importance d'un « simple » signe dans le genre

Nous nous centrerons pour clore cette étude des ponctuations dans l'article sur une marque de ponctuation délaissée par l'analyseur automatique, et pourtant importante ne serait-ce qu'en termes quantitatifs, les guillemets. Il s'agit de poursuivre la caractérisation du genre, en montrant comment l'article, de manière transdisciplinaire ou disciplinaire, appelle certains usages spécifiques, plus ou moins conformes aux caractéristiques attendues dans le discours scientifique, et liés aux enjeux de ces textes, en termes de rapport au lecteur et à l'objet.

Ainsi, les guillemets peuvent permettre de montrer comment une telle entrée formelle, *a priori* loin d'être primordiale, peut servir l'analyse du genre et d'identités disciplinaires, et attester en retour de l'incidence des genres sur l'étude de la langue, en particulier dans la mise en évidence d'emplois relativement stables de ce signe au fonctionnement reconnu comme éminemment complexe.

En guise d'introduction : la complexité des guillemets

Pour être auxiliaires, les guillemets n'en sont pas moins nécessaires à la construction du sens. Traditionnellement, ils se voient assigner une fonction prioritaire de marquage citationnel ; conventionnelle, celle-ci permet de signaler la rupture énonciative introduite par la citation (comme on l'a vu dans la macro-analyse du genre), et, partant, les plans d'énonciation¹ dans le cas du discours direct.

Mais comme le note C. Kerbrat-Orecchioni (1980), tous les mots sont susceptibles de recevoir des guillemets. Aussi est-il convenu de dire que de cette première fonction de marquage citationnel en découleraient d'autres, comme celle qui consiste à encadrer un autonome, un terme emprunté à une langue étrangère, un terme argotique ou un néologisme : plus largement que la seule fonction de citation, « les guillemets visent à isoler un énoncé étranger inséré dans un énoncé principal » (Laufer, 1980 : 83). En cela, les guillemets signalent une distance énonciative du scripteur par rapport aux termes présentés comme rapportés ; mais on leur attribue également la propriété de mettre en relief un terme (Riegel, Pellat, Rioul, 1994 : 94), si bien que la question des effets de sens attachés aux guillemets reste relativement ouverte.

Si la prise en compte des guillemets est indispensable dans le parcours interprétatif du texte, elle le complique singulièrement ; en effet, les guillemets bloquent l'interprétation littérale du segment qu'ils entourent, mais ils ne fournissent en eux-mêmes d'autre instruction que cette rupture (Authier-Revuz, 1981), et peuvent ainsi être associés à la multiplicité d'interprétations évoquées. Ils sont cependant parfois accompagnés d'indices co-textuels (constructions syntaxiques spécifiques, ponctuation, introducteurs lexicaux) qui peuvent résoudre en partie cette multiplicité interprétative, mais le fait est que la polysémie est la caractéristique première de ces signes. Aussi les entreprises typologiques sont-elles extrêmement nombreuses, et aboutissent toutes au constat d'une difficulté, sans compter que la question se complique avec la concurrence de l'italique, comme le montrent, entre autres études, celles de J. Authier (1981, 1989, 1995), R. Martin (1983), et G. François (à venir).

Plutôt que d'envisager comme première fonction le marquage citationnel, duquel découleraient d'autres fonctions assimilées, il convient à l'instar des propositions de J.

¹ En cela, ils sont non seulement nécessaires à l'identification des voix en présence, mais également pour l'analyse de relations textuelles comme la résolution d'anaphores, dans le sens où la continuité thématique souvent exprimée à l'aide d'expressions anaphoriques n'est généralement possible qu'à l'intérieur d'un même plan énonciatif.

Authier-Revuz (1989) de renverser la perspective en concevant les guillemets comme une forme de modalisation autonymique. C'est du moins le parti pris choisi dans le cadre d'un projet mené par le laboratoire Lidilem sur les marques de la subjectivité dans les textes et leur annotation et leur exploitation pédagogique à l'aide d'outils du TAL¹, projet auquel nous avons contribué. L'équipe a souligné que les guillemets ont partie liée avec trois problématiques², celle du signe d'abord et de sa double fonction de mention (fonction autonymique) et d'usage (le signe réfère à autre chose qu'à lui-même), celle de l'hétérogénéité énonciative ou de la polyphonie d'autre part, et celle enfin de la nomination et les différentes facettes de cette activité : catégorisation, expression d'un point de vue, construction des objets du discours.

En prenant ces analyses comme point de départ, il s'agira donc pour nous de montrer à partir d'exemples tirés du corpus quels semblent être les emplois privilégiés des guillemets dans les articles et dans chacune des deux disciplines considérées.

Nous proposons ainsi d'envisager dans quelle mesure les fonctions remplies par les guillemets sont communes ou spécifiques à chaque discipline, et dans quelle mesure elles se conforment aux caractéristiques attendues du discours scientifique. Auparavant, nous fournirons un rapide aperçu du poids des guillemets dans les articles, et plus précisément dans un sous-corpus de 20 articles de LET et de 20 articles de SCL.

Le poids des guillemets

Le comptage des signes doubles ouvrants et fermants que constituent les guillemets montre l'importance des passages entre guillemets dans l'article : on en dénombre en moyenne 63 par texte. En LET, le nombre moyen est de 70,1 passages entre guillemets, et peut être compris entre 25 et 154. En SCL, la moyenne est légèrement inférieure, avec 55,8 passages entre guillemets, et le nombre peut varier entre des bornes plus grandes de 12 et 188 passages entre guillemets.

La différence entre les deux disciplines ne s'avère pas statistiquement significative, ni la question de savoir si le scripteur est doctorant ou non. La pratique des guillemets dans notre corpus est donc relativement homogène en ce qui concerne le poids de ces signes. Il nous faut alors nous pencher sur leurs fonctions ; l'importance des guillemets dans le texte scientifique, dont atteste ce nombre important de passages entre guillemets invite en particulier à se demander dans quelle mesure leur usage se circonscrit ou non à la pratique citationnelle, en tant que caractéristique attendue du genre, et aux autonymes pour les exemples de SCL. Il s'agit plus largement de voir si les deux disciplines attestent, de même qu'au niveau du poids des passages entre guillemets, d'emplois communs ou spécifiques de ces signes.

¹ Projet Lidilem plan pluri-formation piloté par G. Antoniadis et F. Grossmann, 2003-2007, « Développement et exploitation de ressources linguistiques pour la didactique du français à l'aide d'outils TAL ».

² Une typologie des fonctions des guillemets a été établie et testée au moyen d'une annotation en corpus de ces fonctions. Le but est de voir à partir de ces trois problématiques comment les fonctions de signalement d'un autonyme, d'une modalisation, d'une citation, et d'une dénomination se configurent dans les passages entre guillemets, en recourant à une méthode dite d'annotation inter-annotateurs (réalisée indépendamment par trois annotateurs humains), afin de vérifier la stabilité des catégories et de mettre en évidence les zones de désaccord. Les résultats, encourageants, attestent d'un pourcentage de 80% d'accord en moyenne, les plus haut taux de désaccord concernant l'attribution d'une double valeur (polysémie des guillemets) *versus* l'attribution d'une valeur unique (Rinck, Tutin, Hermann, 2004).

Les emplois autonymiques

Une première fonction des guillemets aisément identifiable dans les articles est la fonction exclusivement autonymique des guillemets. Vraisemblablement assez peu attestée dans d'autres discours, elle est bien entendu caractéristique du texte de linguistique comme en atteste son importance dans le corpus d'articles de cette discipline. En voici trois exemples :

Chien DF FUT 3S boire eau « **Ce chien boira de l'eau** » [SCL061]

Rappelons qu'on écrivait jadis :

<i>ynno</i>	pour <i>inno</i>	“ hymne ”
<i>excellencia</i>	pour <i>eccellenza</i>	“ excellence ”
<i>justitia</i>	pour <i>giustizia</i>	“ justice ”
<i>thosco</i>	pour <i>tosco</i>	“ toscan ”
<i>scripto</i>	pour <i>scritto</i>	“ écrit ”
<i>hoggi</i>	pour <i>oggi</i>	“ aujourd'hui ”
<i>extrahere</i>	pour <i>estrarre</i>	“ extraire ”
		[SCL014]

« **Je...** » forcerait à imaginer une phrase multiple composée de trois phrases distinctes [SCL030]

L'usage des guillemets marque une rupture de la syntaxe comme on le voit bien dans le dernier exemple avec l'accord du verbe, et plus largement une rupture dans le statut du signe qui se trouve présenté pour lui-même (valeur autonymique, dite aussi de mention). Ils peuvent ou pourraient s'accompagner de termes métalinguistiques (par exemple, une indication du type « soit la phrase... » pour l'exemple cité *supra* « le chien boira de l'eau » en SCL061).

Ce signalement de la valeur de mention d'une unité linguistique (morphème, lexème, sémème, phrase, etc.) n'est pas le seul fait des guillemets : il peut se réaliser à l'aide des italiques ou des crochets, ou encore d'une disposition spécifique par laquelle l'autonyme, qui sert d'exemple, se trouve décroché dans un paragraphe isolé comme nous l'avons dit en traitant des exemples et des citations. On voit *supra* dans le second extrait tiré de SCL014 (« *ynno* pour *inno* “ hymne ” ») comment les guillemets et l'italique peuvent ainsi faire l'objet d'un usage conjoint à valeur distinctive.

De l'exemple à la citation

L'importance de la fonction autonymique dans les articles de SCL répond à la nécessité de signaler un échantillon de la langue pris en tant que tel. Elle pose donc la question des exemples, et cette entrée formelle permet de rejoindre le constat fait pour ces composantes structurelles : l'exemple en SCL ne se limite pas au strict autonyme. Aussi les passages entre guillemets peuvent-ils signaler qu'on a affaire non seulement à un échantillon de la langue, mais à un échantillon énoncé. Dans l'article SCL036, qui étudie les échanges dans la communication électronique en particulier au moyen des icônes dans les forums, on trouve ainsi l'extrait suivant :

Ainsi coolrasta "**aime** " ou "**sourit à** " tout participant d'apparence féminine. [SCL036]

Le statut donné aux deux passages encadrés par les guillemets est celui d'une citation, bien qu'il s'agisse ici de transcription d'icônes et non d'une réduplication d'un énoncé verbal. Un

échantillon du matériau étudié par l'auteur est cité ici en tant qu'il sert l'analyse de l'usage des icônes dans les échanges entre les participants du forum.

Si l'autonyme strict reste rare dans les articles de LET¹, les passages entre guillemets que l'on y trouve prennent pour beaucoup cette fonction de citation du matériau étudié. L'extrait suivant permet de voir l'omniprésence du matériau étudié, qui se trouve cité sous la forme d'un extrait de quelques lignes et repris sous la forme d'îlots textuels (J. Authier-Revuz, 1989, Grossmann et Boch, 2002) dans le commentaire qui introduit et suit cet exemple :

Toujours à Augsburg, le voyageur est réjoui par le surgissement de jets d'eau qui viennent, par le mystère de quelque mécanisme secret, mouiller par en-dessous les "**cotillons des dames**".

Entre ces deux gardeirs, il y place de dix pas de large planchée d'ais; au travers de ces ais il y a force petites pointes d'airain qui ne se voyent pas. Cependant que les dames sont amusées à voir jouer le poisson, on ne fait que lascher quelque ressort: soudain toutes ces pointes **eslancent** de l'eau menue et roide jusques à la teste d'un homme, et remplissent les **cotillons des dames** et leurs cuisses de cette fraîcheur (...)

On ne saurait douter que cette mode des arrosages clandestins ne délimite en propre, dans le Journal, l'aire géographique de la Suisse et de l'Allemagne: la force des humeurs libère un humour roboratif et sain; des "**élançements**" de l'eau naît un sentiment d'émancipation du corps et de l'esprit. [LET004]

Cette fonction bien connue de marquage de citation par les guillemets est exploitée en SCL dès lors que le matériau relève de pratiques langagières attestées :

En effet, même si les questions sont focalisées sur la variation, les témoins parlent de la norme en la caractérisant comme un véritable " technoclecte ". Elle serait "**plate**", "**impersonnelle**", "**neutre**" (**je cite des termes apparaissant dans la bouche des informateurs**), **uniformisée en un mot, disent la plupart des témoins au détour de l'entretien** [SCL007]

Cette pratique citationnelle est vraie non seulement pour ce qui est pris comme exemple, mais aussi pour ce que l'on a traité dans la macro-analyse du genre comme des citations stricto-sensu de travaux de pairs : dans les deux cas, quand elles sont détachées dispositionnellement dans le texte, elles ne sont généralement pas accompagnées de guillemets, mais c'est le cas quand elles sont suffisamment courtes pour justifier d'être incluses. Des passages de quelques lignes de travaux de pairs peuvent ainsi être cités dans le texte, de même que les exemples peuvent être soit intégrés soit décrochés.

Signalons cependant un usage peut-être moins attendu de guillemets qui servent à mettre en scène une pseudo-citation² :

C'est une façon pour les rappeurs de proclamer "**nous détenons le trésor de votre belle langue, mais notre langue est infiniment plus colorée, plus riche car elle chante les rues, le soleil, l'exil, la joie et la misère ; notre trésor se nomme pluralité**". [SCL009]

La citation, au plan formel, introduit des propos fictivement prêtés aux rappeurs et correspond de fait au discours de l'auteur lui-même : cette mise en scène spécifique lui sert à analyser le comportement et les représentations de la langue qu'ont les locuteurs de manière tout à fait

¹ On peut citer comme exemple le premier et le troisième passage entre guillemets dans « Il y a bien une figure dans ces vers, mais elle consiste plus en une personnification de la ville qui permet alors de donner à " entrailles " son sens " général " (" toutes les parties enfermées dans le corps des animaux ") qu'en un trope portant sur le mot lui-même » (LET068).

² Nous avons vu en traitant du point d'exclamation que ces pseudo-citations étaient également attestées en LET où elles peuvent servir comme dans l'exemple analysé ici à faire parler le matériau étudié.

singulière par rapport aux traits attendus du discours scientifique, en parlant à leur place sous la forme d'une énonciation personnelle caractérisée par l'usage de subjectivèmes affectifs et d'axiologiques.

Nommer et catégoriser

Quand elle se réduit à un îlot textuel intégré dispositionnellement et syntaxiquement au texte, la citation se double d'une fonction dénominative. Les guillemets signalent que le terme est à envisager comme une étiquette symbolique, qui sert la construction d'un appareillage terminologique et conceptuel dans l'écrit de recherche :

C'est en m'appuyant sur la distinction dégagée par John J. Gumperz (1982) entre « **we code** » et « **they code** », que j'ai étudié les productions de groupes marseillais, en posant l'hypothèse que le discours rap était le lieu de (...) [SCL009]

Pouvant faire l'objet de (re)définitions (« *nous entendons par 'pictogramme'...* »), être justifiées,

En effet, **si on considère que** l'approche très méthodique mise en œuvre par L1 est en partie due à l'expérimentation, **on peut en déduire qu'il s'agit là d'une sorte d' 'hypercorrection interprétative'**. **C'est-à-dire que** dans le comportement de L1 le degré d'attention portée au discours (ici à l'interprétation) est supérieur au minimum (Labov, 1976, p.289) et ce degré d'attention exerce une influence non négligeable sur son comportement interprétatif. **Nous évoquons ici l'hypercorrection non pas au sens** de l'usage déplacé d'une règle imparfaitement assimilée (Labov, 1976, p.193) **mais au sens** que L1 outrepassé (...) [SCL003]

ou être discutées en tant que telles :

Ils se situent dans une tranche d'âge étroite de jeunes adultes normalement socialisés et acculturés à l'écrit. Insuffisamment avancés dans leurs études **pour pouvoir être qualifiés de lecteurs " experts "**, ils constituent un groupe relativement homogène de " lecteurs ordinaires" [SCL003]

ces dénominations servent l'activité de catégorisation à l'œuvre dans le discours scientifique¹ ; elles acquièrent, par leur mise entre guillemets, un statut proche de celui de désignateur rigide caractéristique du nom propre (« *la notion de " hors-sujet" »* LET025), qu'elles se présentent comme attestées dans le champ, ou comme une proposition inédite à l'initiative de l'auteur lui-même.

Un autre usage des guillemets fréquent dans les articles de LET, et plus que dans ceux de SCL, concerne l'indication de titres : « *les deux premiers poèmes, " Paquita la Sévillane " de Jan Diaz et " Elégie " de Lousteau* » (LET023). Les titres mobilisent une double fonction dénominative et citationnelle des guillemets, en servant la désignation d'un texte ou d'une partie de texte au moyen du nom qui lui a été assigné.

¹ Elles s'accompagnent souvent de termes métalinguistiques et éventuellement de prédicats d'existence. Mais cette fonction dénominative des guillemets ne sert pas que la conceptualisation, y compris dans l'article où on la trouve, par exemple, dans le syntagme « vache folle » pour désigner la catégorie d'ordre événementiel construite dans le champ social (et dont l'auteur analyse le traitement dans la presse), ou dans la « bise » et l'analyse de son traitement dans le dictionnaire de Furetière : « *les propriétés de la " bise " sont ainsi définies en des termes si proches de ceux de ce vers de Malherbe choisi comme exemple que l'on pourrait croire que c'est au poète même que le lexicographe emprunte son savoir* »; l'usage des guillemets signale que ce n'est pas aux entrées du dictionnaire que l'auteur réfère (il utilise systématiquement les majuscules dans ce cas) mais à la catégorie ontologique désignée.

Signalons un exemple singulier de cet usage des guillemets pour les titres dans l'article suivant,

Parmi d'autres études, les travaux de William Labov sur " **le parler ordinaire** " au sein de groupes de pairs dans les ghettos noirs (Labov, 1993 [1978]) ont fait largement école, aux États-Unis et en Grande Bretagne, mais aussi en France (...) [SCL009]

où l'auteur renvoie à l'ouvrage lui-même par le système auteur+ date, mais en évoque subtilement le titre non pas en tant que tel¹ mais en tant que concept central dont traite cet ouvrage.

Marquer sa distance

Si les fonctions citatives et dénominatives sont attendues dans le discours scientifique et caractéristiques de ce dernier, d'autres emplois peut-être moins conforme à l'image que l'on s'en fait interviennent : comme nous allons le voir, ils intègrent tout en le contredisant le mythe du « mot juste », qui affleure dans les manuels d'aide à la rédaction scientifique en particulier ; on pourrait envisager cette idée du « mot juste » sous deux angles, en termes de correction linguistique (le bon usage) et en termes de coïncidence entre les mots et les choses, coïncidence attendue (par l'interlocuteur) dans ce discours en vertu d'un principe de réalité qui caractérise sa dimension référentielle.

Le cas de la correction linguistique peut être illustré par les néologismes, employés dans les articles et mis entre guillemets avec pour effet de prévenir le lecteur de ce statut linguistique spécifique. Ainsi en est-il de « *arabiser* » (à propos d'une prononciation empruntant des traits caractéristiques de la langue arabe par contraste avec le français dans SCL009), ou de « *pauser* », que l'auteur d'un article sur les pauses dans les imitations d'hommes politiques emploie à maintes reprises dans son texte et systématiquement avec des guillemets dans SCL080, ou encore, en LET, dans « *son expérience de "liseur"* » (LET085) et « *le pyrrhonien parvient au point où il peut "pousser hors " le contenu de ses énoncés: ils s'emportent eux-mêmes* » (LET006). Les guillemets peuvent également intervenir dans le cas d'expressions lexicalisées, mais dont ils indiquent le caractère impropre (compte tenu du genre ou du co-texte) : « *faire " parler " les productions recueillies* » SCL009, ou encore, « *la langue des signes, langue « parlée » par les sourds ou malentendants* »².

Dans d'autre cas, c'est, plutôt que le mot lui-même, le rapport à l'objet qu'il désigne qui est signalé comme inadéquat par l'usage des guillemets – quoique la distinction, proposée par J. Authier-Revuz qui rend compte des différents types de commentaires méta-énonciatifs est loin d'être aisée comme le signale d'ailleurs l'auteur. Des exemples de divers types peuvent en être données :

D'une certaine manière, ce que nous essayons de montrer, c'est que d'un point de vue syntaxique, il n'est pertinent de parler de discours rapporté que si la séquence intégrant discours citant et discours cité fonctionne comme un discours " **normal** ", comme une phrase. L'analyse des exemples suivants devrait inscrire les différents formatages du rapport discours citant/discours cité dans un continuum d'intégration syntaxique. [SCL032]

Pourquoi un écolier "**décide-t-il**" d'utiliser un mot plutôt qu'un autre ? [SCL072]

¹ L'absence de majuscule en atteste, en suggérant ainsi la proximité de l'usage des guillemets pour les titres et pour une fonction plus large de signalement d'une catégorie conceptuelle.

² Hors-corpus, cité dans (Rinck, Tutin et Hermann, 2004).

Ou encore, dans un article consacré aux « *pratiques langagières d'adolescents en milieu urbain* », où l'auteur dit ne pas souscrire à l'appellation « *langage des jeunes* », la mise entre guillemets systématique du terme « quartier » :

Pour ce faire, j'ai enregistré des jeunes locuteurs à Grenoble, dans le "quartier" Chorier-Berriat, un ancien faubourg industriel et ouvrier, dont il convient de donner un rapide aperçu [SCL009]

Les guillemets semblent concerner ici doublement le rapport entre mot-chose, et le rapport à l'interlocuteur ; ils se présentent comme un procédé prévenant, qui devance les attentes prêtées au lecteur : le recours à un commentaire méta-énonciatif signale la distance énonciative du locuteur dans la prise en compte d'un interlocuteur à qui il importe de signaler la non-adéquation du terme mis entre guillemets.

Le texte se charge d'implicites¹ : la mise entre guillemets de « décider » fonctionne ainsi sur un mode allusif, en suggérant une réserve à l'égard d'une conception intentionnelle selon laquelle on « décide » complètement des mots qu'on utilise. Avec « quartier », un autre point de vue est également sous-jacent, par rapport auquel l'auteur marque sa distance : il mobilise l'usage commun du terme, tout en signalant qu'il s'agit d'un discours « autre », en l'occurrence une dénomination dont l'usage dans le champ social est chargé de connotations ; celles-ci justifient ainsi la distinction signalée par les seuls guillemets² entre l'emploi du terme dans ce champ et son emploi scientifique – qui fait l'objet d'une forme de réflexivité.

Ces commentaires méta-énonciatifs suggèrent ainsi une forme de rumeur dans le texte scientifique, une rumeur des textes implicitement désignés par des points de vue signalés comme « autres », qu'il s'agisse des textes produits dans le champ social dans son ensemble ou dans le seul champ scientifique. Ils attestent d'un rapport spécifique au lecteur, car le fait de devancer ses attentes supposées construit l'image d'un interlocuteur doté des pré-requis culturels nécessaires pour repérer dans les points de vue signalés comme « autres » leur source énonciative.

Par rapport aux hypothèses préalables que l'on a pu avoir en abordant les guillemets dans l'article, nous pouvons être surpris de constater que la fonction de précaution que peuvent introduire les guillemets n'est pas représentative dans notre corpus, alors qu'elle paraît fréquente dans l'usage de ces signes ; cet emploi relève aussi de la catégorie du marquage de la distance, et se présente comme un procédé utile pour éviter de s'engager sur le choix d'un terme, en signalant son inadéquation et le fait qu'il ne relève pas du domaine du *je* du locuteur mais est imputable à une instance énonciative « autre ». S'il concerne le degré d'implication du locuteur dans le texte, notre corpus suggère que c'est plutôt l'implication dans le texte du lecteur et des points de vue existants qui est centrale dans la distance énonciative introduite par cet emploi des guillemets.

Ces mises entre guillemets signalent un rapport au lecteur que l'on pourrait qualifier de « savant » et qui se veut aussi un rapport savant à l'objet ; elles montrent comment le texte

¹ De tels commentaires méta-énonciatifs peuvent cependant servir une volonté d'explicitation dans des formes de reformulations didactiques qui laissent transparaître le travail de conceptualisation ; ainsi en est-il dans cet exemple : « *on aurait donc une représentation assez différente de ce qu'on a l'habitude d'entendre. La norme ne serait pas perçue comme une variété de langue sortant de l'ordinaire, particulièrement valorisée en tant qu'instrument technologique de communication (Marie-Louise Moreau, article " Norme " in Moreau éd., 1997), doté de propriétés qui ont été élaborées, historiquement construites, " ajoutées " à un produit de départ " brut " pour le rendre plus adéquat à certaines fonctions sociales, à travers des qualités comme la précision, la concision, l'harmonie, la clarté, etc.* » (SCL007).

² En revanche, l'auteur explicite ailleurs dans le texte son refus du terme banlieue à l'aide d'une négation polyphonique (« *On est donc face à une communauté d'usages, un socio-topo-lecte urbain grenoblois (et non de " banlieue ")* »), qui laisse cependant implicite là encore la source de ce point de vue qui consiste à désigner par *banlieue* ce que l'auteur préfère désigner par le composé savant *socio-topo-lecte urbain*).

scientifique se présente en prise contre le sens commun (cf. par exemple avec « quartier »), mais pourrait manifester cependant aussi certaines doxa à l'œuvre, dans le choix des termes qui font l'objet des commentaires méta-énonciatifs, et, par exemple, dans le refus feint de prendre en charge la qualification de « discours "normal" », soit que cette étiquette puisse être perçue comme floue ou peu conceptualisée, soit qu'elle puisse être jugée peu conforme au parti pris descriptif de la linguistique.

Bilan sur les guillemets

Si l'usage des guillemets ne fait pas l'objet d'une différence disciplinaire en termes de poids de ces marques dans les textes, leurs fonctions dans les deux disciplines se distinguent, tout en présentant des traits communs vraisemblablement caractéristiques de l'écrit de recherche.

La fonction de citation est primordiale en LET, où elle est utilisée pour le matériau étudié principalement, pour les titres de textes (ou de chapitres, etc.) et dans une moindre mesure dans le cas de citations de travaux de pairs. D'autres emplois interviennent de manière plus minoritaire, les guillemets à fonction de dénomination qui signalent l'introduction d'un concept d'abord, les autonymes d'autre part, et enfin des commentaires méta-énonciatifs sur l'adéquation d'un terme, dans le cadre d'interprétations du texte où des formes d'inventions lexicales sont présentées en tant que telles, et comme étant propices à la justesse de l'analyse.

En SCL, il semble plus difficile de dégager un emploi dominant des guillemets. Cette diversité rejoint notamment ce que nous avons dit de la diversité de ce qui est pris comme matériau. On trouve de nombreux autonymes, des citations (du matériau étudié ou de travaux de pairs), des emplois signalant une dénomination, et secondairement des commentaires méta-énonciatifs qui thématisent l'adéquation d'un terme et, indissociablement, sa prise en charge. Les titres sont quant à eux beaucoup plus rares, et extrêmement marginaux par rapport à ce que l'on observe en LET.

Il convient toutefois de signaler le lien qui existe vraisemblablement entre ces usages et les types d'étude, ne serait-ce que parce que certains articles de SCL ne recourent qu'aux autonymes et certains de LET qu'aux citations, et que les emplois pour lesquels la valeur de commentaire méta-énonciatif l'emporte seraient, peut-être, à comprendre comme le fait de certains sous-domaines ou de certains effets d'école.

Ce seul signe de ponctuation *a priori* accessoire se révèle donc constituer une entrée particulièrement intéressante pour l'analyse d'un genre, et en l'occurrence aussi, des identités disciplinaires qui se manifestent dans les articles : à titre de pistes, nous avons ainsi vu comment les guillemets signalent l'activité de catégorisation et la conceptualisation à l'œuvre dans le texte, la recherche stylistique d'une justesse de l'analyse qui passe par des inventions lexicales, la manière dont l'article intègre son matériau dans l'analyse, ou encore comment le texte se constitue comme texte scientifique contre des points de vue doxiques, tout en intégrant une forme de doxa spécifique au champ disciplinaire ou domaniale considéré.

En retour, l'analyse des guillemets dans un genre particulier met en évidence sa complexité. Elle montre que trois types d'instructions peuvent ainsi être signalées par les guillemets et leur co-texte quant à l'interprétation du passage entre guillemets (PEG) :

1. Les guillemets (et le co-texte) indiquent que le PEG est à comprendre comme un autonome exclusivement¹

¹ J. Authier-Revuz a bien montré qu'en tant que formes de commentaires méta-énonciatifs, les guillemets impliquent nécessairement l'attribution d'une valeur de mention au PEG ; cependant, face aux difficultés posées dans les nombreuses typologies des emplois des guillemets, elle souligne l'importance de distinguer dans un

-
2. Les guillemets (et le co-texte) introduisent un commentaire méta-énonciatif qui dédouble le dire en signalant une altérité plus ou moins grande dans la composante modale du PEG (citation ou distance de l'auteur), et qui peut affecter la coïncidence mot-chose ou le rapport à l'interlocuteur¹
 3. Les guillemets (et le co-texte) mettent l'accent sur la fonction dénominative du PEG, i.e. sa fonction de catégorisation symbolique.

Ces instructions se conjoignent et/ou se spécifient dans des configurations relativement stables d'emplois des guillemets dans l'article ; ils impliquent principalement, mais à des degrés divers et dans des configurations variées la question de la référence et une dimension modale, et demanderaient à être comparés avec les observations faites dans d'autres genres de textes, et dans d'autres discours.

Après avoir envisagé le genre de l'article d'un point de vue morpho-syntaxique, en nous centrant également sur ses ponctuations, nous poursuivrons l'analyse de ces paliers micro-textuels en nous intéressant au lexique. Comme celui-ci varie considérablement en fonction de l'objet même de l'article, hormis sans doute pour des traits caractéristiques de la langue vraisemblablement peu productifs pour la caractérisation d'un genre en particulier, nous avons fait le choix de traiter d'une partie spécifique du lexique : nous le désignons en tant que lexique transdisciplinaire, en raison de son rôle supposé spécifique aux textes scientifiques, et central dans le genre de l'article, mais en faisant l'hypothèse qu'il peut cependant faire l'objet d'usages disciplinairement marqués.

5.3. Le lexique transdisciplinaire

Il semble relativement bien admis aujourd'hui que le lexique d'une langue est à aborder « entre langue et discours », selon le titre de l'ouvrage de M.F. Mortureux (1997) consacré à la lexicologie. L'idée que le système lexical se constitue dans ses usages est ainsi exploitée dans nombre d'entreprises dictionnaires, qui s'appuient désormais sur les techniques informatiques de la linguistique de corpus pour améliorer leurs descriptions des mots en contexte, et fournir des données sur les concordances et les collocations.

Le parti pris qui caractérise la présente étude appelle à souligner que l'étude du lexique doit gagner à s'intéresser plus spécifiquement aux genres, en tant qu'instance de normalisation de la langue (Rastier, 2001 : 272) qui permet d'envisager de manière intégrative l'existence d'un système linguistique et la diversité des usages sociaux des textes. Le genre doit ainsi permettre d'analyser le lexique en référence à une phraséologie propre et dans une dimension à la fois syntaxique, sémantique et pragmatique. En retour, comme le lexique varie avec le genre, la caractérisation d'un genre peut prendre comme objet ses aspects lexicaux.

S'agissant de caractériser le genre de l'article de recherche, nous nous intéressons *a priori* à l'ensemble des traits textuels, tout en privilégiant ceux dont l'analyse elle-même suggère la pertinence. Par ailleurs, il nous semble important de mettre l'accent sur ce qui, dans les textes montre en quoi ils sont des faits sociaux et praxéologiques. C'est ainsi que nous avons écarté une étude du lexique spécialisé dans l'article, pour nous centrer sur un

premier temps les emplois autonymiques exclusifs (sans valeur d'usage) des emplois combinant usage et mention.

¹ Comme nous l'avons dit *supra*, on doit la mise en évidence de ces différents types de commentaires méta-énonciatifs à l'apport capital de J. Authier-Revuz (1995) sur ces phénomènes.

aspect du lexique supposé spécifique à l'activité de recherche, et que nous appellerons le lexique transdisciplinaire.

Caractéristique du discours scientifique, ce lexique s'impose pour bâtir dans les textes l'univers propre à l'activité de recherche ; il en pointe les objets et le rapport au réel qui la définit (fait de *notions, unités, principes, fonctions*, etc.), et identifie les opérations cognitives et langagières qui fondent la praxéologie de la recherche (*ses postulats, hypothèses, méthodes, descriptions, exemples*) en tant que champ d'activité visant la production du savoir ; il sert en cela une définition socialement partagée des composantes et ses procédures de cette activité, qu'il s'agisse de présenter la problématique, la méthodologie adoptée, ou de rendre compte des observations faites et les interpréter. Par l'usage de ce lexique, le texte se constitue ainsi en tant que texte scientifique.

Dans ce troisième volet de notre étude morpho-syntaxique et lexicale du genre, l'objectif est donc de décrire l'usage effectif de ce lexique transdisciplinaire en tant que caractéristique attendue de l'article, et de voir dans quelle mesure il est possible de parler de caractéristique transdisciplinaire, ou de caractéristique disciplinairement marquée.

5.3.1. *Un lexique au service de la construction du savoir*

La définition de ce lexique ne va pas sans poser problème, et la seule difficulté au niveau de sa désignation en atteste. En première approche, il est possible de parler de lexique scientifique, pour rendre compte de la spécificité du champ de pratiques concerné, i.e. le « discours » scientifique selon la terminologie de F. Rastier (2002)¹. Le problème tient au lexique de spécialité caractéristique du discours scientifique : nous cherchons au contraire à caractériser l'article en optant pour une entrée transdisciplinaire, qui permette de faire abstraction, au moins dans un premier temps, des spécificités thématiques des textes considérés. Secondairement, la pertinence de cette désignation de lexique scientifique est mise en péril par le fait que le terme « scientifique » renvoie à un mode de construction des savoirs spécifié en fonction des pratiques propres à certaines disciplines ; en ce sens, il se prête peut-être moins bien que celui d'« activité de recherche » aux deux disciplines ici concernées, et notamment aux articles de LET comme cette étude permettra précisément de le montrer.

Le fait de rattacher ce lexique à la construction du savoir qui caractérise l'écrit de recherche a induit dans un premier temps le choix de la désignation « lexique épistémique ». Cependant là encore, le choix du qualificatif « épistémique » reste peu satisfaisant : il introduit un parasitage par rapport à d'autres emplois du terme en linguistique, en particulier avec la notion de modalité épistémique en linguistique de l'énonciation.

Dans ce champ pluriel et foisonnant, la définition des modalités épistémiques est loin d'être stabilisée. En première approche, un terme épistémique est à comprendre comme un « subjectivème » (Kerbrat-Orecchioni, 1980), c'est-à-dire un lexème porteur d'informations quant à l'attitude du locuteur par rapport à son dit ; plus spécifiquement, dans la catégorisation des subjectivèmes de C. Kerbrat-Orecchioni (op. cité) comme dans les analyses plus traditionnelles², qui désormais constituent en quelque sorte le socle commun de la linguistique de l'énonciation, les termes épistémiques sont ceux qui ont trait au statut épistémique du contenu asserté, autrement dit à son statut en tant que savoir : le contenu asserté peut se présenter selon des degrés divers entre le probable et le certain, et entre

¹ Cf. la discussion des catégories de texte et discours dans le chapitre 1, et la présentation du cadre d'analyse proposé par F. Rastier (2002).

² On doit ces analyses traditionnelles héritées en particulier de Thomas d'Aquin à la logique modale (Cervoni, 1992, ed. originale 1987).

l'entendu et le constaté¹. Quoique les discussions restent largement ouvertes sur cette catégorie complexe des modalités épistémiques, elles relèveraient donc, selon C. Kerbrat-Orecchioni (op. cité), d'une évaluation du locuteur quant à la vérité et la certitude de l'assertion.

Cela étant, il faut souligner que, dans une approche polyphonique², le statut de vrai, faux, certain et incertain ne peut être restreint au seul jugement du locuteur, comme le montre par exemple le conditionnel : le problème des degrés de prise en charge de l'assertion devient central, et c'est sans doute alors dans le jeu de dédoublement et de hiérarchisation des points de vue plutôt que dans les seuls subjectivèmes qu'il faut chercher ce jugement interne à l'assertion, et indexable au locuteur.

Parce qu'il sert la construction du savoir, le lexique transdisciplinaire peut dans une certaine mesure être rapproché de cette catégorie des modalités épistémiques : parler d'un *postulat* n'a pas le même statut épistémique – en termes énonciatifs – que parler d'une *hypothèse*, et tous deux se différencient de l'annonce d'un *résultat*. Ces trois termes peuvent être retenus au titre de lexique propre à l'activité de recherche, et posent la question modale du statut de vérité et du degré de prise en charge de l'assertion.

Certains des termes transdisciplinaires seraient donc bien épistémiques en ce qu'ils signalent des statuts divers en termes de construction du savoir, qui permettent de distinguer ce qui est supposé, postulé, constaté, vérifié, etc. de manière lexicalement codifiée au sein de l'activité de recherche. Se manifeste alors l'intérêt de ces aspects énonciatifs pour l'étude de l'argumentation et de la construction dans les textes de la validation du savoir : de même que pour les points de vue introduits par l'auteur au titre d'*hypothèse* ou de *résultat*, les points de vue présentés comme le fait d'une *approche* ou d'une *théorie* n'ont pas la même valeur probatoire.

Malgré l'intérêt qu'il y aurait à approfondir une définition énonciative du lexique transdisciplinaire³ en regard de la notion de modalité épistémique, il nous semble que systématiser un tel rapprochement reste délicat, notamment dans le cas de termes assez généraux retenus au titre du lexique transdisciplinaire, du type *question*, *problème*, *structure*, *fonction*, etc. Notre analyse nous donnera cependant l'occasion d'évoquer le statut épistémique de certains termes et de discuter les critères utilisés pour discriminer ce lexique propre à l'activité de recherche ; auparavant, nous évoquerons rapidement en quoi il intéresse la description de l'article.

5.3.2. Un lexique entre langue générale et langue de spécialité

Le lexique est vraisemblablement l'aspect du discours scientifique le plus thématisé, en particulier dans les champs de l'« English for Academic Purposes » et de l'« English for Specific Purposes », qu'il s'agisse de rendre compte de recherches sur la question, ou de proposer un manuel à vocation didactique. Nous ne passerons pas en revue cette abondante littérature, qui autorise cependant à faire le constat que la question est bien moins traitée en ce

¹ J. Cervoni (1992 : 76) rend compte en des termes proches d'une analyse selon quatre pôles (certain, probable, exclu et contestable).

² C'est ainsi que R. Vion (2005) propose de distinguer les modalisations des modalités sur la base du dédoublement énonciatif (ou polyphonie) caractéristique des premières, qui consistent selon lui en un commentaire réflexif sur le dit dont relèvent les secondes ; il montre ainsi que dans *Pierre viendra demain*, la modalité contribue directement au sémantisme du dit, sémantisme opacifié par la modalisation introduite par l'adverbe dans *Pierre viendra certainement demain*.

³ Des pistes intéressantes sont déjà signalées dans le numéro 102 de *Langue Française* consacré aux *Sources du savoir et leurs marques linguistiques* (Dendale et Tasmowski, 1994).

qui concerne le français¹, ce que le caractère international de l'anglais comme langue véhiculaire dans la communication scientifique justifie aisément.

Plus spécifiquement, le lexique transdisciplinaire sur lequel nous proposons de nous centrer ne va pas sans rappeler les travaux sur le *Vocabulaire Général d'orientation Scientifique*, ou VGOS (Phal, 1971)². Dans la continuité de ce qui avait été fait par G. Gougenheim et son équipe sur le français fondamental (1956 et 1964), ils visent à décrire le vocabulaire du français sur la base de la fréquence de ses usages. Le VGOS repose sur le constat de difficultés des étrangers qui étudient en France les sciences et les techniques, difficultés qui ne concernent pas la communication ordinaire de la vie quotidienne, ni les concepts spécifiques aux domaines de spécialité, mais l'expression scientifique. Le VGOS a donc pour objet d'inventorier ce vocabulaire intermédiaire entre le vocabulaire usuel et les lexiques de spécialité, vocabulaire intermédiaire dédié à l'expression scientifique.

L'introduction d'un niveau intermédiaire montre la difficulté qu'il y a à définir autrement ce lexique que par rapport aux deux catégories du lexique général et du lexique de spécialité. Quoiqu'elle amène les auteurs à discuter le bien-fondé de cette bi-partition lexique général/ lexique de spécialité, le VGOS serait selon eux inclus dans le vocabulaire général, et caractéristique des textes scientifiques quel que soit leur domaine de spécialité. Mis en évidence selon des critères de fréquence et de répartition dans des textes scientifiques de genre et de domaines divers, le VGOS reste cependant marqué par des spécificités disciplinaires : on y trouve en effet des termes comme *tangente*, *température*, *trapèze* ou encore *unité de longueur*. Il est vrai que ces termes sont bien transdisciplinaires, et que leur emploi déborde aussi la seule expression scientifique. Cependant, au regard de certaines disciplines, en particulier dans les sciences humaines et sociales, ces termes apparaissent, à côté de termes plus généraux comme *hypothèse* par exemple, comme un vocabulaire nettement spécifique aux disciplines considérées.

De loin bien plus modeste dans ses objectifs et ses méthodes que les travaux sur le VGOS, notre étude se fonde sur la même idée d'un lexique relevant du lexique dit « général », mais présentant un usage spécifique à l'activité de recherche ; autrement dit, dans notre optique, ce lexique est supposé commun à l'ensemble des genres de textes en usage dans ce champ de pratiques (ou dans ce « discours », selon la terminologie de F. Rastier), et en cela transversal à différents domaines de spécialité.

Nous proposons de caractériser le genre à partir de cette entrée, tout en faisant l'hypothèse qu'en vertu d'une identité disciplinaire du genre, chaque discipline fait un usage spécifique³ de ce lexique transdisciplinaire : ses usages, en termes quantitatifs et qualitatifs varieraient en partie d'une discipline à l'autre. Contrairement au VGOS, il ne s'agit donc pas de dresser un inventaire exhaustif de ce lexique mais plutôt de sonder l'usage de quelques termes du double point de vue de leur fréquence et de leurs concordances dans les articles de SCL et de LET.

Nous analyserons d'abord la distribution dans le corpus d'articles d'un certain nombre de termes sélectionnés en référence à d'autres genres et domaines du discours scientifique et au corpus lui-même comme nous allons le préciser dans la section suivante. Nous

¹ Citons cependant à titre d'exemple, et de manière bien entendu non exhaustive le numéro 17 de *Langue Française* consacré aux Vocabulaires techniques et scientifiques (Guilbert et Peytard, 1973), l'ouvrage bien connu de Kocourek (1982), *La langue française de la technique et de la science*, et les travaux actuels de A. Tutin (2005) et de O. Kraif et A. Tutin (2006) sur les collocations et la phraséologie scientifique en français.

² Les travaux actuels sur l'anglais qui se rapprochent le plus de ce projet-là sont sans doute ceux menés par A. Coxhead sur l'« Academic Word List » (par ex., 2000, 2002).

³ I.e. disciplinairement marqué.

compléterons l'analyse quantitative par un examen des occurrences en contexte, de manière à comparer les usages dans les deux disciplines en discutant notamment des frontières entre emplois généraux, spécialisés, et transdisciplinaires. Poursuivant notre objectif de description des facteurs de variabilité du genre, nous chercherons également à voir si le statut de l'auteur a un effet sur ces usages lexicaux.

5.3.3. Constitution d'une liste de termes transdisciplinaires

Le point de départ de cette étude du lexique transdisciplinaire a été de constituer une liste de lemmes, avec deux exigences :

- se baser sur des corpus de textes scientifiques de genres et de disciplines variés pour déterminer ce qui dans leur lexique est propre à l'activité de recherche et présente un caractère transdisciplinaire ;
- retenir des termes pertinents – au regard de la spécificité des textes et des disciplines représentées dans notre corpus – en termes de fréquence d'une part et de qualité « transdisciplinaire » par rapport aux usages généraux et spécialisés des mêmes unités formelles (nous en donnerons un exemple ci-après avec le terme de *mesure*, non retenu en raison de son usage principalement général dans le corpus).

Dans un premier temps, nous avons eu recours à une liste de 47 substantifs (lemmatisés) transdisciplinaires, établie par G. Bloquet dans le cadre d'une étude des collocations nom-verbe dans les écrits scientifiques (2004, cf. aussi Bloquet et Tutin, 2004). Cette liste a été constituée sur la base d'un critère de fréquence, auquel s'est ajoutée également une sélection manuelle des termes en contexte en fonction d'un critère sémantique, lié à la qualité transdisciplinaire de l'emploi du terme par rapport à des emplois généraux ou spécialisés.

Le recours à cette liste permet d'avoir des données sur d'autres disciplines et d'autres genres du discours scientifique : le corpus utilisé¹ par G. Bloquet (*ibid.*) présente en effet une hétérogénéité tout à fait intéressante de ce point de vue-là : il réunit des articles, des rapports et des thèses des domaines du traitement automatique du langage et de la médecine principalement, mais aussi du traitement de l'information, du traitement de l'image, de la génétique, de la linguistique et de l'épidémiologie. Quoique encore limitative, cette diversité de domaines et de genres fournit des données externes à notre corpus, et nécessaires pour l'étude de ce lexique propre au discours scientifique pris dans son ensemble.

Nous avons alors procédé à un examen du lexique en usage dans notre corpus, en vue d'adapter la liste de termes transdisciplinaires à ses spécificités. Deux critères de sélection nous paraissaient en effet devoir être respectés :

- d'une part, à partir de la quantification de l'ensemble des lemmes du corpus à l'aide du logiciel d'analyse, un critère quantitatif de fréquence et de répartition des termes a été appliqué, pour éviter de retenir de la liste initiale des termes trop peu représentés (par exemple, *base*, *test* et *tester* ont ainsi été écartés), et pour ajouter à cette liste certains termes (en particulier, il nous a semblé utile d'adjoindre aux substantifs de la liste de G. Bloquet quelques verbes et adjectifs) ;
- d'autre part, les occurrences ont fait l'objet d'un examen en contexte en vue de croiser le critère quantitatif avec un critère d'ordre sémantique : celui-ci consiste à sélectionner des termes dont l'usage en corpus répond à la définition du lexique transdisciplinaire. C'est ainsi

¹ Il s'agit du corpus du projet Emergence, projet du Lidilem piloté par O. Kraif et A. Tutin (2004) et consacré à *l'Acquisition semi-automatique de traductions d'expressions semi-figées pour l'élaboration d'outils d'aide à la rédaction scientifique et technique*.

que nous avons complété la liste initiale, mais également supprimé certains termes, comme par exemple les substantifs *modèle* et *mesure* qui présentent un usage essentiellement général : la majorité des occurrences du lemme *mesure* dans notre corpus n'a pas le sens d'évaluation quantitative (comme dans *mesure physique*), mais plutôt celui de pondération, ou encore celui de moyen (comme dans *prendre la mesure de*).

Dans le même sens, certains termes pourtant fréquents n'ont pas été retenus car leurs emplois dans notre corpus s'avèrent trop spécifiquement liés aux objets propres aux deux disciplines considérées : c'est le cas de *figure*, *règle* ou encore *objet* (par ex. dans *objet direct/indirect* dans les articles de SCL).

Enfin, il convient de préciser qu'un dernier critère, de type lexicographique, a été utilisé : nous avons parfois préféré des termes de la même famille (noms, verbes, adjectifs) à la diversité des termes représentés ; ainsi avons-nous retenu *traitement* avec *traiter*, ou encore, l'adjectif *méthodologique*, bien qu'il soit peu fréquent, avec *méthode* et *méthodologie*, de même que pour les autres adjectifs. Quoique tout à fait discutable tant du point de vue des termes retenus que de ceux qui ont été écartés, la liste qui contient au final 100 lemmes nous semble globalement représentative de ce lexique transdisciplinaire, et s'avère relativement opératoire dans le cadre de cette approche exploratoire de l'article.

Le tableau ci-dessous liste les cent lemmes retenus ; pour chaque lemme est précisée sa catégorie (A pour adjectif, N pour nom, V pour verbe) ; certains lemmes ont en effet une forme ambiguë : pour *savoir*, c'est le verbe que nous avons retenu, non le nom, pour *problématique*, c'est le nom, etc. Au total, on compte 6 adjectifs, 62 noms et 32 verbes.

analyse N	conclusion N	discussion N	investigation N	problème N	tableau N
analyser V	condition N	donnée N	matériau N	procéder V	taxinomie N
appeler V	conséquence N	étude N	méthode N	proposer V	tendance N
approche N	considérer V	étudier V	méthodologie N	propriété N	terme N
argument N	constat N	examen N	méthodologique A	question N	théorie N
caractérisation N	constater V	examiner V	montrer V	questionnement N	théorique A
catégorisation N	constituer V	exemple N	notion N	recherche N	traitement N
choisir V	critère N	expérimental A	objectif N	réflexion N	traiter V
choix N	décrire V	expérimentation N	observable A	relever V	travail N
classification N	définir V	explication N	observation N	représenter V	typologie N
commentaire N	définition N	expliquer V	observer V	résultat N	unité N
comparaison N	démarche N	fonction N	paradigme N	savoir V	valeur N
comprendre V	démonstration N	hypothèse N	phénomène N	schéma N	vérifier V
concept N	démontrer V	idée N	présenter V	signifier V	viser V
conception N	description N	impliquer V	présupposé N	systematique A	voir V
conceptuel A	désigner V	indiquer V	principe N	systematiser V	
concevoir V	déterminer V	interprétation N	problématique N	systeme N	

Figure 5.13. Liste des 100 termes transdisciplinaires retenus.

Le comptage des occurrences des lemmes est effectué à l'aide de Cordial Analyseur sur les 220 articles du corpus, dans le corps de l'article (i.e. l'article à l'exclusion des notes,

annexes et bibliographies). La base de données ainsi constituée adjoint à ces 100 lemmes la variable « tous termes confondus » pour étudier le poids global dans les textes des termes retenus. L'analyse statistique décrit les distributions des valeurs et teste la significativité de la différence entre les textes des deux disciplines (LET et SCL), et en fonction du statut de l'auteur (doctorant et non doctorant). Un concordancier permet alors l'examen des occurrences en contexte, et la mise en évidence de co-occurents en fonction de leur nombre et de leur position relative au lemme transdisciplinaire.

Nous présenterons d'abord les résultats liés à notre hypothèse principale, celle d'un usage disciplinaire du lexique transdisciplinaire, en nous centrant sur les aspects quantitatifs, pour discuter ensuite l'emploi de certains termes. Pour finir, nous nous intéresserons, plus brièvement, à la question de savoir si l'usage de ce lexique varie avec le statut de l'auteur.

5.3.4. Un plus grand nombre d'occurrences tous termes confondus en SCL

L'analyse de la fréquence et de la répartition des lemmes transdisciplinaires dans le corpus fait apparaître que l'usage de ces termes se distingue dans les deux disciplines, quant au nombre d'occurrences par articles tous termes confondus et quant au nombre d'occurrences de certains termes. Nous verrons par la suite en quoi ces usages se distinguent également en termes de concordances.

Bien que les occurrences de l'ensemble des termes transdisciplinaire se répartissent dans chacun des articles du corpus, de grandes différences entre articles se manifestent : le nombre moyen de termes retenus comme transdisciplinaires est de 154,3 occurrences, mais on en trouve 501 dans l'article en contenant le maximum, et 11 dans l'article qui en contient le minimum. Comme ces différences peuvent s'expliquer par la variabilité au niveau de la longueur des textes, il convient d'étudier la fréquence relative des termes. Celle-ci atteste d'une diversité effective dans l'usage du lexique transdisciplinaire ; une fois le ratio sur le nombre de mots du texte ramené à un article moyen de 5000 mots, le minimum enregistré est de 7,7 occurrences et le maximum de 328,5, pour une moyenne de 151,4 occurrences par article.

L'usage du lexique transdisciplinaire, tous termes confondus, atteste d'une différence entre les articles des deux disciplines : le nombre d'occurrences est bien plus élevé en SCL, avec une moyenne de 186,6 termes par article contre 116,2 en LET. C'est de cette supériorité globale de l'usage des termes transdisciplinaires en SCL dont rend compte le graphique suivant, qui représente la distribution en fréquence du nombre d'occurrences tous termes confondus dans chaque discipline.

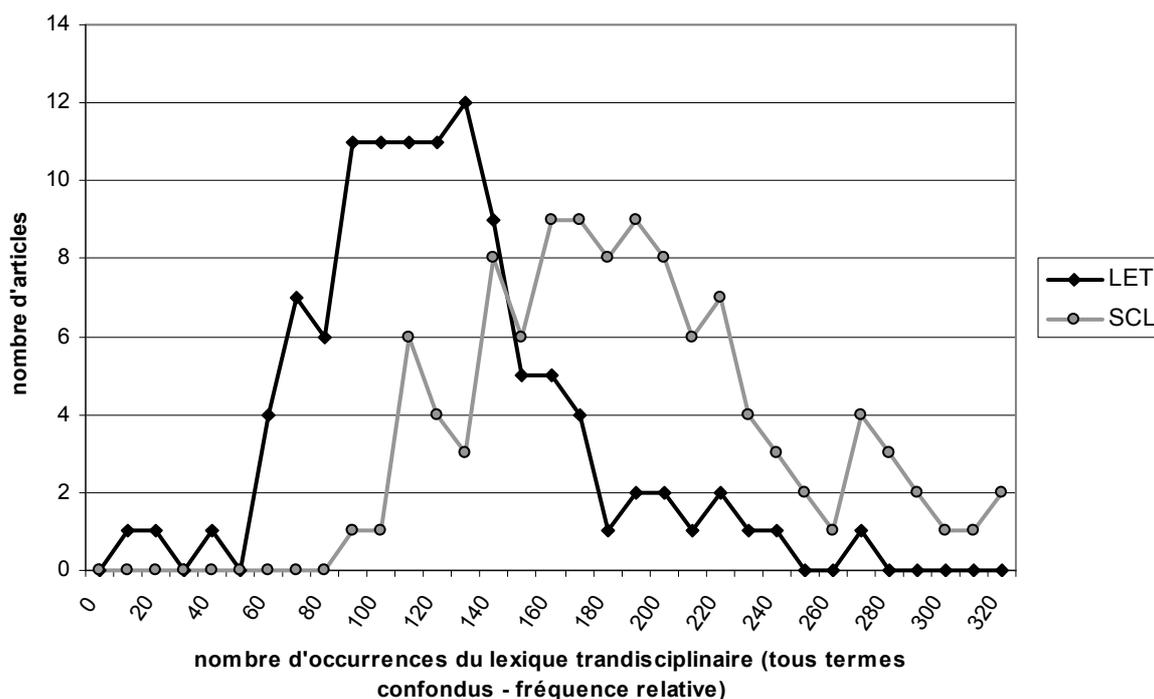


Figure 5.14. Distribution en fréquence du nombre de termes transdisciplinaires par article (tous termes confondus) : comparaison disciplinaire.

Comme le montre le graphique, le nombre minimum de termes transdisciplinaires (en fréquence relative) est le fait d'un article de LET, et le nombre maximum celui d'un article de SCL. En termes de tendances centrales, une opposition se manifeste entre un usage abondant de ces termes, caractéristique des articles de SCL et un usage moindre, caractéristique des articles de LET. Cela étant, les distributions dans chaque discipline montrent également que ce lexique est partagé, qu'un certain nombre d'articles de LET et de SCL présentent à ce niveau un profil similaire, et que quelques articles de LET dépassent les valeurs enregistrées pour certains articles de SCL.

La différence s'avère cependant statistiquement significative¹ : conformément à notre hypothèse, la discipline a un effet déterminant sur le nombre d'occurrences des termes retenus au titre du lexique transdisciplinaire. La différence affecte les indices de tendance centrale, plus élevés en SCL qu'en LET, mais également les indices de dispersion qui indiquent une variabilité sensiblement plus importante entre les articles de SCL, déjà observée à d'autres niveaux du texte. Cette différence disciplinaire observable dans l'usage global du lexique transdisciplinaire tous termes confondus se vérifie comme nous allons le voir pour l'emploi d'un nombre important de termes.

5.3.5. Des termes inégalement employés en LET et en SCL

Si l'on s'intéresse au nombre d'occurrences terme par terme, on note de nouveau une variabilité entre articles – ne serait-ce que parce qu'il existe pour chaque terme un article dont ce terme est absent. Soulignons aussi l'importante variabilité en fonction des termes retenus, y compris quand on les étudie en fréquence relative ; certains sont très rares (*systematiser* est représenté 12 fois dans l'ensemble du corpus, soit 10,7 fois en fréquence relative, *methodologie* 20 fois avec une fréquence relative de 19,5) tandis que *voir* et *exemple* totalisent respectivement 1838 et 1888 occurrences, soit 1828 et 1915,5 en fréquence relative.

¹ *F* de Fisher = 108,584, *p* < 0001.

Pour la grande majorité des termes, le nombre d'occurrences est plus élevé en SCL qu'en LET¹, et les analyses statistiques permettent de conclure que les 45 termes indiqués ci-dessous y sont significativement plus fréquents.

analyse	constater	exemple	observation	recherche
analyser	constituer	fonction	observer	relever
approche	critère	hypothèse	phénomène	résultat
catégorisation	définir	impliquer	présenter	système
concept	déterminer	indiquer	problématique	tendance
conceptuel	discussion	interprétation	problème	typologie
conclusion	donnée	montrer	proposer	unité
condition	étude	objectif	propriété	vérifier
considérer	examiner	observable	question	viser

Figure 5.15. Liste des 45 termes significativement plus fréquents dans les articles de SCL.

Dans les articles d'études littéraires, qui se caractérisent négativement par une présence plus faible des termes ci-dessus, seuls trois termes sont significativement plus fréquents : « commentaire », « idée » et réflexion ».

La discipline est ainsi un facteur de différenciation de l'usage de près de la moitié des termes de la liste. En SCL, la plus grande importance du nombre moyen d'occurrences par article par rapport aux LET va de pair avec des indices de dispersion plus élevés. Cette variabilité semble pouvoir être liée au fait que les termes retenus peuvent présenter quelques emplois non proprement transdisciplinaires, en particulier un usage spécialisé, d'ordre conceptuel ; la forte présence d'un terme s'expliquerait alors par des spécificités thématiques des articles², mais elle est valable aussi pour des termes plus généraux, tels que *étude*, *question*, ou encore *montrer*.

La variabilité intra-disciplinaire doit ainsi pouvoir être rapportée, à l'instar de la variabilité entre les deux disciplines, à des identités disciplinaires plurielles, y compris au sein d'un même champ. Il est possible également qu'elle s'explique par des préférences d'ordre idiomatiques comme nous chercherons à le voir dans la partie consacrée à l'étude des styles d'auteur.

Les différences observées quantitativement appellent à analyser les emplois en contexte ; nous ne nous intéresserons ici qu'à quelques termes, dont les occurrences permettent de discuter la diversité des emplois (i.e. les emplois transdisciplinaires par rapport aux emplois généraux ou spécialisés des mêmes unités formelles), et de mettre en évidence le caractère disciplinairement marqué de certains usages.

¹ 83 termes sur 100 sont plus fréquents en SCL, contre 17 en LET (nombre total de termes par discipline en fréquence relative).

² Ce peut être le cas pour le terme *exemple*, ou pour *interprétation* dont on trouve 92 occurrences dans un article sur l'interprétation.

5.3.6. Concordances des termes transdisciplinaires : de quelques traits communs et des spécificités disciplinaires

Sans entrer dans le détail des différentes concordances de chacun des termes, nous formulerons ici quelques remarques quant à la diversité des emplois, pour dégager ensuite des tendances dans chaque discipline.

Les distinctions que nous proposons pour rendre compte des usages observés, loin d'être applicables à tous les termes ni à tous les contextes, sont à envisager comme des outils de travail, et vont permettre de compléter l'analyse quantitative des occurrences. Il s'agit de s'intéresser à des caractéristiques phraséologiques, à partir des co-occurents lexicaux ou grammaticaux des termes transdisciplinaires.

Nous nous en sommes tenus à l'analyse des concordances ; faute d'un nombre suffisant d'occurrences, il ne s'est pas avéré possible en effet de déterminer statistiquement des collocations, sauf à entendre ces dernières dans une acception large¹ qui inclut les constructions syntaxiques privilégiées ; on repère en effet certains articles et certains pronoms en contexte droit ou gauche des substantifs. Du reste, les verbes qui ont l'avantage d'être particulièrement représentés, comme *être* ou *avoir*, ont également un statut grammatical plutôt que lexical, et ne sont sans doute que peu productifs pour l'étude d'un genre en particulier.

De fait, sans établir des profils collocatifs, nous avons cependant quantifié en certains cas les concordances, et utilisé en particulier un critère de fréquence pour corroborer les distinctions proposées quant aux usages des termes transdisciplinaires, à l'aide notamment des articles et pronoms.

La diversité des emplois : les usages spécifique, général et transdisciplinaire

Comme nous l'avons précisé dans l'introduction à cette étude, nous avons pris la précaution en constituant la liste de termes transdisciplinaires d'écarter ceux dont la majorité des emplois correspondraient à des usages spécialisés. Certains des termes retenus présentent cependant, à côté d'un emploi principalement transdisciplinaire, des emplois en tant que concept ; d'autres à l'inverse présentent quelques emplois d'ordre général ; dans les deux cas, il nous paraît important de les signaler pour montrer la difficulté posée par les frontières entre ces catégories d'usages lexicaux, et mieux définir ainsi la spécificité des usages proprement transdisciplinaires.

Un usage conceptuel

Les termes retenus comme transdisciplinaires présentent pour certains un usage conceptuel dès lors qu'ils ne renvoient pas à des composantes et des procédures de l'activité de recherche, mais à un appareillage terminologique spécialisé.

¹ A. Tutin et F. Grossmann montrent en effet que la tendance actuelle consiste à limiter les collocations aux associations lexicales, mais qu'une acception large, attestée depuis 1950 environ, demeure, en particulier chez les fonctionnalistes anglais (par ex. dans Sinclair, 1991).

Ainsi *expérimental* est-il quelquefois utilisé en LET quand il est question du *roman expérimental* ; de même, *démonstration* est employé dans un des articles de SCL en tant que concept désignant une forme de citation :

[Clark et Gerrig] conçoivent les **démonstrations** comme nécessairement fondées sur une relation de ressemblance perceptuelle, c'est-à-dire matérielle, entre le support et l'objet de la démonstration, et sont ainsi conduits à opposer radicalement les procédés du style direct et du style indirect [SCL029]

La méthode de constitution de la liste des termes est telle que les quelques occurrences au statut de concept restent rares dans le corpus, mais celles-ci peuvent vraisemblablement renforcer ou nuancer en partie les différences observées dans l'usage exclusivement quantitatif de ces termes.

Il est toutefois intéressant de souligner que parfois, l'usage conceptuel, plus spécifiquement lié au rattachement disciplinaire (voire sous-disciplinaire ou domaniale) et l'emploi transdisciplinaire ne sont pas dissociables ; c'est le cas pour le terme de *système*, dont le statut conceptuel hérité de la linguistique structurale (le *système linguistique*, *phonologique*, etc.) se confond avec un usage transdisciplinaire au sens d'« ensemble structuré » et reste proche de l'usage général.

L'usage transdisciplinaire comme spécification de l'usage général

Si l'on met de côté les usages conceptuels et en cela spécialisés de certains des termes transdisciplinaires, il reste qu'un certain nombre de termes présentent un usage transdisciplinaire d'une part, et plus général d'autre part. Ainsi, pour le terme *démonstration* par exemple, il peut être question de démonstrations propres à l'activité de recherche – qu'il s'agisse de la recherche présentée dans l'article ou de références à d'autres travaux – mais également de la démonstration en tant qu'opération intellectuelle ou acte de parole non réservé aux chercheurs. C'est ainsi qu'en SCL, outre l'emploi conceptuel de *démonstration* cité ci-dessus, il est possible de distinguer les occurrences du terme dans les deux extraits suivants :

Dès lors, puisque cette épreuve académique est conçue en vue d'une démonstration de compétences à évaluer, il ne faut plus s'étonner alors que la discipline elle-même, la stylistique, soit définie par ses « contacts » avec les disciplines qui fournissent les compétences en question. [SCL047]

Il existe des **démonstrations** empiriques qu'une relation de dépendance inverse est possible [SCL087]

Quoique très proches sémantiquement, la différence entre un usage transdisciplinaire et un usage non spécifiquement transdisciplinaire se manifeste dans les concordances de ces termes. C'est ce dont témoignent par exemple les contextes où est employé *hypothèse* dans les extraits suivants, tirés d'articles de LET :

Ainsi, le récit représente les effets de l'imagination comme faculté, sa nature, ses fonctions, ses pouvoirs ; en même temps, il expose diverses fictions du monde, des **hypothèses** imaginaires sur la forme du monde. [LIT055]

À l'extrême limite, toutes les **hypothèses** présentées pourraient n'être que pure plaisanterie, et les égarements de l'imagination pourraient révéler la vanité de toute philosophie ; [LIT055]

L'écriture étant pour eux l'espace de liberté où les esprits se décroissent, où les charmes des **hypothèses** permettent des audaces inconcevables ailleurs. [LIT022]

Dans notre corpus, les hypothèses de recherche ne sont jamais qualifiées d'« imaginaires » ni de « pure plaisanterie » ; c'est aussi ce que montre l'usage de *vérifier*, co-occurent relativement fréquent d' *hypothèse* de recherche dans les articles de LET comme dans ceux de SCL, mis entre guillemets dans l'extrait suivant :

Ainsi, le Narrateur « **vérifie** » en s'élevant vers le Soleil **les hypothèses de la science récente** : « Je connus très distinctement, comme autrefois j'avais soupçonné en montant à la Lune, qu'en effet c'est la Terre qui tourne [...]. [LIT055]

Les co-occurents d'*hypothèse* justifient de le considérer ici comme transdisciplinaire, mais la mise entre guillemets du terme, qui joue le rôle de commentaire méta-énonciatif portant sur la relation mot-chose¹ signale une non-adéquation pointée par l'auteur entre la vérification proprement scientifique d'une hypothèse et la construction littéraire des objets de connaissance.

Les constructions syntaxiques semblent également pouvoir intervenir dans le partage de ces usages généraux et transdisciplinaires : pour le terme *résultat*, quoique la distinction entre ces deux usages ne soit pas toujours possible², il semblerait que l'usage général tende à privilégier la reprise anaphorique sans déterminant (*X, résultat de...*), la complémentation par *de* ou *d'*, et le tour *être le résultat de*. Les résultats relatifs aux travaux de recherche apparaîtraient davantage précédés d'un démonstratif (*ce/ ces résultat(s)*), ou, dans une moindre mesure, de possessifs. Avec *les* et *le* cependant, on ne peut conclure à la supériorité d'un usage sur un autre, de même qu'avec *comme* et *pour* assez fréquemment aussi antéposés à « résultats ». Concernant le verbe *être*, l'abondance de son usage est davantage caractéristique du premier emploi que du second, où la spécification se manifeste par l'usage de verbes variés (avec *être*, *obtenir* surtout, *confirmer* et *montrer* ensuite).

Dans tous les cas il y a bien un sème commun aux deux usages général et transdisciplinaire d'un terme, mais, et c'est le critère que nous avons employé, il nous semble possible de voir dans l'usage transdisciplinaire un sème du type « propre à la recherche », qui viendrait s'adjoindre au premier pour apporter une spécification quant au domaine d'activité et au statut épistémique ainsi codifié. Valable pour de multiples autres lemmes, tels que *constat*, *observation* et *observer*, *objectif*, *discussion*, *démontrer*, *étude*, *expérimental* et *expérimentation*, *investigation*, *matériau*, *méthode*, *résultat*, *travail*, etc., cette distinction entre un usage général et un usage transdisciplinaire est à envisager comme un outil d'analyse, mais il convient de souligner encore une fois que les usages restent sémantiquement proches dans les deux types de contextes. Par ailleurs, il est des termes auxquels elle s'applique difficilement, l'emploi transdisciplinaire semblant se confondre avec un emploi général (comme avec *phénomène*, *système*, etc.).

Un usage transdisciplinaire propre aux articles de LET, dédié à l'œuvre littéraire, sa production et sa réception

L'examen des usages multiples, plus ou moins généraux ou plus ou moins spécialisés des termes retenus comme transdisciplinaires suggère une spécificité des articles de LET dans les

¹ Nous nous référons à l'analyse des commentaires méta-énonciatifs développée par J. Authier-Revuz (1995).

² Par exemple dans « les mesures de durée que nous présenterons plus loin sont le résultat d'une analyse acoustique sur Signalyze » [SCL077], la complémentation de *résultat* fait que l'usage général d'« effet » ou de « produit » se conjugue à celui de « produit spécifique d'un travail de recherche ».

emplois transdisciplinaires ; se manifeste en effet un usage indifférencié des termes transdisciplinaires pour traiter de travaux de recherche et des textes littéraires étudiés.

Il faut commencer par souligner que les termes transdisciplinaires servent par définition aussi bien à désigner le travail de recherche (ses opérations cognitives et langagières et ses produits, les travaux de recherche) que l'objet même de ces travaux : certains termes sont plutôt employés concernant l'objet (*principe, propriété, fonction, valeur, etc.*), les autres, plus nombreux, concernent les deux, et le plus souvent, il serait bien périlleux de chercher à les distinguer (pensons, par exemple, à *appeler*, ou à *exemple*).

Cela semble possible toutefois avec certains termes ou du moins certaines de leurs occurrences (et avec toutes les précautions que demande ce type de distinction, insistons sur ce point) ; en SCL par exemple, il peut être question de *démarche identitaire* et de *démarche empirique*, d'*objectif* poursuivi par le locuteur d'un énoncé et d'*objectif* poursuivi par l'auteur de l'article, de *conceptions* sous-jacentes à tel ou tel modèle théorique et des *conceptions de l'orthographe* que se font les enseignants. Le terme comme l'article ou l'extrait d'article où il est employé doivent référer à des textes et/ou à des sujets.

Cette condition est remplie en LET¹, où il est ainsi question d'*approche, objectif, démarche, description, recherche, conceptions, matériau, données, observations, hypothèses* de Montaigne (ou autre) ou encore d'*expérimentations* livrées par le texte littéraire, de ce qu'il *montre, démontre, implique* et *indique*, et la liste des termes retenus comme transdisciplinaires et ainsi mobilisés est loin d'être exhaustive. Nous remarquons ainsi qu'en LET cohabitent parfois jusqu'à trois usages différents : un usage transdisciplinaire, un usage plus général, et un usage réservé aux textes littéraires. C'est le cas par exemple pour le lemme *travail* :

Enfin préciser — et d'abord pour moi-même — la genèse de ce **travail**, son utilité pour la connaissance de Montaigne, mais aussi ses limites. [LIT021]

Le **travail** littéraire de Christine ne peut en effet être séparé de la demande de la cour pour laquelle elle écrit. [LIT029]

Pour Malot, comme pour Pestalozzi, le **travail** manuel présente un intérêt économique aussi bien qu'éducatif. [LIT091]

Cette nouvelle distinction que nous introduisons entre un usage réservé ou non aux textes littéraires se combine avec la question de savoir dans quelle mesure l'usage est plus ou moins général ou plus ou moins transdisciplinaire. Les occurrences de *travail* citées ci-dessus le montrent, comme celles de *démarche, objectif*, ou de *conceptions* évoquées plus haut, un continuum des usages se manifeste, mais qui nous semble permettre de distinguer :

1. un usage général d'un usage dit transdisciplinaire, relevant de cet usage général mais apportant, en contexte, une spécification du type « qui a trait à la recherche, son objet ou ses opérations »
2. un usage transdisciplinaire concernant le travail de recherche (opérations langagières, procédures) et un usage intermédiaire, qui sert à traiter de l'objet de la recherche ; cependant certains termes sont tels que ces deux usages se confondent systématiquement en un seul.

Ces deux oppositions permettent de souligner qu'il n'est pas anodin que les œuvres littéraires soient décrites dans les articles de LET au moyen des termes qu'on a retenus comme transdisciplinaires en tant qu'ils réfèrent à un travail de recherche. Ainsi, le lemme *étude*, à

¹ Du fait du matériau pris pour objet d'étude (Cf. dans le chapitre 4, la partie consacrée aux exemples et citations).

côté d'un usage général, figé au pluriel (*faire des études*), y est employé, au singulier ou au pluriel, pour désigner indifféremment un travail de recherche universitaire, l'« *étude de l'histoire antique de Montesquieu* » [LIT051], ou encore « *l'étude comparative [faite par Montaigne] des trois affections qui le concernent particulièrement : l'amitié, l'amour et les livres* » [LIT016]. De manière similaire est évoquée l'« *hypothèse expérimentale* » de Montesquieu dans le livre XVI de *L'Esprit des Loix* :

Que dans cette **hypothèse expérimentale**, ce soit un corps masculin que Montesquieu choisisse d'immerger dans un milieu « chaud et enfermé » ne doit assurément rien au hasard : la mise en valeur du facteur déterminant de l'air chaud et confiné n'en sera que plus probante. [LIT049]

Cette hypothèse de Montesquieu est évoquée de nouveau plus loin en ces termes :

Tout se passe comme si, par cette **expérimentation virtuelle**, Montesquieu invitait tacitement à la conclusion que l'effet de l'air chaud doit, *a fortiori*, être d'autant plus fort sur une femme, le postulat étant que les fibres du corps féminin sont, naturellement, plus relâchées que celles du corps masculin [LIT049]

La virtualité qui qualifie l'expérimentation menée par Montesquieu s'oppose ici à l'expérimentation matérielle actualisée dans certaines disciplines scientifiques. Il est clair cependant qu'il y a continuité et non rupture entre ces usages d'*expérimentation*, comme il peut y avoir continuité entre la sphère du littéraire et celle du scientifique – citons pour exemple les historiographes ou les auteurs d'essais – d'autant que leur définition mutuelle dépend pour beaucoup des époques.

Traiter du travail des auteurs littéraires en des termes transdisciplinaires revient à lui attribuer des propriétés communes sinon identiques avec le travail du chercheur tel qu'on en nomme les différents aspects ; c'est reconnaître que les opérations intellectuelles caractéristiques de l'un ne sont pas éloignées de celles de l'autre et du texte littéraire lui-même ; c'est montrer que les deux activités partagent un certain nombre de traits (des matériaux, des objectifs, des méthodes, des démonstrations, etc.) et représentent deux modes d'intellection somme toute relativement proches, qui poursuivent la même visée de connaissance du monde.

Tendance marquée à la systématisation de l'usage des termes transdisciplinaires en SCL versus en LET

A partir du nombre d'occurrences étudié *supra*, il est possible de dire que comme dans les articles de LET, les emplois de termes transdisciplinaires pour rendre compte de l'œuvre littéraire sont élevés, proportionnellement, le nombre d'occurrences strictement réservées à l'évocation d'un travail de recherche académique s'amointrit. Or, le nombre d'occurrences des termes transdisciplinaires doit déjà être relativisé au regard des usages qu'on a qualifiés de plus généraux. Ces derniers ne sont jamais majoritaires, puisque, rappelons-le, c'est un des critères qui a présidé à la constitution de la liste, mais il n'en reste pas moins que pour certains termes, *travail* notamment, ou encore *résultat*, l'usage général est loin d'être négligeable. Les articles de SCL, soumis à ce même constat sur les usages non spécifiquement transdisciplinaires, en présentent cependant un taux moindre, et sont par ailleurs, nous l'avons dit, moins concernés par un usage dédié aux sujets et aux textes étudiés.

Ainsi, pour les termes *expérimental* et *expérimentation*, plus de la moitié des occurrences dans les articles de LET concernent une expérimentation consistant à « faire l'expérience de », sans que l'expérience ait un statut scientifique, et se rapportent à l'œuvre

littéraire ou à l'essai, ainsi qu'à leur production et réception, comme on le voit dans les quelques occurrences citées ci-dessous :

Dans le mouvement moderniste le nationalisme était surtout une tranchée pour lutter en faveur de la liberté de création et de l'**expérimentation esthétique**. [LIT008]

Il s'agit d'un "essai" au sens strict du terme, **exercice de pensée sans frontières et sans bride, expérimentation ludique et rigoureuse** tout à la fois d'une liberté scabreuse [LIT009]

[...] pour désigner l'**attitude d'attente avide de celui qui se fait fête de s'ouvrir, ne serait-ce qu'à titre expérimental, à la variété des éléments d'un texte** qui, pour quelque raison que ce soit, auraient pu retenir son attention : ses redondances, ses accumulations, ses agrammaticalités, ses discontinuités, ses obscurités. [LIT020]

Si de telles audaces restent nécessairement limitées au sein d'un roman, il est d'autres lieux où Balzac a le loisir de **se livrer plus librement à une véritable expérimentation sur les limites des jeux possibles entre univers fictif et monde réel**. [LIT023]

L'analyse comparée des concordances d'*hypothèse* dans les deux disciplines corrobore ces observations faites sur les lemmes *expérimental* et *expérimentation*. Si les verbes ne paraissent pas être particulièrement différenciateurs, les noms et adjectifs suggèrent que l'usage transdisciplinaire s'associe en LET à d'autres usages : tandis qu'en SCL, les lemmes adjectivaux les plus fréquemment employés en co-occurrence d'*hypothèse* sont *linguistique, autre, sémantique, lexical* et *suivant*, en LET, à côté de *même* et *première*, on trouve *atomiste* mais aussi *magique*.

Dans le même sens, pour les noms co-occurents d'*hypothèse*, on trouve en SCL *segmentation, langue, type, question, exemple* – termes transdisciplinaires pour les trois derniers, en LET *imagination* et *monde* d'abord, *article, phénomène* et *corps* ensuite. Malgré des tendances communes fortes, les co-occurents font état d'un usage moindre des transdisciplinaires en LET, au profit d'usages plus généraux ou littéraires (par ex. *imagination, magique, monde*).

D'autre part, si l'on trouve des occurrences dites transdisciplinaires, comme *la physique expérimentale* ou *le protocole d'expérimentation anatomique*, celles-ci ne réfèrent point au travail de recherche présenté dans l'article lui-même, contrairement à ce qui apparaît dans les articles de SCL où l'expérimentation, quoique rarement sans doute, est une des méthodes d'observation en usage. Pour le terme de *méthode*, en dépit d'un grand nombre d'occurrences transdisciplinaires (*méthode historique, critique, exégétique, comparative*), seules quelques-unes (10 occurrences) ont trait à la recherche présentée dans l'article ; il en est de même pour *travail* (8 occurrences), ainsi que pour *résultat* (5 occurrences).

Ces données sur les différents usages du terme permettent de spécifier les résultats obtenus par la seule analyse du nombre d'occurrences dans les deux disciplines : s'il faut insister sur la continuité des usages des termes transdisciplinaires dans les deux disciplines, non seulement comme nous l'avons vu l'usage des termes transdisciplinaires y reste globalement beaucoup plus restreint qu'en SCL, mais il est restreint encore par un plus grand taux d'usages généraux et d'usages dédiés à l'objet de la recherche, et enfin, parce que les occurrences concernent très peu la recherche présentée dans l'article.

En SCL au contraire, l'usage de l'appareil terminologique transdisciplinaire pour désigner la recherche présentée dans l'article est particulièrement fréquent, à tel point que pour certains termes, cet usage se présente comme tendanciellement systématique.

Ainsi, pour le terme *hypothèse*, terme significativement plus fréquent en SCL rappelons-le, les occurrences transdisciplinaires ne sont concurrencées que par un emploi très

peu fréquent, de type général, relatif à une attitude langagière ou cognitive de sujets dont l'article étudie les pratiques langagières. Plus de la moitié des emplois concernent de fait la recherche présentée dans l'article. En témoignent les co-occurrences les plus fréquentes d'*hypothèse*, en particulier les pronoms personnels et possessifs se trouvant dans son entourage gauche ou droite, immédiat ou non ; pour les déterminants, à côté de *cette* et *des*, respectivement au premier et second rang en SCL (et inversement en LET), on trouve en SCL un grand nombre d'occurrences pour *notre* (immédiatement après *une* et *les*). Ces occurrences sont plus faibles en LET ; il en est de même pour les occurrences de *nous* dans un contexte gauche ou droite à 4 mots de distance. *On* arrive ensuite en SCL, alors qu'en LET, *il*, *son* et *sa* l'emportent sur *nous* et *nos*. *Mon* et *je* sont en SCL plus fréquents que *sa*, *son* ou *leur*, mais *il* et *elle* encore davantage, à équivalence de « on » en contexte droit.

L'étude de ces quelques aspects des concordances les plus fréquentes mériterait toutefois d'être poursuivie par des analyses statistiques sur des corpus plus larges, et la prise en compte conjointe des contextes droit et gauche de 1 à 5 mots d'intervalle ne saurait qu'en donner une idée très générale. Cependant, couplée aux analyses quantitatives, qui montrent en particulier que 35% des articles de SCL ne contiennent aucune occurrence du terme contre 71% des articles de LET, on voit se profiler une caractéristique forte des articles de SCL, selon laquelle le terme d'*hypothèse* est un de ces termes transdisciplinaires particulièrement prisé, que l'article appelle à convoquer de manière presque systématique pour évoquer le travail de recherche qui y est présenté, et ce au moyen d'une prise en charge énonciative souvent personnelle.

C'est peut-être aussi en ce sens que l'on peut interpréter alors dans l'extrait suivant le fait que l'auteur ajoute à sa supposition un commentaire entre parenthèses :

[...] son emploi à l'oral amène à supposer (**hypothèse** à vérifier d'une manière systématique) qu'il peut signaler une reprise du dire » [SCL025].

Il s'agit ici pour l'auteur de modaliser son assertion en désambiguïsant le verbe *supposer*, pour préciser que le procès équivaut à une hypothèse ; autrement dit, par l'usage du terme transdisciplinaire, l'auteur confère un statut épistémique à sa supposition, comme en atteste l'usage avec *hypothèse* de *vérifier* et *systématique*. Si l'énoncé pris isolément ne permet pas à lui seul de dire en quoi l'injonction « (*à vérifier*) » est prise en charge comme un objectif que se donne l'auteur, la précision entre parenthèses se présente comme appelée par le genre de l'article en SCL, comme une explicitation requise vis-à-vis du lecteur.

5.3.7. Bilan sur les usages disciplinaires des termes transdisciplinaires : des épistémologies plurielles

L'analyse des concordances permet ainsi d'affirmer l'importance dans l'article en SCL d'un appareil terminologique transdisciplinaire, qui comparativement reste somme toute peu utilisé en LET. Ces termes désignant sa propre recherche, ses composantes et ses procédures (*recherche*, *travail*, *étude*, *approche*, *hypothèse*, *matériau*, *donnée*, *résultat*, *méthode*, *methodologie*, *objectif* – et la liste pourrait encore être complétée) seraient autant de normes d'écriture du genre et éléments en cela essentiels de la définition de l'activité de recherche dans le champ des SCL.

Il est possible d'y voir l'effet des épistémologies propres à chaque discipline, le recours à l'expérimentation étant absent des LET, de même qu'il n'est le fait que de quelques rares sous-domaines des SCL. Mais ces normes d'écriture contribuent elles-mêmes à définir des épistémologies et des identités disciplinaires, en suggérant la manière dont elles se situent par rapport à d'autres champs scientifiques, et par rapport à des modèles socialement

dominants de la science ; les termes transdisciplinaires pourraient même en ce sens être envisagés comme des caractéristiques entrant dans une définition canonique de la science.

Par comparaison, les articles de LET se caractérisent d'abord par cette absence d'un usage aussi important (en nombre d'occurrences et en nombre de termes concernés) du lexique transdisciplinaire, et tout particulièrement pour évoquer la recherche présentée dans l'article. Leurs traits positifs sont que trois termes relativement généraux, *commentaire*, *idée* et *réflexion*, sont plus présents qu'en SCL où, pour le dire rapidement, on leur préférerait sans doute les termes de *problème*, *question*, *recherche* ou *analyse*, ou encore ceux de *questionnement* ou de *problématique*.

L'usage conjoint des termes transdisciplinaires pour traiter d'activités de recherche comme de l'œuvre littéraire, voire des activités de production et d'interprétation de ces œuvres pour certains termes est une autre caractéristique forte de l'écriture de l'article en LET, et de l'identité épistémologique des études littéraires – si tant est qu'on puisse parler d'une identité au singulier.

Comme nous l'avons vu déjà au chapitre 4 à propos de la structure du genre, les articles en LET manifestent une forme d'indépendance à l'égard des canons stéréotypiques de la scientificité. On le voit au nombre d'occurrences des termes transdisciplinaires, aux termes adoptés ou non, ainsi qu'à leurs co-occurents. Quand il est possible de conclure à un usage proprement transdisciplinaire du terme, cet usage se présente de manière moins figée, dans des associations lexicales peut-être moins attendues¹. Il convient bien entendu de relativiser cette assertion, compte tenu du faible nombre d'occurrences envisagées et de l'absence de test quant à la significativité de ce trait des articles de LET par opposition à ceux de SCL. Simplement à titre de piste qui mériterait d'être creusée, nous retiendrons cette idée de variantes déplaçant les usages canoniques des termes transdisciplinaires, en commentant pour exemple l'emploi d'*hypothèse* dans l'extrait suivant, dernière phrase de la conclusion d'un article de LET :

Ainsi se justifie l'**hypothèse (délirante)** qui a motivé cet article: « Antonio Tabucchi » est peut-être tout simplement un ultime avatar de Pessoa, hétéronyme posthume et imprévu, déniché tardivement au fin fond de la « malle pleine de gens ». [LIT099]

Alors que l'auteur se conforme à un usage standard, tel qu'on a pu l'observer en SCL, par l'emploi du terme *hypothèse* pour qualifier sa propre recherche (phénomène rare en LET nous l'avons dit), il est intéressant de voir que cet emploi s'accompagne d'une mise en parenthèses par laquelle l'auteur évalue sa propre hypothèse au moyen d'un axiologique fort, inattendu au regard des co-occurents d'*hypothèse* relevés dans le corpus. Les parenthèses, ponctuation énonciative forte, donnent lieu ici à un dédoublement énonciatif par lequel se voit commenté le terme d'*hypothèse* – et son statut épistémique.

Un tel tour n'est pas sans effet sur l'image associée à l'instance de prise en charge du texte : l'auteur revêt la figure d'un chercheur, sujet épistémique à l'origine d'hypothèses, mais qui relativise cependant la pertinence scientifique de son hypothèse, l'évaluant sur le mode d'un décalage ; dans le *délire* attaché à la prise en charge d'une telle hypothèse, la figure de l'auteur est celle d'un chercheur qui déplace les valeurs cartésiennes et les modèles canoniques de la science, en reconnaissant sa part de déraison ou d'extravagance presque artistique, tout en attribuant un statut proprement scientifique à son hypothèse.

Intuitivement, l'emploi d'un tel qualificatif ne nous semble pas envisageable en SCL, que ce soit pour une hypothèse propre à l'auteur ou pour discréditer une thèse adverse. Les co-occurents axiologiques les plus fréquents dans les articles de SCL sont *générale*, *globale*,

¹ Mais une telle interprétation en termes d'attentes peut être envisagée elle-même comme un effet disciplinaire.

forte, provisoire, simple, intuitive et séduisante, mais encore une fois, compte tenu du nombre trop restreint d'occurrences pour l'analyse de différences entre articles de ce point de vue et pour celle des collocations, nous entendons simplement avec ces remarques pointer quelques-uns des traits textuels, qui peuvent intéresser en particulier l'analyse de la figure de l'auteur dans l'article.

Parmi d'autres phénomènes locaux qui rejoignent peu ou prou le commentaire apporté ci-dessus par *délirante* au terme *hypothèse*, on relève un emploi du même terme dans un contexte où il prend une valeur dépréciative :

N'allons pas dans les hypothèses pour expliquer la réticence de l'histoire littéraire de nos jours ; **repartons plutôt** à zéro et constatons que les venditions sont de petits poèmes de longueur inégale, de 4 à 10 vers, généralement octosyllabiques, à rimes plates. [LIT044]

La négation syntaxique renforcée par l'adverbe *plutôt* dans des propositions à l'impératif signale ici une forme d'opposition relevant du dialogisme : or il n'est pas rare – et c'est vrai dans les deux disciplines – qu'une hypothèse adverse soit dépréciée ; ici, le point de vue qui fait l'objet de la négation n'est pas montré comme étant attesté dans le champ disciplinaire, il apparaît au contraire comme un point de vue factice. L'opposition concerne l'hypothèse en tant que supposition hasardeuse, plutôt qu'une hypothèse scientifique, et l'on voit ici la différence de statut épistémique (énonciativement parlant) entre les deux emplois du terme.

L'usage d'*hypothèse* ici rejoint ainsi ce que nous avons dit des usages généraux des termes transdisciplinaires dans les articles de LET : la dépréciation contribue à définir la démarche de l'auteur en tant que démarche scientifique qui rompt avec le sens commun, mais en même temps, l'usage général du terme d'*hypothèse* contredit l'univocité attendue de ce terme dans le texte scientifique et par contraste avec les usages transdisciplinaires ; il manifeste en quelque sorte une indifférence à l'égard de son statut épistémique dans le discours scientifique.

Si l'on s'en tient aux occurrences spécifiquement transdisciplinaires des termes retenus, on observe en SCL également des commentaires sur ces termes, en particulier un commentaire très explicite affectant les termes *approche* et *méthodologie* dans SCL007. Nous reproduisons ici le début de la partie introductive, intitulée « Fable d'une approche », où intervient le commentaire :

Il n'est pas possible ici de décrire dans le détail le positionnement théorique sous-jacent à ce travail [...]. Néanmoins, il est nécessaire de donner quelques points de repères, ne serait-ce que par souci de transparence, et au risque que cela soit un peu caricatural, la brièveté aidant.

Le choix du terme « approche »¹ traduit une recherche de souplesse, de relativisation des dogmatismes théoriques, une ouverture au métissage épistémologique, et une modestie dans les attentes que l'on peut entretenir face à un travail de recherche, ce que ne connoterait pas nécessairement « méthodologie ». [SCL007]

L'opposition entre *approche* et *méthodologie*, et le statut épistémique que leur accorde respectivement l'auteur se poursuit encore :

L'approche utilisée ici est fondée sur une fréquentation des terrains, qui enseigne que l'on n'y fait d'abord que ce qu'il est possible d'y faire, et que, par conséquent, les enquêtes « rigoureuses » au sens de « codifiées », « standardisées », « reproductibles », « interchangeables » le sont probablement en payant un prix, bien décrit par exemple par L. Mondada (1998). Comme on ne sait guère fiablement mesurer les avantages et inconvénients des différentes approches possibles, on postulera que celle

¹ L'auteur fait référence au titre de la partie, « Fable d'une approche ».

utilisée ici a autant d'inconvénients que celles se réclamant de la rigueur (au sens ci-dessus), et des avantages globalement comparables. [ibid.]

Plus largement donc, cherchant à justifier le choix d'une approche *de terrain*, attentive à des phénomènes non standardisables, l'auteur en vient en discuter la diversité des approches, en défendant qu'elles peuvent toutes être légitimes et n'ont pas à être hiérarchisées sous prétexte de « *défense de territoires scientifiques* » (ibid., paragraphe suivant). Ce qu'il nous paraît intéressant de pointer ici, c'est que, *via* l'opposition entre *approche* et *méthodologie*, il est possible de voir des valeurs attachées à ces termes et à leur statut scientifique transdisciplinaire ; pour cet auteur, la *méthodologie* connoterait, pour reprendre ses termes, *rigueur*, *reproductibilité*, etc. ; ces critères peuvent être envisagés comme propres à une définition de la science fondée sur les modèles dominants des sciences dures et expérimentales.

On voit donc se manifester dans ce commentaire, comme dans les usages transdisciplinaires ou non d'*hypothèses*, que l'article est, de manière plus ou moins explicite, porteur de valeurs attachées à la recherche, de définitions différentes de la recherche, et renvoie à des manières différentes de faire qui peuvent être comprises comme des positionnements extrêmement divers par rapport à des canons de scientificité.

Plus même, on voit se dessiner des oppositions quant à ces définitions de la recherche, oppositions explicitement thématiques dans l'article de SCL cité ci-dessus, qui propose d'éviter de voir en ces différences des inégalités. L'article serait ainsi le lieu où la présentation de la recherche, impliquant aussi sa légitimation, s'accompagnerait nécessairement mais de manière plus ou moins explicite d'une définition même de l'activité de recherche, dans ses composantes et ses procédures, et en tant que champ social.

La tendance que suggère notre corpus, selon laquelle les articles de LET seraient assez éloignés d'un usage canonique de l'appareillage scientifique – ici abordé sous l'angle du lexique – et d'un modèle dominant de scientificité existerait aussi en SCL ; au terme de cette analyse lexicale, c'est toutefois la tendance inverse, celle d'une tentation de scientificité, qui se présente comme une caractéristique centrale de l'écriture du genre dans ce champ des SCL, bien que l'étude mette en évidence aussi la diversité, voire les oppositions des modèles au sein d'une même discipline.

5.3.8. Des différences liées au statut de l'auteur?

Nous clôturerons cette analyse du lexique transdisciplinaire dans le genre de l'article en nous intéressant à notre hypothèse secondaire quant aux facteurs de variation du genre, celle selon laquelle le statut de l'auteur pourrait induire un usage spécifique du genre. Nous proposons d'évaluer l'effet de ce statut sur l'usage du lexique, envisagé ici du seul point de vue quantitatif, en nombre d'occurrences des termes de la liste.

Signalons d'abord que le statut de l'auteur n'influe pas de manière significative sur le nombre d'occurrences des termes pris dans leur ensemble. Aussi pouvons-nous conclure à une relative stabilité du genre dans l'usage de ce lexique, avec des différences qui concernent prioritairement des identités disciplinaires. Le détail de chaque terme fait quant à lui apparaître des différences, mais ne contredit pas complètement le premier constat dans le sens où les termes discriminants les textes des deux groupes « doctorant » et « non doctorant » restent assez peu nombreux, et sont presque négligeables au sein des articles de LET, contrairement à la variation inter-disciplinaire qui affecte au total rappelons-le 48 des 100 termes considérés.

Concernant le poids de chaque terme, un premier test consiste à évaluer l'effet du statut de l'auteur indépendamment de la discipline d'appartenance. On observe ainsi que le terme « présumé » est complètement absent des articles des doctorants, qu'ils soient de LET ou de SCL, mais la différence avec les non doctorants n'est statistiquement pas significative. Elle l'est pour 11 termes : 2 sont en moyenne significativement moins usités par les doctorants, et 9 le sont au contraire davantage. Le détail de ces caractéristiques marquées par leur usage plus abondant dans les textes des non-doctorants ou inversement dans ceux des doctorants est présenté dans le tableau ci-dessous.

Corpus pris dans son ensemble	
Usage plus important dans les textes des non-doctorants	Usage plus important dans les textes des doctorants
principe théorie	constater constituer étude méthodologie objectif observer résultat tableau tendance

Figure 5.16. Les usages contrastés du lexique épistémique : l'effet du statut de l'auteur.

Des différences liées au statut de l'auteur se manifestent également au sein de chacune des deux disciplines, pour 2 termes en LET et pour 15 termes en SCL.

En LET, les 4 termes qui font l'objet d'un usage différencié en fonction du statut du scripteur sont tous davantage représentés dans les textes de doctorants :

Articles de LET	
Usage plus important dans les textes des non-doctorants	Usage plus important dans les textes des doctorants
	constituer exemple

Figure 5.17. Les usages contrastés du lexique épistémique : l'effet du statut de l'auteur dans les articles de LET

En SCL, sur les 15 termes dont l'usage est marqué en fonction du statut de l'auteur, 10 sont davantage employés dans les textes des non-doctorants, et 5 dans les textes des doctorants :

Articles de SCL	
Usage plus important dans les textes des non-doctorants	Usage plus important dans les textes des doctorants
argument	constater
conclusion	étude
définir	méthode
définition	observer
exemple	résultat
idée	
interprétation	
notion	
problème	
propriété	

Figure 5.18. Les usages contrastés du lexique épistémique : l'effet du statut de l'auteur dans les articles de SCL.

Comment interpréter alors ces résultats ? Le premier test montre que certains termes distinguent les articles des doctorants et ceux des non doctorants indépendamment de la discipline dont relève leur texte. Or, s'il existe des tendances communes aux doctorants quelle que soit la discipline, alors que celle-ci est un facteur important de différenciation des pratiques, c'est peut-être que les types d'articles et d'étude qu'ils produisent sont spécifiques. Envisagée en terme d'usage plus ou moins important de quelques termes transdisciplinaires, la spécificité de leurs articles peut tenir à ce que les doctorants rendent compte dans leurs articles de travaux d'analyse et de recherches qui se caractérisent par leur empirisme : c'est ainsi que pourrait s'expliquer leur usage plus massif des termes *constater*, *étude*, *méthode*, *objectif*, *observer*, *résultat*, *tableau*, *tendance* dans les articles de SCL ou les deux disciplines confondues.

L'usage plus important dans les textes des non-doctorants de « principe » et « théorie » les deux disciplines prises ensemble, et de *argument*, *définir*, *définition*, *idée*, *notion*, *problème*, *propriété* en SCL signifierait dans le même sens une opposition des textes des deux groupes d'auteur fondée sur une opposition entre études empiriques et études théoriques, distinction déjà discutée dans son intérêt et ses limites à propos des exemples dans l'analyse de la structure du genre.

Face à une telle hypothèse interprétative, la prudence est bien entendu de mise, car il n'est pas possible de prétendre ne pas se tromper en affirmant que les termes cités seraient davantage caractéristiques pour les uns d'articles rendant compte de recherches empiriques, et pour les autres d'articles plus théoriques, dédiés à un état des lieux sur une notion ou un champ d'investigation, ou présentant simplement davantage de discussions quant aux paradigmes en jeu.

Cela étant, cette distinction entre des types d'études semble se vérifier par l'examen des articles produits par les doctorants en SCL ; en LET, il semble plus difficile d'envisager une distinction de types d'articles sur un axe théorique-empirique – qu'il resterait soulignons-

le à déterminer ; en retour, l'absence de différence en LET quant au type d'étude mené par les deux groupes d'auteurs pourrait expliquer qu'un nombre presque négligeable de termes discrimine les deux groupes d'auteur, et qu'aucun ne se présente comme plus utilisé par les non-doctorants.

Il est possible par ailleurs que les textes des doctorants tendent davantage à se rapprocher de modèles canoniques du genre. Les termes qu'ils emploient plus fréquemment en SCL, *étude*, *résultat*, avec aussi, dans les deux disciplines, des termes comme *objectif*, *méthodologie* indiqueraient que les doctorants se réfèrent à des standards en vigueur, définissant leur travail de recherche – et celui d'autres chercheurs – en recourant à un appareillage transdisciplinaire canonique, comme obligés d'évoquer en ces termes leurs *objectifs* et les *méthodes* employées.

Finalement, il semble possible d'avancer que les doctorants auraient davantage recours que les non doctorants à un appareillage terminologique servant à décrire quelques aspects bien identifiés de leur travail de recherche ou d'autres travaux de recherche ; les termes qui caractérisent positivement leurs textes (*objectif*, *méthode*, *résultat*) se présentent comme des composantes standards, qui ne vont pas sans évoquer les entrées du plan « IMRD » (*Introduction, Materials and Methods, Results, Discussion*) canonique de l'article scientifique.

Les termes discriminant les deux groupes des textes de doctorants *versus* ceux des non-doctorants signaleraient des différences quant aux types de recherches présentées dans les articles, en particulier en SCL. Les travaux de doctorants intégreraient ainsi une composante empirique signalée comme telle, tandis que ceux des non-doctorants accorderaient une place plus importante à l'identification de *théories*, la discussion de *notions* ou la mise en évidence de *problèmes*.

Cependant, la seule entrée du lexique transdisciplinaire ne suffit pas à en dire davantage, et nous retiendrons prioritairement de cette analyse l'usage disciplinaire qui est fait de ce lexique. Dans la mesure où sauf pour quelques termes en SCL notamment, celui-ci fait l'objet d'emplois relativement stables quel que soit le statut du scripteur, ce lexique se présente comme une entrée intéressante pour caractériser le genre dans une composante centrale des textes.

Les fortes différences entre les disciplines signalent par ailleurs l'intérêt de ce lexique pour la mise en évidence d'orientations épistémologiques plurielles, voire adverses, et ce au sein d'un même champ comme nous l'avons vu pour les SCL. Si ce lexique fait la scientificité des textes, il montre dans ses usages la diversité des pratiques, et des positionnements possibles par rapport à cette idée de scientificité ; son statut épistémique quant à lui suggère la diversité des modes de validation du propos dans l'article, avec la figure d'un auteur qui en LET peut aller jusqu'à revendiquer une part de *délire*, et tend du moins à discuter la scientificité qu'il n'a de cesse d'affirmer en SCL, et ce quelle que soit par ailleurs l'*approche* scientifique que le linguiste entend représenter.

5.4. Des caractéristiques attendues et figées, et des usages contrastés : bilan sur la morpho-syntaxe, la ponctuation et le lexique du genre.

L'analyse linguistique de l'article montre l'intérêt d'un profil général de la phrase, de la question des temps verbaux et des marques de personne et de non-personne, de la ponctuation et du lexique transdisciplinaire pour affiner la caractérisation du genre. Elle permet de pointer ses traits centraux, conformes aux caractéristiques attendues, et de montrer que c'est aussi aux traits marginaux, résultant de prescriptions négatives, que le genre doit sa stabilité. Une force centripète du genre se manifeste dans ces aspects centraux ou marginaux relativement figés et dans des profils de textes partagés entre les différents groupes. Cette unité du genre permet en retour de mieux cerner certains faits linguistiques. Loin d'une approche qui en gommant les spécificités des genres aboutit à des inventaires par trop hétéroclites, une étude au sein des genres pointe des usages récurrents et des fonctions spécifiées.

La variabilité observée montre des usages contrastés qui tiennent aux différences entre les groupes (SCL et LET, doctorants et non-doctorants) ; ainsi le palier méso-textuel du genre intègre-t-il certaines de ses caractéristiques structurelles : les modèles de textualité décrits à partir d'une macro-analyse se trouvent corroborés par l'étude de la phrase, par certaines des ponctuations et par les balises méta-textuelles mises en évidence à partir des temps verbaux. On retrouve l'idée d'un guidage du lecteur plus important en SCL qu'en LET, et chez les doctorants par rapport aux non-doctorants, confirmé encore au niveau des marques de personne. Mais cette rhétorique du genre qui implique à des degrés contrastés son lecteur repose sur une phraséologie largement routinisée.

Les différences entre les groupes ne subsument pas la diversité du genre ; des modèles scripturaux contrastés servent le rapport à l'objet (son traitement historique, les modes d'analyse du matériau étudié, la dimension polémique de certains objets), et font état de types d'études possiblement spécifiés en SCL sur un axe entre le théorique et l'empirique ; plus largement des orientations épistémologiques plurielles se profilent, qui montrent comment la validation du savoir peut emprunter des cheminements variés.

Affiner la description du genre permet ainsi d'étudier comment les usages déplacent, ou relativisent fortement les caractéristiques supposées de l'écrit scientifique, et de voir que les textes se positionnent diversement par rapport aux canons de scientificité. La question de la figure de l'auteur semble prioritairement concernée, et atteste de l'intérêt de la morpho-syntaxe, de la ponctuation et du lexique sur cette question que nous retrouverons avec les styles, et de manière plus centrale dans la dernière partie d'analyse.

La transparence réputée d'une écriture scientifique centrée sur son objet est nettement réductrice, et reste pourtant visée. La rhétorique du texte est une rhétorique de la découverte, qui implique le lecteur à « partir explorer » avec le scripteur le matériau étudié, et à prendre part active dans la construction du savoir. Un goût de l'écriture se manifeste, au service toutefois de l'objet, dans la recherche d'un mot juste qui donne matière à penser. La fonctionnalité n'est jamais déjouée, mais la part de jeu avec le texte est une possibilité, exploitée à des degrés divers, de manière sans doute largement idiomatique.

C'est cette variation spécifique du genre, celle qui tient à l'usage que chaque auteur en fait, que nous proposons à présent d'étudier ; à partir des mêmes variables que celles ici traitées, nous avons mis en place un dispositif spécifique qui permet de cerner en quoi l'écriture de l'article n'interdit pas l'existence de styles, propres à un auteur. S'il s'agit de pointer des choix scripturaux caractéristiques des textes d'un même auteur, c'est aussi sa pratique de la recherche qui fait la spécificité de ses textes dans le champ.

Chapitre 6.

Le genre et ses usages singuliers, les styles : une entrée pour l'analyse de la diversité interne du genre.

La description du genre de l'article du double point de vue de sa stabilité et de sa variabilité appelle à s'intéresser à l'usage singulier que chaque auteur en fait, et que nous conviendrons d'appeler son style. Existe-t-il des styles en sciences, dans l'écriture d'un article ? En quoi l'écriture du genre chez un même auteur est-elle singulière, autrement dit qu'est-ce qui fait son style, et quel est son degré de singularité ? Dans quelle mesure ce facteur de variabilité des pratiques entre-t-il ou non en conflit avec la notion de genre ?

L'objectif est d'avancer dans la description du genre, en intégrant ce nouveau facteur potentiel de variation, et en nous centrant sur les variables linguistiques étudiées dans la partie précédente (la morpho-syntaxe, les ponctuations et le lexique). L'étude nécessite un corpus et une méthodologie d'analyse spécifique pour étayer l'hypothèse des styles et c'est la raison pour laquelle cette partie isolée lui est dédiée.

Les enjeux de l'étude menée sont multiples. Il s'agit de poursuivre l'analyse des textes dans leur diversité, de voir ce qui fait la spécificité des textes produits par un même auteur, et dans quelle mesure ils se distinguent de ceux des autres. Par ailleurs, c'est la question du rapport entre styles et genre qu'il s'agit de mieux cerner, l'hypothèse ici examinée étant que le genre fait l'objet d'usages singuliers qui ne remettent toutefois pas en question son unité. Une étude de la diversité des textes du même genre qui prend comme point de départ sa variation stylistique doit également permettre d'affiner ses caractéristiques supposées, éventuellement de les dépasser et de mettre en évidence quelles caractéristiques textuelles font l'unité du genre, i.e. quelles caractéristiques sont communes aux textes d'un même genre, et quelles caractéristiques structurent de manière relativement stable sa variabilité.

6.1. Le style en sciences : une question occultée

La question des styles en sciences peut paraître paradoxale. Tandis que la littérature échapperait aux genres, selon une perspective qui oppose modèles du genre et créativité langagière du sujet (cf. chapitre 1), l'écriture de la science échapperait quant à elle à cette créativité. L'écriture scientifique est en effet généralement traitée en termes de conventions et de contraintes, voire de « routines scripturales » (De Nuchèze, 1998). Le caractère figé et routinisé de l'écriture des sciences interdirait l'écart stylistique en tant qu'écart formel qui

produit un effet expressif ou esthétique ; en outre, le style tend à impliquer un auteur, et une intention expressive ou esthétique à produire un tel effet.

Dans l'écrit scientifique, ce que l'on pourrait appeler la « part du scripteur »¹ tend à être occulté, en raison d'une part de la dimension collective de l'activité scientifique et de la conception du sujet qu'elle implique, et d'autre part du caractère impersonnel et objectif attribué à l'écrit scientifique. Communément entendu comme propre à l'écrivain, le style serait une dimension neutralisée de textes censément objectifs, qui honorent le seul *logos* : le scripteur de ces textes « sans locuteur »² est même conçu comme « non-auteur »³, et n'accéderait pas en cela au statut d'écrivain⁴.

La dimension stylistique de l'écriture de la science, quand elle n'est pas taxée de fioriture, voire de déviance journalistique⁵, ne semble être admise que dans la mesure où elle est au service de qualités fonctionnelles que l'on pourrait gloser ainsi : soyez clair, soyez concis, et soyez pertinent. Cette idée que la fonctionnalité de l'écriture scientifique exclut toute recherche stylistique est forte dans l'image que les étudiants se font de l'écrit de recherche, comme l'a mis en évidence Y. Reuter (1998). Elle fait obstacle selon cet auteur à l'entrée dans ce type d'écrit, en se greffant sur un ensemble plus large de représentations-obstacles, qui consistent à nier le processus de production de l'écrit.

Les raisons de ce déni du travail stylistique dans l'écrit scientifique méritent d'être regardées de plus près ; ce faisant, c'est la question du style, et celle de l'écriture scientifique qu'il faut revisiter pour que la problématique du style d'auteur en sciences prenne sens.

Le déni du travail stylistique dans l'écrit scientifique repose d'abord sur une définition du style historiquement fondée sur des critères esthétiques et tendanciellement assimilée à des qualités artistiques.

Dans une stylistique que l'on pourrait dire générale, l'analyse consiste à évaluer une différence : le statut stylistique particulier d'un mot ou d'une autre unité linguistique tient à sa mise en relation avec les autres unités, sur le plan syntagmatique et paradigmatic, d'une part. D'autre part, l'écart stylistique est à comprendre comme le souligne L. James en fonction d'« un code général stylistique et rhétorique qui informe notre analyse »⁶ (Cressot et James, 1991 : 10, ed. originale 1947). Le degré zéro du style correspondrait donc à la langue dans son caractère le plus figé, une langue neutre, purement informative, où rien d'autre que la fonction référentielle (Jakobson, 1963) n'attirerait le regard du lecteur. Inversement, le style caractériserait mieux que tout autre le discours littéraire.

¹ En reprenant le titre de l'étude de M. Visoz (1989) : *La part du lecteur*.

² G. Philippe (2002).

³ Cf. D. Pontille (2001, 2004, 2006), qui entend renoncer à l'exclusion de cette notion d'auteur en sciences ; il montre que valide, elle doit cependant être envisagée sociologiquement à l'encontre d'une conception littéraire de l'auctorialité, sous l'angle d'une conception collective de l'action, de la responsabilité énonciative et de l'autorité en sciences.

⁴ Statut qui semble sociologiquement réservé aux essayistes et aux auteurs littéraires reconnus comme tels dans le champ des productions culturelles.

⁵ Que l'on songe en particulier à certains titres identifiés comme « accrocheurs » qui reposent sur le mécanisme polyphonique du détournement, et jouent par là sur un effet de connivence avec le lecteur : de mise dans le discours journalistique, ils peuvent être en usage dans le discours scientifique – peut-être davantage dans les communications orales qu'écrites – mais sont généralement prohibés dans les prescriptions méthodologiques destinées aux étudiants entrant dans l'écrit de recherche.

⁶ C'est ainsi que l'on peut expliquer une autre acception communément répandue du style, indépendante de qualités artistiques, dans des expressions comme « style télégraphique » ou « style journalistique ». Non moins significative quant à l'« esthétique de la production verbale » telle qu'elle est communément appréhendée, cette acception revient à envisager le style en termes de spécificités phraséologiques des textes qui correspondent vraisemblablement à des codifications génériques.

Cela étant, le style implique aussi un sujet à l'origine de choix linguistiques au service d'une esthétique et d'une expressivité. Dans cette perspective, l'analyse stylistique va interpréter le fait de style comme résultant d'une intention stylistique ou comme une pratique idiomatique de la langue, non nécessairement intentionnelle. La comparaison qui permet de conclure à un fait de style n'a plus alors pour référent le message réduit à sa fonction informative, mais un ensemble de textes du même auteur ou d'auteurs différents (Cressot et James, op.cité).

C'est cette optique qui caractérise les études stylométriques. La notion d'auteur y intervient, au moins en termes d'unité et de spécificité des pratiques textuelles d'un individu. Se donnant pour objet l'analyse quantitative et statistique de traits stylistiques, les études stylométriques envisagent ces derniers comme des caractéristiques textuelles – souvent restreintes à la dimension lexicale – qui sont propres à un auteur et distinguent son écriture entre toutes. Cette définition structuraliste du style comme « une manière d'être distinctive »¹ trouve en la statistique un outil adapté : le fait de style, en tant qu'élément jugé différent, n'est plus seulement appréhendé en tant que tel par le lecteur ou l'analyste. C'est l'évaluation statistique qui vérifie un écart significatif, en mesurant et comparant les caractéristiques des textes de divers auteurs ou d'un même auteur. Si plusieurs œuvres d'un auteur comportent les mêmes caractéristiques, l'auteur fait un usage récurrent d'un style particulier. Si l'œuvre d'aucun autre auteur ne possède les mêmes caractéristiques, on peut dire que son style est unique (Holmes, 1994).

En s'intéressant prioritairement au discours littéraire² et dans une moindre mesure au discours politique³, de telles études ont ainsi cherché à identifier ce qui fait la spécificité d'un auteur en comparant des textes, à seule fin descriptive, ou pour dater les textes, en déterminer la paternité et débusquer les apocryphes. Elles reposent donc sur ce postulat que les textes d'un même auteur présentent des caractéristiques qui le distinguent de ceux d'un autre au point de pouvoir être conçues comme indicielles de l'identité du locuteur qui en est à l'origine⁴. Leur absence d'intérêt pour le discours scientifique s'explique par ce postulat, qui contredit le statut du sujet dans le texte scientifique, et plus largement peut-être par les objectifs mêmes des études du style : l'esthétique et l'expressivité du fait stylistique (en tant que fait intentionnel ou en tant que pratique idiomatique) doit permettre de cerner d'une certaine manière l'essence de l'auteur et/ou de l'œuvre, le talent du premier, l'esprit et la portée de la seconde, ou, en un mot, leur « génie »⁵.

Par ailleurs, l'idée d'exclusion de la dimension stylistique dans l'écriture scientifique peut s'expliquer sous l'angle du rapport entre l'activité de recherche et sa mise en forme dans l'écrit. Ce rapport n'est généralement pas questionné, vraisemblablement sous l'effet conjoint de représentations de l'écrit et de représentations de la science. On l'a dit des représentations que les étudiants se font de l'écrit de recherche (Reuter, 1998), mais nombre de travaux l'ont

¹ M. Cressot (1974) cité par D. Monière et D. Labbé (2002).

² Cf. par exemple les travaux de C. Muller (1968) sur le théâtre classique, ceux d'E. Brunet (1978) sur J. Giraudoux, de B.M. Kylander (1995) sur Molière et de C. Bernet (1983) sur Racine.

³ Cf. par exemple l'étude du général de Gaulle par J.-M. Cotteret et R. Moreau (1969), et par Léon (1971), celle de J. Roche (1971) sur les candidats aux élections présidentielles de 1965 et 1969 ou encore celle de Labbé (1983) sur François Mitterrand et de D. Monière et D. Labbé (2002) sur trois premiers ministres québécois, Duplessis, Bourassa et Lévesque.

⁴ Cette hypothèse explique les fréquentes controverses que suscitent les entreprises stylométriques, qui demandent en effet à se questionner sur la pertinence des traits retenus d'une part et d'autre part sur celle des calculs probabilistes des proximités.

⁵ Soulevant ainsi l'inévitable question de savoir si « le génie d'un auteur peut se traduire en données statistiques aussi crues que la taille de son vocabulaire, ou la longueur des mots, phrases et paragraphes qu'il emploie » (J. Gani, cité par J. Hamann 1996).

montré¹, c'est vrai de toutes les représentations de l'écrit, quels que soient les usagers et les usages concernés ; elles tendent à passer sous silence le processus de sa production au profit d'une focalisation sur le seul produit fini.

Concernant l'écrit scientifique, il nous semble avoir ceci de spécifique que le produit écrit est perçu comme la transposition relativement mécanique du processus de la recherche. Il est difficile de comprendre exactement ce dont procède une telle représentation ; le processus de la recherche, en tant que processus praxéologique, est passé sous silence, et c'est ce silence qu'ont précisément entendu rompre ceux qui se sont attachés à décrire la « vie de laboratoire » (Latour et Woolgar, 1979). Hors de ces travaux, cette dimension praxéologique telle qu'on se la représente se trouve reconstituée à partir de ce qu'en livre le produit écrit, comme une activité linéaire dont chacune des composantes serait menée de manière parfaitement rationnelle.

Il est intéressant à cet égard d'observer le traitement de cette question dans les manuels à usage méthodologique ; si l'on prend pour exemple un ouvrage de référence tel que le *Manuel de recherches en sciences sociales* de R. Quivy et L. Van Canpendhoudt (1998), on remarquera qu'il aborde comme l'indique son titre le processus même de la recherche : les auteurs en proposent un modèle récursif, dont l'objet est bien de dépasser une conception linéaire de la démarche de recherche pour mettre en évidence les incessants va-et-vient entre l'empirisme et le théorique, entre les hypothèses et les analyses, entre les observables et leur interprétation.

Le rapport entre ce processus et sa mise en forme dans l'écrit est thématiquement dans l'ouvrage, en termes de reconstruction de la démarche de recherche, mais l'analyse du processus et de sa reconstruction tendent à en donner une image relativement canonique, marquée par des valeurs scientifiques de rigueur et de rationalité. Cette image que l'on pourrait qualifier de normative a sa part de fonctionnalité, *a fortiori* dans un discours à visée didactique comme l'est celui des manuels : elle participe du socle commun de conventions sur lesquelles se fondent l'activité et le groupe ; en ce sens, elle peut se trouver renforcée par la mission éducative des discours qui se proposent d'aider l'acculturation du sujet au champ de pratiques considéré.

La primauté de la fonctionnalité du texte scientifique explique qu'il soit défini comme un discours exclusivement dédié à son objet, et/ou comme la transposition fidèle d'un processus de recherche, qui consisterait en une série d'opérations rationnelles, calquées sur celles dont rend compte le produit écrit. Perçu sous l'angle des conventions qui le caractérisent, l'écrit scientifique véhicule un ensemble de représentations partiellement fondées ou non par lesquelles le style se trouve occulté ou nié, qu'on l'envisage en termes d'écart linguistique ou en termes de pratiques spécifiques à un sujet. La question du style en sciences prend pourtant sens dès lors que l'on restreint la labilité de la notion à une définition textuelle du style, en tant que singularité des textes d'un même scripteur, et que l'on cesse d'envisager le style de manière indépendante, pour l'intégrer dans un cadre d'analyse des pratiques scripturales.

6.2. Une définition du style relative aux genres

L'intérêt pour la notion de style dans le discours littéraire prioritairement incombe à divers facteurs, qui se rattachent de près ou de loin au caractère singulier de l'œuvre et à la créativité du sujet. Son analyse peut ignorer la notion de genre, quand les textes d'un même auteur sont étudiés sous l'angle de leur homogénéité et de leur spécificité stylistique indépendamment de

¹ Cf. par exemple D. Bourgain, (1988), M. Dabène, (1987), I. Delcambre et Y. Reuter, (2002), R. Guibert, (1989).

leurs différences génériques¹. Plus largement, les deux notions paraissent *a priori* diamétralement opposées : quand l'une renvoie au locuteur et à la singularité de son expression linguistique, l'autre repose sur l'intersubjectivité et montre comment les pratiques langagières respectent des modèles textuels stables, selon le double principe du dialogisme bakhtinien (cf. chapitre 1).

Genre et styles ne sont pas pour autant contradictoires : le « déjà-là » des genres, constitutif de la création verbale, n'exclut pas la singularité du sujet (cf. chapitre 1). Analyser cette singularité demande à tenir compte des genres : une analyse réalisée sans référence au genre permet au mieux de dégager des types linguistiques de textes ; elle n'autorise pas à conclure que leurs ressemblances et différences sont le seul fait de leur paternité, comme le montre E. Brunet à propos des débats suscités par les entreprises stylométriques quant à la paternité des textes littéraires : « pourquoi chercher à confondre Molière et Corneille, quand le partage du même genre comique et de la même expression versifiée suffit à expliquer la ressemblance ? »². Inversement, dans quelle mesure un auteur conserve-t-il ses spécificités quand il change de genre³ ?

Aussi est-ce une définition du style relative au genre qui intéresse notre étude ; nous intégrons la notion à un cadre global d'analyse des pratiques scripturales et considérons avec F. Rastier (2001) les styles en tant qu'usages singuliers du genre. La notion de style doit ainsi permettre de rendre compte de la diversité des textes : elle est un des facteurs potentiels de variabilité du genre et renvoie à l'idée que les textes d'un même auteur au sein d'un même genre présentent des propriétés distinctives qui caractérisent son style.

C'est l'hypothèse que nous nous proposons de tester sur l'article de recherche, genre *a priori* fortement contraint et peu à même d'autoriser la « marge de variation » offerte au scripteur. Deux objectifs guident l'étude, qui est menée statistiquement :

- évaluer la validité de la notion de style en sciences, en observant si émergent d'un corpus d'articles d'éventuels styles d'auteurs qu'il nous reviendra de caractériser ;
- analyser la diversité des usages du genre et ses facteurs de variation, en appréciant la variabilité stylistique du genre et son degré de stabilité.

Plus précisément, trois entrées seront privilégiées pour rendre compte des usages singuliers que chaque auteur fait du genre et de l'impact de la variation stylistique sur le genre.

- 1) **La caractérisation des styles**, et la mise en évidence de « lieux-pivots », ou variables textuelles privilégiées de la variation stylistique : en comparant les textes d'un même auteur à d'autres textes, on peut décrire les traits qui font la spécificité des premiers ; avec plusieurs auteurs, on peut alors dire si certains traits sont particulièrement soumis à des usages stylistiques.
- 2) **L'analyse du degré de distinctivité des styles** : une fois validée statistiquement l'existence de styles, et décrit ce qui fait la spécificité des textes d'un même auteur, il s'agit de voir dans quelle mesure cet usage est spécifique. Nous proposons d'aborder cette question sous deux angles, celui de l'homogénéité et celui de la singularité des articles d'un même auteur par rapport à d'autres textes du même genre.

¹ Biais analysé comme tel dans E. Brunet et C. Muller, 1988.

² E. Brunet cité par T. Mézaille (2003).

³ Ibid. Notons par ailleurs que l'étude de E. Brunet, consistant à comparer le vocabulaire des différents textes de Victor Hugo, montre que les différences de vocabulaire se superposent à celles des genres : sur la classification graphique obtenue, "at the bottom are grouped all the poetic collections, while the novels and plays are placed in the upper half, without merging too much into each other" (1991 : 76).

-
- 2) **Le retour sur le genre** à partir de l'analyse de sa variation stylistique. Face à la diversité des textes se pose la question de la stabilité du genre. Elle peut être traitée du point de vue de ce qui est commun aux textes, s'opposant alors à l'idée d'usages singuliers. Elle tient aussi à ce qui structure de manière stable la diversité des textes du même genre ; en ce cas, les usages stylistiques reflètent des usages textuels spécifiques, marqués les uns par rapport aux autres, et qui définissent des tendances distinctes, caractéristiques du genre dans sa variabilité.

Nous nous intéresserons aux seuls articles de sciences du langage, sans tenir compte des variations jusqu'ici analysées entre les disciplines et en fonction du statut de l'auteur. Ce choix se justifie par la spécificité de l'étude, et par la lourdeur du protocole à mettre en œuvre pour étayer l'hypothèse des styles dans l'article.

6.3. Méthodologie pour l'étude des styles

Pour décrire et valider statistiquement l'effet de la variable « auteur » sur les textes, encore faut-il disposer d'un corpus qui s'y prête. Par ailleurs, il importe de se doter d'une méthodologie spécifique permettant la caractérisation textuelle des styles et de la variabilité du genre. C'est en collaboration avec C. Poudat (Laboratoire Coral, Orléans) que nous avons entrepris cette étude du style dans l'article, nos travaux de doctorat appelant à tenir compte de cette dimension du style et de ses relations avec le genre de l'article.

En dépit de la spécificité de leurs objets respectifs, nos deux études ont en effet en commun de chercher à décrire le genre de l'article en linguistique, en se fondant sur des méthodes de linguistique de corpus. La variation stylistique du genre représente une entrée particulièrement intéressante pour mieux cerner ce qui fait son unité et sa diversité, avec d'autres axes de contraste : les différences inter-génériques dans l'étude de C. Poudat (2006a), et la dimension contrastive entre les articles en français et en anglais, les différences entre les articles de SCL et ceux de LET, et la variabilité liée au statut de l'auteur dans notre cas. Dans les deux cas, l'analyse des styles permet d'une part de s'intéresser à un facteur de variation capital, puisqu'il se présente comme le pôle opposé d'un genre conçu en termes de modèle figé. Il permet alors de spécifier quelles caractéristiques textuelles sont concernées par la variation des usages, et comment se structure la variabilité du genre.

Pour mener à bien une telle analyse, il s'avérait intéressant de mutualiser nos corpus d'étude du genre et d'élaborer un corpus commun spécifiquement dédié à l'étude des styles. L'étude nous a également permis de réunir les jeux de descripteurs – les variables textuelles prises en compte – et les outils d'analyse (logiciel d'analyse automatique des textes, quantification et traitement statistique des variables). Le dispositif ainsi constitué répond aux objectifs d'analyse, et rend possible aussi la confrontation des protocoles respectifs.

L'étude a donné lieu à deux publications¹, mais nous en proposons ici une version différemment structurée ; nous cherchons à mettre en relation les résultats de l'étude avec les aspects envisagés dans les parties précédentes, et nous essayons également de développer la réflexion sur le lien styles et genre, et sur la stabilité et la variabilité de ce dernier.

6.3.1. Le corpus : un corpus « genre » et un corpus « auteurs »

Valider en corpus la notion de style demande à constituer un corpus comportant un nombre suffisant de textes par auteur. Ce critère entre en conflit avec les critères présidant à la constitution d'un corpus pour l'étude du genre (cf. chapitre 3) : celui-ci demande à neutraliser

¹ C. Poudat et F. Rinck, 2006, et à paraître.

la variable « style ». Aussi l'analyse s'est-elle fondée sur deux corpus comparables, distincts du seul point de vue de cette variable : un corpus « genre » et un corpus « auteurs ». Ces deux corpus sont étudiés séparément puis associés et comparés pour analyser les styles et ce qu'ils disent du genre.

Le corpus « genre » résulte de la réunion du corpus de C. Poudat (*op. cité*) et du nôtre. Il est génériquement et domanialement homogène : ne sont retenus que des articles de revues linguistiques, afin de limiter l'impact encore mal maîtrisé des domaines, ou sous-disciplines scientifiques sur le genre. On notera toutefois que diverses sous-disciplines linguistiques y sont représentées (sémantique, syntaxe, phonologie, etc.), et cette représentativité difficilement évaluable est postulée sur la base d'une approbation par des experts du champ. L'assemblage des deux corpus est permis par leur compatibilité, au niveau des critères génériques¹ de leur constitution : conçus pour répondre à un même objectif de description du genre, les critères employés vérifient – sinon une identité exacte² – une proximité, et celle-ci peut ainsi être envisagée comme indice de leur pertinence pour l'analyse des genres. Une fois supprimés les éventuels doublons, le corpus se constitue de 330 articles extraits de 54 numéros de revues (soit 18 revues) francophones de sciences du langage, publiés entre 1998 et 2003. 332 auteurs y sont représentés.

A ce corpus de référence pour l'étude du genre, s'ajoute un corpus « auteurs » dédié quant à lui à l'étude des styles, qui rassemble 130 articles de 15 linguistes français accrédités dans le champ, soit entre 5 et 8 textes par auteur. Hormis ce critère spécifique, les autres reprennent ceux du corpus « genre ». Les textes sont demandés directement aux auteurs compte tenu de la nécessité de recueillir plusieurs textes écrits par chacun d'eux³. Un nettoyage des textes est nécessaire pour s'assurer qu'ils sont conformes à la version effectivement publiée.

Un autre problème se pose cependant, celui du genre de texte : rares sont les auteurs qui peuvent nous fournir un nombre important d'articles de revue parus dans les 5 années stipulées. Nous devons donc faire entorse à ce critère en intégrant au corpus des textes de genres dont on postule qu'ils sont suffisamment proches de l'article de revue pour ne pas induire de biais dans l'analyse du style en tant qu'usage singulier du genre : l'article publié en tant que contribution à un ouvrage collectif, et l'acte de colloque. Intuitivement fondé, le postulat de la proximité de ces genres avec celui de l'article de revue demeure néanmoins une limite du corpus, qu'on tentera cependant de contrôler en testant la pertinence générique de ce dernier.

Etant donné les particularités de l'étude, nous avons choisi d'écarter les exemples des textes, fortement susceptibles de corrompre leurs caractéristiques morphosyntaxiques. Pour éviter de briser la continuité syntaxique, ce qui nuirait à la pertinence de l'analyse, nous avons choisi de n'isoler que les exemples délimités dans les textes, i.e. introduits sous la forme d'un décrochement topographique. Comme nous l'avons vu à propos des exemples (dans l'analyse des composantes structurelles), et dans l'analyse de la morpho-syntaxe et de la ponctuation, il reste que, même de manière moindre une fois les exemples ôtés, les textes portent nécessairement la trace de leur matériau.

¹ Ces critères visent une homogénéisation générique nécessaire pour contrôler les facteurs de variabilité du genre (par ex., la date de parution, le domaine comme nous venons de le dire, etc.).

² On note en effet quelques différences : dans notre cas, nous avons dû mettre à l'écart les textes signés par plusieurs auteurs ; par ailleurs, le critère de date de publication est légèrement plus restrictif dans notre étude (articles parus entre 1999 et 2003) que dans celle de C. Poudat (entre 1998 et 2003) ; concernant les revues représentées, c'est la sélection de C. Poudat qui semble plus restrictive, dans le sens où elle entend s'appuyer sur des revues notoires, plus diffusées dans le champ, alors que nous avons retenu aussi bien des revues nationales que des revues locales.

³ La revue et le numéro dont relèvent les articles sont donc extrêmement hétérogènes au sein du corpus « auteurs », davantage que dans le corpus « genre ».

6.3.2. Les traits textuels retenus

Les études stylométriques ont tenté d'associer aux styles d'auteur des traits lexicaux, morphosyntaxiques ou phonologiques¹ en fonction des genres concernés. L. Lebart et A. Salem (1994 : 243) montrent que la plupart recourent à des indices synthétiques construits à partir des longueurs des mots, des phrases, la fréquence des mots-outils, la richesse du vocabulaire, ce à quoi on peut ajouter les thèmes et champs lexicaux. Si globalement, c'est donc la dimension lexicale qui est privilégiée (Holmes, 1994 : 91), on soulignera avec R.W. Bailey que les traits stylistiques doivent pour être analysés comme tels présenter les propriétés suivantes : « they should be salient, structural, frequent and easily quantifiable, and relatively immune from conscious control » (cité par Holmes, 1994). Conséquence directe des possibilités offertes par l'outil informatique, les analyses lexicométriques du style, largement dominantes, se trouvent aujourd'hui complétées par la linguistique de corpus, dans des études portant sur d'autres niveaux de la textualité².

Nous cherchons à analyser les variations stylistiques du genre de l'article à partir des mêmes variables que celles dont a fait l'objet notre précédente partie, à savoir la morpho-syntaxe et la ponctuation³, ainsi que le lexique transdisciplinaire, et nous y ajoutons le lexique général.

Dans la mesure où les systèmes d'annotation morphosyntaxique varient d'un étiqueteur à l'autre, la mutualisation des ressources nous a permis d'éprouver dans un souci d'objectivation les variations stylistiques du genre de l'article au moyen des variables que nous utilisons et de celles utilisées par C. Poudat (2006a). L'ensemble se compose donc de 97 variables extraites du jeu d'étiquettes du système Cordial (version 8 de Cordial Analyseur, Synapse développement), et de 140 descripteurs dédiés à l'étude du discours scientifique⁴, étiquetés par TnT (Trigrams'n'Tags, Univ. Saarlandes, DE, Brants) – et manuellement vérifiés. À côté des variables morpho-syntaxiques de Cordial traitées dans notre chapitre précédent (et incluant les ponctuations), les variables du second jeu de descripteurs sont morphosyntaxiques, voire sémantiques, et incluent des caractéristiques supposées du discours scientifique (types de connecteurs, *il* impersonnels *vs.* anaphoriques, modaux, symboles, indices de structuration et de renvois dans le texte, etc.) et linguistique (* et ?, signes d'acceptabilité linguistique, abréviations linguistiques comme *SN*, *SV* ou *GN*, etc.).

Nous avons adjoint à ces variables morpho-syntaxique le lexique transdisciplinaire précédemment étudié, qui sert la mise en scène de la recherche dans ses composantes et ses procédures, et qui, par hypothèse, est transversal aux sous-domaines considérés (soit 100 descripteurs lemmatisés, du type *hypothèse*, *méthode*, *résultat*⁵). Enfin, nous avons ajouté les mots signifiants des textes, bien que nous ayons privilégié des variables *a priori* indépendantes de la thématique des textes. La question du rapport entre le style et les sous-domaines scientifiques dont relèvent les articles et leurs auteurs demandera toutefois à être discutée.

On obtient donc quatre jeux de catégories, les variables Cordial, les variables TnT, le lexique et le lexique transdisciplinaire.

¹ Sur ce point sans doute plus méconnu, cf. P. Léon (1971), et, plus récemment, l'étude de H. Schlichter (1999) qui s'intéresse aux aspects phonétiques et acoustiques des imitations d'hommes politiques.

² On peut citer la tentative originale de V. Beaudoin et F. Yvon (2004) d'identifier genres et auteurs à partir du décompte métrique des vers qui les caractérise.

³ Notamment D. Malrieu et F. Rastier (2001), C. Poudat, (2003).

⁴ C. Poudat, 2006a.

⁵ Cf. la partie 5.3. sur le lexique transdisciplinaire

6.3.3. Analyses et méthodes statistiques

Ces quatre jeux de catégories ont servi de fondement à plusieurs analyses statistiques, qui consistent à comparer les observations faites sur les textes, pour déterminer :

- les spécificités aux niveaux de la morphosyntaxe, du lexique et du lexique transdisciplinaire des textes de chaque auteur par rapport aux autres textes ;
- le degré de spécificité de ces styles en termes de classification et de positionnement des textes en fonction des variables morpho-syntaxiques et des termes transdisciplinaires : il s'agit d'évaluer en fonction de la partition des textes en quoi un style est homogène (les textes d'un même auteur présentent des caractéristiques suffisamment homogènes pour être rassemblés dans une même classe) et en quoi il est singulier (les textes d'un même auteur sont isolés des autres dans une classe qui leur est exclusivement dédiée).
- les axes de contraste internes au corpus « auteur » et au corpus « genre », axes factoriels constitués de la corrélation de traits textuels et rendant compte de la diversité interne des corpus indépendamment de la variable stylistique : on cherche à voir si les facteurs qui structurent les différences entre les textes restent stables quel que soit le corpus considéré.

Ces analyses ont été effectuées à l'aide du logiciel DTM (*Data and Text Mining*), logiciel académique développé par L. Lebart¹, et sur lequel se fonde la méthodologie d'analyse de l'article dans l'étude de C. Poudat (*op. cité*).

Spécifiquement dédié à l'analyse de données textuelles, ce logiciel entend répondre aux objectifs du *Text Mining*, que l'on peut traduire par « fouille des textes » ou « analyse statistique exploratoire des données textuelles ». Le *Text Mining* se fonde généralement sur l'analyse multidimensionnelle des textes. Ses objectifs recouvrent les aspects lexicométriques (étudier la richesse du vocabulaire en particulier), l'analyse descriptive, qui permet de définir des proximités entre les formes et les documents et de réaliser des typologies et enfin la construction de modèles statistiques capables de prédire l'appartenance d'un nouveau document à une catégorie déjà définie.

Directement issu du *Data Mining*, le *Text Mining* présente une démarche similaire ; sa particularité réside dans les étapes spécifiques de préparation des données dans la mesure où elles nécessitent un travail de structuration, consistant à extraire du texte des unités statistiques (Benzécri, 1970). Une fois le travail de préparation des données effectué, le *Text Mining* repose sur une analyse descriptive des données multidimensionnelles, qui intègre des méthodes statistiques traditionnelles et des outils de visualisation, et mobilise ainsi des fonctionnalités usuelles d'un outil de *Data Mining*. Deux grandes familles de méthodes sont employées de manière complémentaire par le logiciel DTM, comme dans toute entreprise de *Text Mining* (Benzécri et al., 1981, 1992) :

- les méthodes factorielles qui consistent essentiellement en des représentations graphiques caractérisant les liens entre les différents critères d'analyse (notamment, l'analyse en composantes principales l'analyse factorielle des correspondances, etc.) ;
- les méthodes de classification qui produisent un regroupement d'objets ou d'individus décrits par un certain nombre de variables ou de caractères (classification de type nuées dynamiques, classification ascendante hiérarchique), et servent donc l'analyse descriptive comme la prédiction sur l'appartenance d'un texte à une classe.

Par ailleurs, les programmes développés par L. Lebart dans DTM ont ceci d'intéressant qu'ils se veulent constamment évolutifs au regard des savoirs nouveaux en statistique ; le logiciel existe ainsi en diverses versions en fonction des traitements statistiques qu'il intègre. Les plus récentes perfectionnent les méthodes d'analyse des données

¹ <http://www.enst.fr/egsh/lebart/>

multidimensionnelles évoquées *supra* en utilisant des procédures de validation des visualisations graphiques produites¹. Il s'agit d'une part de procédures de validation externe, qui mobilisent la méta-information disponible sur les données (i.e. des informations externes sur le corpus, mais qui ne figurent pas dans le tableau de données sur ce corpus, par exemple des informations dictionnairiques sur un terme, ou encore des variables externes ou indépendantes) ; d'autre part, DTM intègre des techniques de validation internes, techniques récentes essentiellement fondées sur des méthodes de rééchantillonnage, en particulier la méthode Bootstrap sur laquelle nous reviendrons.

DTM présente enfin un avantage d'ordre ergonomique, celui de fournir les résultats des entreprises de classification sous la forme de représentations graphiques qui intègrent une description automatique des classes : l'interface graphique est telle qu'un simple pointage de la classe permet de faire apparaître les variables entrant dans sa caractérisation. Il dépasse ainsi les limites des analyses factorielles plus traditionnelles, qui demeurent très difficilement interprétables dans les études fondées sur un grand nombre de variables, car elles requièrent de l'analyste de retrouver *a posteriori* les variables constitutives de ces facteurs.

Si l'utilisation d'un logiciel spécifiquement dédié à la mise en œuvre des méthodes statistiques multidimensionnelles était requise, les analyses fournies par DTM ont été relayées pour certains aspects de l'étude par des tests statistiques réalisés avec le logiciel Statview (SAS Institute INC, 1992-1998) : il produit des statistiques générales, qui décrivent les observations faites en corpus (statistiques descriptives) et les comparent (statistiques inférentielles) en vue de déterminer la significativité d'une variable externe – en l'occurrence la variable auteur – sur la distribution des mesures (i.e. en termes d'effets observables sur les textes).

Complémentaires, les différentes analyses effectuées permettent d'étudier le style et ses relations au genre sous divers aspects et de corroborer ou de spécifier les résultats obtenus en les croisant. Il s'agit donc de répondre aux limites intrinsèques des analyses statistiques des données textuelles, souvent sujettes à caution quant à leur pertinence et leur degré de validité. Le choix de DTM, produit direct de l'implémentation des travaux de L. Lebart, nous a permis de bénéficier de l'expertise de son concepteur, au niveau des traitements de données à effectuer, des techniques d'analyse à mettre en œuvre et de la validation des visualisations produites par les méthodes factorielles et les méthodes de classification. Le recours à Statview est moins productif pour l'analyse du style, puisqu'il ne vise pas la classification des textes et la visualisation, mais il sert à associer aux procédures de validation implémentées dans DTM une validation fondée sur d'autres méthodes statistiques plus générales².

Comme résumé dans le tableau ci-dessous, nous procéderons en trois étapes. La première question abordée consistera à voir de quoi sont faits les styles : on propose de mettre en évidence les plus discriminants des traits caractéristiques, obtenus par la caractérisation différentielle des styles au sein du corpus « auteur », et par comparaison statistique avec le comportement de ces variables dans le corpus « genre ». On s'intéressera alors à la question de savoir dans quelle mesure les styles se distinguent, question que l'on aborde selon deux

¹ L. Lebart (2004) regrette que cette validation soit encore peu en vigueur dans la production des visualisations graphiques en Text Mining, faute selon lui de procédures disponibles et de demande pressante de la part des utilisateurs.

² Nous avons eu recours au test d'inférence statistique de Kruskal Wallis. Cf. le chapitre 3 pour des détails sur ce test non paramétrique, qui permet d'évaluer l'effet d'une variable indépendante à plus de 2 variantes (ici la variable « style » : 15 auteurs) sur une variable linguistique continue.

démarches, une entreprise de classification réalisée à partir de l'ensemble des descripteurs sur le corpus auteur, et une validation de cette analyse et de sa visualisation par la méthode du Bootstrap appliquée avec un des jeux de variables (jeu TnT) sur les deux corpus « auteur » et « genre ». Au terme de cette analyse des styles, on reviendra sur le genre pour évaluer sa permanence en fonction des auteurs des articles : il s'agira d'examiner dans quelle mesure le genre reste stable du point de vue de ce qui est commun aux textes et du point de vue des axes de contraste structurant la diversité de l'article en SCL, pour discuter finalement la variation stylistique en regard de la variabilité du genre.

De quoi sont faits les styles? (partie 3.4.)
Caractérisation différentielle des textes du corpus "auteurs" (DTM)
Validation des traits discriminants les textes d'un auteur par comparaison avec le corpus "genre" (Statview)
(Variables Cordial, TnT, lexique transdisciplinaire et lexique général)
Le degré de distinctivité des styles (partie 3.5.)
Classification des textes du corpus "auteurs" avec (DTM - toutes variables)
Validation des classes avec la méthode du Bootstrap (DTM - variables TnT)
Retour sur le genre et sa stabilité (partie 3.6.)
Analyse en composantes principales dans le corpus "genre" et dans le corpus "auteurs" (DTM - variables TnT)

Figure 6.1. Vue synoptique de l'analyse des styles et de la diversité interne du genre.

6.4. De quoi sont faits les styles ? Les traits stylistiques discriminants

Pour décrire les spécificités stylistiques des 15 auteurs retenus, nous avons soumis le corpus « auteurs » à une analyse en axes principaux (analyse en composantes principales et analyse des correspondances) et à une classification ascendante hiérarchique : il s'agit d'appliquer *via* DTM les outils de calcul et de visualisation des tables de données pour montrer quelles sont les caractéristiques des textes d'un même auteur, i.e. ce qui les distingue des autres textes du corpus et fait leur spécificité, au niveau de la morpho-syntaxe, du lexique, et du lexique transdisciplinaire.

Nous avons alors complété cette caractérisation différentielle des styles au sein du corpus « auteurs » par une validation dans le corpus « genre » : dans Statview, on compare les textes d'un même auteur aux 330 textes du corpus « genre », pour voir si l'on retrouve dans les variables qui les distinguent¹ les traits de son style mis en évidence par la comparaison des textes des 15 auteurs. Seules les variables validées en tant que caractéristiques des textes de l'auteur au regard des normes du genre sont considérées comme des traits stylistiques.

Une fois les spécificités stylistiques d'un auteur identifiées dans DTM et précisées par Statview, on détermine sur quelles caractéristiques les styles s'appuient préférentiellement, comme autant de lieux privilégiés de la variation stylistique du genre.

En annexes (annexes 4) est fourni un inventaire recensant les spécificités stylistiques mises en évidence pour les textes de chacun des auteurs. Plutôt que de décrire ici chacun des styles, nous proposons de les aborder de manière transversale, sous l'angle des variables discriminantes qui les caractérisent. Nous montrons de quoi sont faits les styles et ce qu'ils peuvent dire de la diversité interne du genre en nous intéressant successivement au rôle joué par le lexique, par les traits morpho-syntaxiques et par le lexique transdisciplinaire dans la

¹ Une variable distinctive est considérée en tant que telle dans Statview lorsque le résultat du test statistique montre au niveau de cette variable une différence significative entre les deux groupes de données (les textes d'un auteur *vs* les textes du corpus « genre »), différence qui ne saurait être liée au hasard de l'échantillonnage.

caractérisation différentielle des styles au sein du corpus « auteur » et par comparaison avec le corpus « genre ». Auparavant, il nous faut signaler un problème de biais, qui résulte de la comparaison des textes nécessaire à l'étude des styles.

6.4.1. Un biais dans l'analyse des styles

Le style est défini en termes d'écart, ici statistiquement testés, qui font la spécificité des textes d'un même auteur par rapport aux textes du même genre ; en conséquence, les spécificités mises en évidence posent la question de savoir si on a effectivement affaire à des traits de styles ou à des biais résultant de la méthode d'analyse. En effet, parmi les traits discriminant les textes d'un même auteur par l'usage marqué qu'il en fait (usage significativement plus abondant ou moins abondant par rapport aux textes du même genre), nous remarquons des caractéristiques qui sont liées d'une part à la thématique spécifique des textes, d'autre part aux exemples.

Nous avons précisé que la méthodologie adoptée pour l'étude des styles nous a fait préférer des variables *a priori* indépendantes de la thématique des textes, hormis pour le lexique général (l'ensemble des mots signifiants) ; par ailleurs nous avons supprimé des textes les exemples décrochés dispositionnellement (i.e. détachés dans un paragraphe spécifique), mais nous avons déjà montré dans nos précédentes parties d'analyse que le matériau de l'article reste inextricable. Comme nous le verrons tout au long de cette partie, la caractérisation automatique des textes et leur classification intègrent nécessairement des données sur la thématique et sur le matériau étudié dans ces textes.

L'exemple le plus frappant en est sans doute celui de A13 : *folle* et *vache* constituent ainsi les premières places des variables positives discriminant son style. Viennent ensuite *scientifiques*, *Libération* et *science*, où l'on retrouve le poids du corpus (*Libération*), conjoint à celui de l'objet étudié (discours médiatiques et scientifiques) : l'auteur analyse le traitement journalistique, en l'occurrence dans *Libération*, d'objets comme la *vache folle*.

Si le lexique des textes intègre des données sur sa thématique et son matériau, le poids du matériau dans les textes se manifeste également au niveau morphosyntaxique. Certains styles se caractérisent en effet par le poids de variables marginales au niveau du genre, par exemple les termes identifiés comme étrangers, qui représentent la caractéristique la plus forte des textes de A4 : l'auteur travaille sur l'ancien français et cite de longs passages d'œuvres de l'époque.

Les marqueurs d'oralité, tels que les exclamatifs, les interrogatifs, et les marques de 2^{ème} personne sont un autre exemple, mieux représenté dans le corpus, de ces variables qui parce qu'elles sont marginales dans le genre ressortent comme des caractéristiques marquées de certains styles et de certains textes. C'est le cas notamment du style de A2 et de A6.

Le style de A2 se distingue par l'usage abondant que l'auteur fait des pronoms *tu* et *vous*, des disjoints *toi* et *vous*, et des clitiques *vous* et *te*, ainsi que des points d'interrogation, des interjections et des slashes, autant de marques qui semblent tenir principalement à la présence de segments relevant du corpus oral étudié par l'auteur.

De manière similaire, le style de A6 se distingue par les points d'exclamation, les deux-points et les tirets et par un usage fort de la 1^{ère} et de la 2^{ème} personne du pluriel, qui va de pair avec une proportion forte de verbes à ces personnes, caractéristiques qui sont à comprendre comme un effet du corpus conversationniste employé par l'auteur. Ainsi, le poids du corpus dans la caractérisation des styles justifie l'extraction des exemples des textes que nous avons effectuée, mais en montre aussi les limites.

On serait tenté bien entendu de conclure qu'il faut mettre de côté ces biais, pour s'en tenir à une description du genre et de ses usages singuliers qui soit indépendante de la

spécificité thématique des textes et de la spécificité de leur matériau. Le problème est qu'une étude des styles et des genres qui entend se fonder sur les caractéristiques textuelles est forcée d'admettre que les textes intègrent des traits liés à leur thématique et à leur matériau.

L'analyse fait ressortir ces traits dès lors qu'ils sont marqués par rapport aux caractéristiques du genre ; ainsi, si les traits de l'oral ressortent comme une caractéristique marquée des textes qui utilisent un tel corpus, il n'en va pas de même par exemple pour un article qui prendrait comme matériau des extraits d'articles scientifiques. De fait, il importe d'évaluer ces biais en tant que tel, en reconnaissant que ce qui fait la spécificité de certains usages du genre (que ces usages soient le fait d'un auteur ou de plusieurs auteurs du champ) peut tenir d'une part à des spécificités thématiques marquées, et d'autre part à des spécificités liées à un matériau dont les caractéristiques linguistiques sont marquées par rapport aux caractéristiques du genre.

Cela étant, on conviendra aussi qu'il n'est pas possible d'éliminer complètement ces biais, sauf à modifier profondément les caractéristiques des textes alors que ce sont elles que l'on cherche à cerner ; l'enjeu reste de s'intéresser aux caractéristiques stables et aux caractéristiques marquées du genre, de discuter de ce qui fait sa diversité (a-t-on affaire à des styles d'auteur à proprement parler ? A un effet du matériau ? etc.), et de voir ce que les usages singuliers (ou caractéristiques marquées dans l'usage que chaque auteur fait du genre) peuvent nous dire du genre quant à sa stabilité. C'est ce que va nous permettre d'aborder la mise en évidence d'usages singuliers du genre au niveau lexical, usages fortement liés aux thématiques des textes comme nous l'avons signalé.

6.4.2. Style et lexique : une entrée pour l'étude des concepts et des épistémologies

Le lexique joue un rôle relativement important dans la caractérisation d'usages singuliers du genre (i.e. des styles). On l'a dit, les traits lexicaux discriminants montrent le poids du matériau (cf. *vache* et *folle* dans le style de A13) et des thématiques des textes. Encore faut-il que le corpus et la thématique soit suffisamment spécifiques par rapport aux traits du genre, et qu'ils ne soient pas le fait d'un seul texte de l'auteur, mais caractérisent l'ensemble (ou une grande partie) de ses textes par rapport aux textes du même genre.

Ainsi le style de A12 manifeste-t-il sa spécificité par les concepts qu'il mobilise dans ses textes : *sémiotique*, *politique*, *symbolique*, *espace*, *public* sont des caractéristiques qui distinguent positivement ses textes des autres textes du corpus, par l'usage abondant que l'auteur en fait. Ces variables lexicales positives rendent clairement compte des principaux thèmes abordés par l'auteur, qui semblent très spécifiques dans la mesure où la plupart des autres auteurs sont caractérisés négativement par ces mêmes thèmes. Ainsi, *politique* et *sémiotique* caractérisent positivement les seuls textes de cet auteur et représentent respectivement les caractéristiques négatives de 8 et 11 auteurs.

Tous les auteurs sont loin d'être concernés par l'importance du lexique en tant que caractéristique positive de leur style. Si A12 est le cas le plus frappant, A4 se distingue également par l'usage important de *syntagme*, *proposition*, *grammaticalisation* et *topique* notamment, et A3 par celui de *langues*, *langue*, *français*, *anglais*, *pays*, autant de termes qui reflètent encore une fois les thématiques privilégiées des auteurs dans l'ensemble (ou une grande partie) de leurs articles.

On peut faire l'hypothèse que les auteurs dont les styles se caractérisent par des termes essentiellement rattachés à la dimension thématique et conceptuelle des textes ont des domaines d'investigation relativement à la marge dans le corpus. Ces termes sont des candidats-concepts : parmi l'ensemble des mots signifiants, ils représentent un lexique de spécialité ; cette spécialité est suffisamment rare dans le corpus pour être identifiée comme

une caractéristique marquée des textes de l'auteur par rapport à l'ensemble des articles du corpus « auteur » et du corpus « genre ».

Une telle approche, menée sur des corpus de taille plus importante, doit ainsi permettre de fonder en corpus une cartographie du champ disciplinaire, les concepts discriminants donnant à voir dans les caractéristiques positives et négatives des textes des paradigmes qui structurent le corpus, et, partant, la discipline. Pour revenir à l'exemple le plus parlant, le caractère particulièrement discriminant des termes *politique* et *sémiotique*, sur-représentés chez A12 et sous-représentés chez les autres auteurs, signale la place marginale des problématiques politique et sémiotique dans la recherche linguistique.

Le graphique qui suit rend compte des mots significatifs des textes (nom, verbes, adverbess et adjectifs) sur les deux premiers axes factoriels. Il permet ainsi de voir où se positionnent les auteurs par rapport aux termes qu'ils emploient : la singularité de A12 (*politique*, *sémiotique*, etc.) est particulièrement frappante (facteur 1), et A4 (*grammaticalisation*, *topique*, etc.) se démarque également en positif sur le facteur 2 :

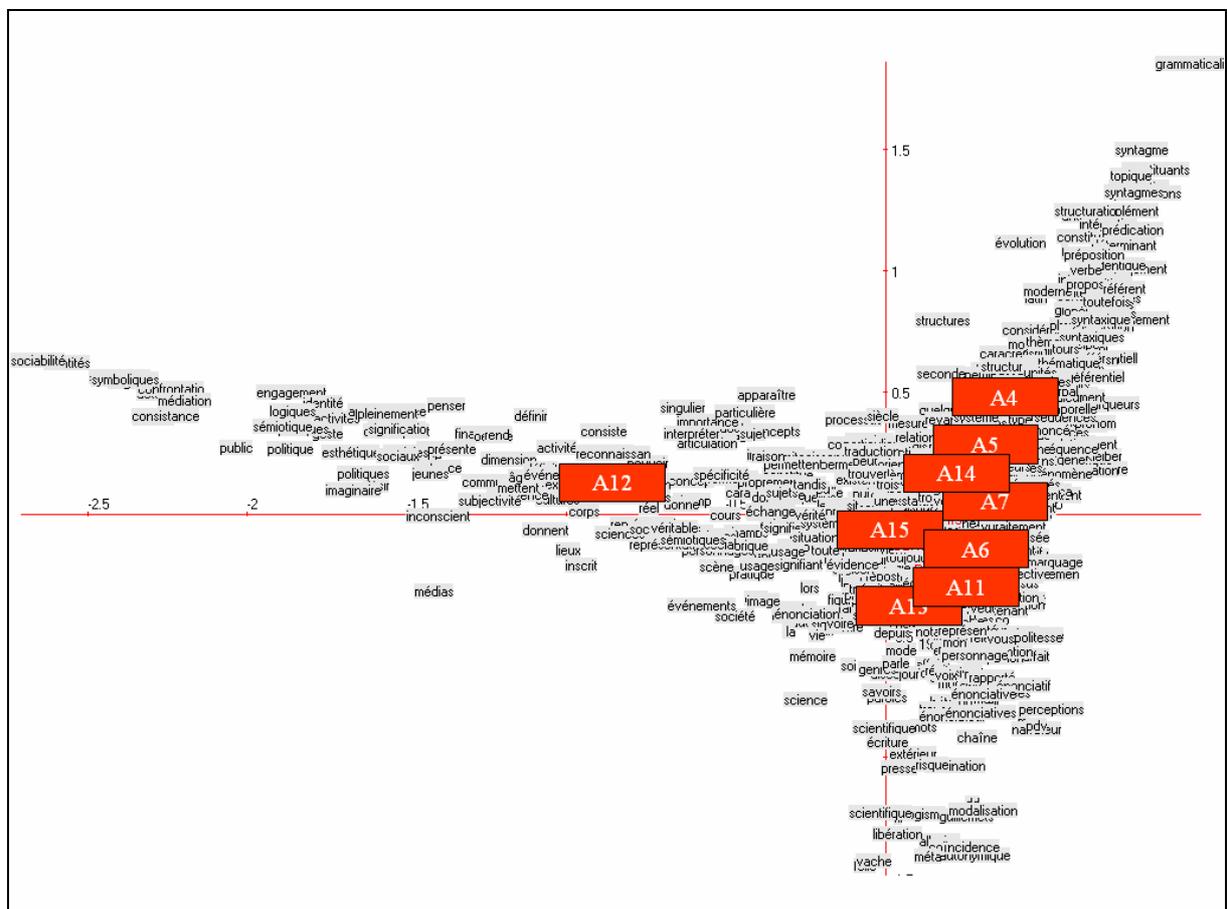


Figure 6.2. Styles d'auteur et lexique.

La distribution des mots significatifs sur le graphique suggère trois tendances opposées qui pourraient renvoyer à des pôles de recherche en linguistique – sans doute de manière quelque peu caricaturale ici, car il faudrait augmenter encore le corpus. On voit en effet se distinguer :

- les approches grammaticales (en haut, à droite), avec des termes tels que *syntagme*, *syntactique*, *structures*, *constituant*, *déterminant*, etc., et le rôle d'attracteur joué par *grammaticalisation*, tout en haut de ce pôle,
- l'analyse de discours (en bas, à droite), avec les termes *énonciation*, *modalisation*, *science*, *écriture*, ou encore *dialogisme*, *politesse*, *point de vue*, etc.,

- et les analyses politiques et sociales, qui se distinguent sur le facteur 1 (à gauche), avec *identité, sociabilité, engagement, politique, jeunes*, etc.

A titre de piste pour des analyses ultérieures, l'analyse de la diversité interne du genre au niveau lexical doit donc potentiellement permettre de rendre compte des thématiques prisées dans un champ, en l'occurrence celui des sciences du langage¹, et de la manière dont il se structure sur la base de tendances adverses.

Pour en revenir aux styles, intéressons-nous à présent aux traits morpho-syntaxiques ; si A12 est l'auteur le plus singulier au niveau lexical, il semble également se distinguer nettement à ce niveau-là, ce qui suggère qu'il existe des degrés divers de distinctivité des styles, et nous reviendrons sur cette question au cours de notre analyse.

6.4.3. *Style et morphosyntaxe*

Au niveau morpho-syntaxique se manifestent, comme nous l'avons montré, des spécificités liées à un matériau dont les caractéristiques linguistiques sont marquées par rapport aux caractéristiques du genre (les corpus oraux conversationnalistes). Si l'on exclut cet aspect-là, il reste cependant à discuter ce qui fait que les textes d'un même auteur présentent une morpho-syntaxe spécifique par rapport aux autres textes du corpus « auteur » et du corpus « genre » : a-t-on affaire à des styles d'auteur à proprement parler ? Nous allons voir que les caractéristiques morpho-syntaxiques entrant dans la caractérisation des styles permettent d'envisager que la spécificité des usages du genre peut tenir d'abord à des types d'études (et plus précisément à la dimension historique du texte ou à la part prise par l'exemplification), et également à des particularités d'ordre idiolectales.

Des usages du genre marqués par le type d'études menées par l'auteur

Des textes comportant une dimension historique

L'étude des styles au niveau morpho-syntaxique rejoint ce que nous avons mis en évidence dans notre précédent chapitre, à propos des usages contrastés du genre. On remarque en effet que les temps verbaux sont particulièrement discriminants en matière de styles, et plus particulièrement les temps du passé. Ceux-ci se trouvent associés à d'autres indices d'un mode historico-narratif marqué par rapport aux tendances centrales du genre, et caractéristique des textes de certains auteurs. Nous en traiterons à partir des deux cas de A9 et A3.

Les textes de A9 doivent leur singularité à un usage important de l'imparfait et du plus-que-parfait, du passé simple, ainsi que des dates et des noms propres. Ce mode historico-narratif n'est finalement pas étonnant, car l'auteur est spécialiste de linguistique historique. Le présent est en revanche sous-représenté (un tiers de moins en moyenne pour le présent et plus de deux fois moins de modaux au présent que dans les autres textes), et nous retrouvons donc avec les styles l'opposition entre présent d'une part et imparfait et passé simple de l'autre mise en évidence dans l'analyse du genre menée au chapitre 5.

On relève également dans les textes de A9 une proportion plus importante de marques de troisième personne (et plus précisément de *il* anaphoriques et de déterminants possessifs), qui nous semble caractéristique des textes historiques. Ils signalent en effet une progression

¹ C. Poudat (2006b) développe cette idée en visant à établir en corpus une typologie des concepts de linguistique.

thématique spécifique, dans des textes qui se focalisent sur un acteur ou un événement et en font un thème plus constant à partir duquel s'organise le texte

Il y a un cas Albert Dauzat. Soixante années durant, **il** fut un des acteurs majeurs du champ de la linguistique et jamais **il** n'y occupa une position centrale. Les témoignages d'admiration et de gratitude ne **lui** ont pas été ménagés et pourtant **il** reste en marge, par **ses** origines et son cursus, par **ses** objets de prédilection qui n'étaient pas de ceux que l'Université reconnaît volontiers. C'est que, si Dauzat est un universitaire, **il** l'est en dehors des lieux de reconnaissance académique, hors des facultés de lettres et d'abord de la première d'entre elles, la Sorbonne. **Il** n'est parvenu ni au Collège de France ... [A9-074].

Le mode historico-narratif peut cependant être indépendant du passé simple, comme nous l'avons vu dans notre précédente partie et comme nous le voyons avec le style de A3 (déjà signalé pour son usage de *langue, langues, français, anglais*). L'auteur emploie à l'instar des textes journalistiques de très nombreux imparfaits narratifs :

La séparation et la non communication entre les approches de Saussure et dans une moindre mesure de Meillet, d'une part, et celle de la Grasserie ou de Van Gennepe d'autre part, nous montrent que la linguistique structurale dominante **s'est construite** en marginalisant les faits qui n'**entraient** pas dans le cadre ou dans le modèle. C'est-à-dire qu'elle se **donnait** la langue qui l'**arrangeait** : ou bien les faits **rentraient** dans le modèle ou bien on les **jetait** (ou on les **rejetait** vers la psycholinguistique, la sociolinguistique, approches considérées comme périphériques).[A1-001]

La singularité des textes ainsi mise en évidence montre un style historique qui tient à l'objet de l'auteur (une analyse historique du champ de la linguistique). Mais nous avons vu aussi¹ que le mode historico-narratif ne s'explique pas seulement par l'objet : il requiert certes une analyse rétrospective, mais aussi un traitement historicisant de l'objet.

Si A9 et A3 sont tous deux concernés par ce style historique, l'analyse des occurrences de verbes aux temps du passé permet d'en dire davantage quant à leur domaine de spécialité. L'analyse historique du champ de la linguistique caractérise mieux les textes d'A9 que ceux de A3, qui concernent les politiques linguistiques. S'ils intègrent une analyse historique du champ de la linguistique comme nous l'avons vu avec l'exemple cité *supra*, ils manifestent également un usage d'un mode narratif dans d'autres circonstances, par exemple pour traiter de la situation plurilingue de l'Égypte à la charnière du dix-neuvième et du vingtième siècle (A3-002), ou du statut des langues dans les relations internationales (A3-003).

L'auteur privilégie donc des objets qui comportent une dimension historique et affectionne leur traitement historicisant. Un style historique, lié à un type d'étude et son traitement, peut donc se présenter comme un style d'auteur, en termes de pratiques spécifiques de la recherche et en termes de pratiques scripturales. Chez A3, la dimension historique signalée par l'usage de l'imparfait va ainsi de pair avec un usage fort des noms propres comme chez A9, mais contrairement à ce que l'on a noté chez ce dernier, les dates ne sont pas discriminantes. A3 recourt abondamment aux connecteurs de temporalité reliant les événements décrits, et ses textes comprennent par ailleurs peu de niveaux de structuration et de listes. Dans le même sens, si A3 et A9 ont en commun l'imparfait, le style du second associe à ce temps le passé simple, que l'on retrouve comme spécificité de A4.

Des textes diversement exemplifiés

Si les styles de A3, A4 ou de A9 attestent d'un style historique dans les textes de ces auteurs, d'autres styles sont marqués d'un point de vue morpho-syntaxique par des traits qui suggèrent un recours abondant à l'exemplification.

Ainsi en est-il du style de A1 : on observe dans ses textes significativement moins de numéraux ordinaux et de numéraux de datation mais un usage plus abondant de numéraux de

¹ Cf. la partie sur les temps verbaux au chapitre 5.

renvoi. Cette tendance résulte manifestement d'un usage abondant d'exemples numérotés dans ses textes. La forte présence de tirets serait alors à mettre au compte des indices de structuration qui servent à introduire les exemples, et c'est ce que l'on retrouve dans le style de A6, dont les textes présentent de nombreux extraits du corpus étudié par l'auteur (corpus oral que nous avons évoqué à titre de biais dans l'analyse des styles). Le style de A8 est concerné également par les numéraux à fonction de renvoi qui s'associent à un usage important d'acronymes (de type *SN*, *SV*), et d'abréviations linguistiques (*prep.*), et l'examen de ses textes confirme l'importance donnée par l'auteur à l'exemplification et à l'analyse syntaxique et énonciative de ces exemples.

Chez A1, cette large place faite à l'exemplification se manifeste aussi par l'abondance de connecteurs d'exemplification. Ces derniers cependant ne sont pas indiciels en eux-mêmes d'un taux fort d'exemplification dans les textes ; le cas de A15 est à ce titre frappant : le style de l'auteur se distingue par un usage massif de ce type de connecteurs (deux fois plus en moyenne que dans le reste du corpus), mais à y regarder de plus près, on ne trouve dans ces textes que de très rares exemples, si l'on s'en tient du moins à une définition des exemples restreinte à la mention autonymique d'unités linguistiques. Nous en avons parlé dans le cadre de l'étude des exemples au chapitre 4, l'usage de *par exemple* dans les textes de cet auteur permet de souligner la complexité de ce que l'on entend par « exemple » dans l'article en linguistique et dans l'écrit de recherche de manière générale.

Or, un genre définit précisément un rapport normé entre signifiant et signifié au palier textuel : par **exemple**, dans le genre de l'article scientifique, le premier paragraphe, sur le plan du signifiant, correspond ordinairement à une introduction, sur le plan du signifié ; dans le genre de la nouvelle, il s'agit le plus souvent d'une description. [A15-088]

Ainsi, les différents marqueurs d'exemplification suggèrent, en fonction de leur nature et de leurs combinaisons dans les textes, des tendances quant aux types d'études privilégiées par un auteur : une association de connecteurs d'exemplification et d'indices de structuration comme les numéraux de renvoi et les tirets signalent une approche plus descriptive chez A1, qui cite des énoncés à titre d'autonymes (de même que A2 et A8 évoqués *supra*) tandis que le recours à des formules du type « par exemple » alimente chez A15 une discussion théorique et un positionnement doctrinal qui ne mobilisent pas de référence à des énoncés, mais plutôt à des observations faites sur les usages de la langue.

L'analyse des caractéristiques marquées dans l'usage qu'un auteur fait du genre se présente ainsi comme une entrée intéressante pour dépasser une analyse de l'exemple restreinte aux autonymes. Elle signale des tendances régissant la diversité interne du genre en fonction de ce qui est pris comme matériau par l'auteur. Il est possible ainsi d'affiner l'opposition entre des textes plus théoriques et des textes plus empiriques. Nous avons discuté des limites qu'il y a à cerner cette opposition avec le seul trait distinctif des exemples autonymes, ou avec l'exemple *versus* la citation (cf. chapitre 4) ; on voit ici que des critères multiples interviennent, tels que :

- certains textes se distinguent par la trace de leur corpus (A1 pour le recours à des exemples d'énoncés, A2 pour l'oral, ou encore A13, évoqué à propos de *vache* et *folle*) ;
- d'autres par l'analyse d'exemples (recours à des exemples numérotés et formalisation linguistique chez A8) ;
- d'autres enfin, comme c'est le cas chez A15, par des exemples qui ne correspondent pas à des occurrences des phénomènes langagiers mais à des résultats de travaux de recherche, en cela comparables sur le plan fonctionnel à des citations.

Des styles au sens de spécificités idiolectales

Les spécificités stylistiques observables au niveau morpho-syntaxiques ne tiennent pas cependant qu'au type d'étude privilégié par l'auteur. Elles attestent semble-t-il aussi de pratiques idiomatiques du genre, montrant comment le style en sciences rejoint une acception littéraire. On le voit au rôle discriminant des ponctuations, des aspects énonciatifs, et de traits qui semblent pouvoir être rattachés à la question de l'argumentation.

Les ponctuations, la complexité syntaxique et les relations logiques

Le rôle particulièrement discriminant des ponctuations dans la caractérisation différentielle des styles s'explique en partie comme un biais lié au matériau étudié par l'auteur (ponctuations énonciatives comme par exemple les exclamatifs), en particulier le matériau oral, dont les caractéristiques linguistiques ressortent comme caractéristiques marquées par rapport à celles de l'article.

Cependant, les ponctuations peuvent constituer un trait stylistique indépendant du corpus de prédilection des auteurs ; l'usage de la virgule chez A12 en est un exemple, qui montre que les ponctuations sont un trait textuel particulièrement susceptible de varier avec l'auteur. A12, dont on a parlé à propos de l'usage abondant dans ses textes des concepts marginaux de *politique* et de *sémiotique*, utilise 1,5 fois plus de virgules que la moyenne des autres textes, ce qui confère à ses textes un rythme spécifique.

C'est par la main que le corps signifie, car c'est la main qui, à la fois, établit la relation entre le corps et les choses ou entre mon corps et celui de l'autre, et permet d'orienter *le corps*, de lui donner son sens, puisque c'est par rapport à elle que mon corps distingue la gauche et la droite. La main, en ce sens, *institue le code du corps*, c'est elle qui fait du corps une forme symbolique, que l'on va pouvoir lire, interpréter, à laquelle on va pouvoir reconnaître une signification en en reconnaissant l'orientation.
[A12-091]

On remarque également dans son style un emploi plus important de points, alors que les phrases de l'auteur sont de taille plus importante que chez d'autres ; s'il n'est pas lié à la taille des phrases, cet usage marqué peut donc s'expliquer par les proportions plus faibles de points-virgules, de deux-points et de points de suspension. Ces deux exemples, A2 et A12, d'usages spécifiques de la ponctuation montrent que le style en sciences ainsi décrit dans notre étude rejoint ce qu'il est convenu d'appeler le style dans le cas des auteurs littéraires. Le style ici ne dépend plus ni du corpus ni de l'objet ni du type d'étude mené par l'auteur, mais bien de « choix » scripturaux, au sens d'un usage personnel, conscient ou non, des possibles de la langue.

Par ailleurs, comme l'auteur affiche une prédilection pour des phrases longues et pour l'usage abondant des virgules, ce profil singulier fait écho à ce que nous avons dit dans le chapitre 5 des différences entre les disciplines et en fonction du statut de l'auteur à ce même niveau de la longueur des phrases et des virgules. Nous avons vu en effet que les articles de LET (*versus* ceux de SCL) et les articles des non-doctorants et des non-doctorants de SCL (*versus* ceux des doctorants et des doctorants de SCL) affichaient cette même prédilection. On peut ainsi affiner encore l'analyse de la diversité des textes, car on voit ainsi que parmi les non-doctorants de SCL, certains affectionnent significativement plus que d'autres les phrases longues et les virgules.

Comme les textes de l'auteur ici concerné, A12, se distinguent également par des concepts très marqués par rapport aux tendances centrales du genre (*politique* et *sémiotique*),

il est permis aussi de voir dans cet usage spécifique de la phrase et des virgules un effet disciplinaire, à l'instar de la différence observée entre les articles de LET et ceux de SCL. Il est possible alors d'avancer dans la définition des styles, et de se demander si le style en tant que spécificité idiolectale ne pourrait pas être compris comme un effet de la formation d'un auteur ; l'ancrage disciplinaire de sa formation déterminerait des pratiques d'écriture spécifiques, et peut-être aussi des pratiques spécifiques de la recherche.

Le profil d'A12 montre avec d'autres que les indices de complexité syntaxique sont un lieu-pivot privilégié de la variation du genre. On peut regrouper dans cette catégorie les indices de longueur de phrases et de proposition (en nombre de mots), et tout ce qui concerne les subordinées et les subordinants. A nouveau, on voit ici comment le style en sciences reste proche des spécificités stylistiques observables dans les textes littéraires : certains auteurs affichent une préférence pour des phrases et des paragraphes courts (comme A5) tandis que d'autres, tel A7 notamment multiplient les phrases complexes, comme le montre un taux fort de phrases comportant une subordinée, de subordinants, d'adverbes et connecteurs toutes catégories confondues et plus précisément, de connecteurs de causalité, de disjonction, de présupposition, et de certitude. Ces derniers signalent également l'importance de la dimension argumentative comme trait qui varie avec l'auteur.

A l'inverse de A7, les textes de A2 (déjà évoqués à propos de leur corpus oral) se distinguent par la présence significativement plus faible de phrases avec une subordinée, de subordinées, de subordinants¹, et de divers types de connecteurs (de conclusion, de disjonction, reformulation, de causalité, d'exemplification). La sous-représentation de ces traits, vérifiée avec les jeux TnT et Cordial, semble pouvoir s'expliquer en lien avec une sur-représentation des deux-points (quasi deux fois plus que dans les autres textes), très employés par l'auteur pour justifier ou exemplifier son propos :

D'autre part, les noms de lieux se référant à la voirie urbaine (*rue*, mais aussi *place*, *boulevard*, etc.) ne sont pas les seuls à introduire un complément illatif/locatif sans préposition : d'autres sortes de lieux servent de support au même type de construction. [A2-011]

Tout texte exprime un *ethos* : il possède une manière de se dire et de s'énoncer qui fait sens (Maingueneau, 1993). [A2-013]

Ce recours fort aux deux-points et la faible présence de connecteurs suggèrent que le style de A2 est ainsi caractérisé par une non explicitation de certaines relations logiques, implicitement marquées par les deux-points². L'usage que cet auteur fait du genre suggère l'intérêt des ponctuations pour analyser la diversité de ses usages, et en particulier de l'argumentation.

Le style de cet auteur se caractérise également par des ponctuations à valeur énonciative, pour partie dus à un biais (le corpus oral), mais pas uniquement : un examen en contexte des points d'interrogation dans les textes de A2 où ils sont particulièrement fréquents montre une pratique récurrente de la question rhétorique, dont nous avons parlé dans notre précédent chapitre. Ce trait marqué est à rattacher au caractère particulièrement discriminant des aspects énonciatifs dans les usages du genre que chaque auteur fait.

¹ Et en particulier de pronoms relatifs.

² Cf. l'analyse que nous avons faite de ce trait qui discrimine en SCL les textes des doctorants et des non-doctorants (chapitre 5).

Les ponctuations, en s'associant à d'autres marques, signalent que la syntaxe est particulièrement soumise à l'usage que chaque auteur en fait, en termes de complexité de la phrase et de relations logiques des propositions. Elles suggèrent aussi l'intérêt de la dimension énonciative des textes pour l'analyse des styles et de la diversité interne du genre.

L'importance de la dimension énonciative dans la singularité de l'écriture de l'article se vérifie ne serait-ce qu'au niveau du choix offert au scripteur pour un mode de prise en charge plus personnel ou plus impersonnel. D'abord, le style d'un auteur, décrit par comparaison de ses textes avec les autres textes du corpus « auteur » et avec ceux du corpus « genre » peut se caractériser par une préférence marquée pour un mode. Ainsi en est-il pour le mode délocuté, avec la sur-représentation de la 3^{ème} personne au niveau des verbes (singulier et/ou pluriel) (dans les textes de A5, A8, A12, A13, A14, A15), ou un usage important du *on* (A13). Inversement, certains auteurs affectionnent le mode personnel, avec l'usage important de *je* (A1 et A2), et de *nous* (A3, A7, A10, A11)¹.

D'autres styles peuvent se caractériser négativement par un usage moindre d'une marque de personne ou de non-personne. C'est le cas par exemple dans les textes de A9 qui délaissent le *on*.

Par ailleurs, certains styles affichent à la fois une préférence pour une marque et une aversion pour une autre. Par exemple, A11 utilise *nous* mais délaisse *je* ; A12 et A15 affectionnent la 3^{ème} personne mais délaissent également *je* ; A13 et A14 affectionnent la 3^{ème} personne et délaissent *nous*². La double tendance peut aller jusqu'à signaler un choix net dans l'alternative entre le mode de prise en charge personnel et le mode délocuté. Ainsi, les textes de A1 se caractérisent positivement par l'usage de *je* et négativement par une sous-représentation de la troisième personne. Dernier fait notable, à l'inverse du choix exclusif d'un mode au détriment de l'autre, chez A4 la sur-représentation des formes de *je* rivalise avec un usage fort également du « il » impersonnel. Un double effet d'effacement et de marquage de la personne dans les textes se manifeste dans l'usage singulier que cet auteur fait du genre.

L'examen des textes des auteurs dont le style se caractérise par un usage fort du *nous*, (A3, A7, A10, et A11) et de ceux où il est sous-représenté (A13, A14) montre l'affinité de cette marque de personne avec les balises méta-textuelles dont nous avons déjà traité : la forte présence de *nous* dans le style d'un auteur pourrait ainsi être liée à l'usage de ces balises, qui prennent la forme d'expression lexicalement figées. On le voit par exemple dans les textes de A3 qui comprennent notamment de très nombreux *nous verrons*.

(14) bilinguisme que l'on retrouve, comme **nous** allons le voir, sur les plaques des rues et des voitures, les enseignes des boutiques, etc. On y trouve tout un discours qui nous³ dit que cette ville a quelque chose de particulier [A3-021]

¹ Il peut s'agir aussi des autres formes du pronom personnel (*me* par ex.), des pronoms et déterminants possessifs correspondants, et enfin des marques flexionnelles de personnes au niveau des verbes.

² Voir la note précédente.

³ Le *nous* ici n'a pas le statut de balise méta-textuelle ; en sa qualité de *nous* générique, il participe d'un rôle d'implication du lecteur plutôt que de guidage. Avec le *on* au début de l'extrait cité, on voit la frontière diffuse de marques qui ne sont pas tant celles de la personne du locuteur que celles d'une implication mutuelle de l'auteur et du lecteur. (Cf. le chapitre 5, et la première partie du chapitre 7 pour de plus amples discussions de cette question).

L'utilisation massive de ce type de balises s'observe également dans des textes où domine « je ». La diversité du genre au niveau des marques de personne montre donc d'abord qu'il existe des textes attestant d'un recours mixte au *je* et au *nous*, constat déjà mis en évidence au chapitre 5 et qu'il nous reviendra d'approfondir dans notre dernière partie d'analyse (chapitre 7).

Par ailleurs, indépendamment du choix pour l'un, l'autre ou un mode mixte d'usage de *je* et de *nous*, le recours à ces balises semble donc en lui-même constituer un trait stylistique. On le voit chez A2 où la forte présence de « je » est corrélée à celle du futur dans les nombreuses annonces que fait l'auteur de ce qui va suivre, en début d'article mais également en son corps (« *je me limiterai...* », « *je prendrai pour exemple...* »). Comme nous l'avons vu, le style de cet auteur se caractérise également par l'usage de points d'interrogations dans des questions rhétoriques ; les prédilections idiomatiques dans ses textes signalent donc un rapport spécifique au lecteur.

Dans le style de A7, des balises méta-textuelles interviennent également. Le *nous* est clairement « writer-oriented » : la personne linguistique renvoie à la personne énonciative du locuteur. Chez cet auteur, le *nous* s'associe aussi, sous la forme du déterminant possessifs (notre, nos) à un lexique transdisciplinaire, dans des mentions récurrentes comme « notre travail », « notre propos », « notre conception », « notre hypothèse ». Les variations dans l'usage des marques de personne, usage particulièrement discriminant d'un auteur à l'autre, rendent ainsi tangible une énonciation diversement représentée dans les textes ; elle peut se caractériser prioritairement par une mise en scène de son auteur, ou à un autre pôle par un rapport didactique au lecteur.

Ces usages singuliers permettent ainsi d'affiner la description morpho-syntaxique du genre (cf. partie précédente) en montrant au-delà du seul emploi de *je* et de *nous* comment les textes sont plus ou moins « writer- » ou « reader-oriented » en fonction des associations syntagmatiques de ces marques de personne.

La dimension argumentative des textes

Le fait que les aspects énonciatifs fonctionnent comme un lieu-pivot privilégié de la variation stylistique du genre montre également que ce qui fait la spécificité des textes d'un auteur caractérise l'usage singulier que chaque auteur en fait peut être lié avec la dimension argumentative des textes ; on l'a suggéré à propos de la syntaxe de la phrase, et on le voit également dans l'usage marqué d'adjectifs réfléchis, qui peuvent dans certains cas signaler une tendance à l'emphase (« *condition même de sa production* », « *l'image qu'elle en dessine en elle-même par ses non-coïncidences* », « *non seulement à repérer exactement mais même à localiser* »).

Cette tendance caractérise le style de A7, dont on vient de montrer l'usage d'un *nous* auto-centré (« *nous rappellerons* », « *notre hypothèse* », etc.). Ces textes se distinguent aussi par les négations, marques de polyphonie indicielles d'une argumentation forte, sinon polémique. Au contraire le style de A2 se distingue par son usage faible (1,5 fois moins en moyenne que dans le reste du corpus) ; sur ce point différent aussi de A7, l'auteur recourt au *je* plutôt qu'au *nous*, et ses textes sont davantage tournés vers le lecteur si l'on en juge à l'usage des balises méta-textuelles.

Les connecteurs rendent compte également de la dimension plus ou moins argumentative et polémique des textes : chez A7, l'importance de la négation s'accompagne d'un usage marqué de connecteurs argumentatifs (notamment des connecteurs de présupposition et de certitude). Le style de A8 se distingue quant à lui par la forte présence de connecteurs d'opposition. Au contraire, ils sont délaissés dans les textes d'A12 (quasi trois

fois moins, alors que ce sont les troisièmes connecteurs les plus représentés du genre), de même que les connecteurs de concession. On pourrait donc y voir une tendance à des textes globalement peu polémiques, les connecteurs de conclusion, de doute, d'addition et de conséquence l'emportant dans la visée explicative et argumentative des textes de l'auteur.

L'usage du conditionnel semble, de manière *a priori* paradoxale, particulièrement prisé des styles dont d'autres traits (négation, connecteurs) suggèrent une dimension argumentative forte, voire tendancielle polémique. L'effet de nuance introduit par le conditionnel peut en lui-même représenter un trait discriminant (dans les textes de A3, A5, et A9) mais l'examen des textes montre qu'il est souvent utilisé pour contrebalancer des assertions fortes,

On voit que le traitement de la polysémie de l'adj. sec en termes d'homonymie **serait** particulièrement desséchant! Car on y **perdrait** tout un échantillonnage de glissements de sens dont nous sommes largement conscients dans l'usage langagier. [A5-047]

marquées, comme ici par la modalité exclamative, ou encore par le mode déontique : de manière générale, parmi les formes au conditionnel dans les textes de ces auteurs, on trouve de nombreuses fois le verbe *devoir*.

On retrouve cette caractéristique dans le style de A4, qui illustre le mieux le rôle de la dimension argumentative dans la discrimination de styles et d'usages distincts du genre. Nous avons vu que les concepts utilisés par A4 sont spécifiques (*grammaticalisation, topique*), mais ses textes se distinguent tout particulièrement aussi au niveau morpho-syntaxique. Ils se caractérisent par un emploi plus important de modaux, et de modaux au conditionnel en particulier¹. Ils indiquent deux tendances dans ses textes, en termes d'argumentation : d'une part, son style indique un penchant pour les assertions fortes, de type déontique (*falloir, devoir*) ; celui-ci va de pair aussi avec un emploi très marqué de connecteurs de justification (seconde caractéristique de ses textes observée avec le jeu TnT). D'autre part, son style manifeste une tendance à la nuance, dont on relève d'autres indices, notamment l'usage globalement important du conditionnel (deux fois plus en moyenne que dans le reste du corpus) et des connecteurs de doute.

En termes énonciatifs, cette double tendance marquée par des corrélations de traits est particulièrement intéressante : elle montre qu'on n'a pas d'un côté une modalisation qui marquerait une nuance dans l'assertion et de l'autre une modalisation qui favoriserait des assertions plus péremptoires, mais qu'au contraire le recours abondant à l'un va de pair avec le recours abondant à l'autre. Or, nous avons vu aussi du style de cet auteur qu'au niveau de la prise en charge énonciative, ses textes se distinguent également par deux tendances opposées, mais là encore non contradictoires : ils se caractérisent à la fois par un usage fort du *je* et par un usage fort également du *il* impersonnel. Un double effet d'effacement et de marquage de la personne dans les textes se manifeste ainsi.

Avec les autres constats faits sur le poids des aspects énonciatifs et polyphoniques dans la variabilité des usages du genre, le style de A4 suggère donc des pistes intéressantes pour l'analyse énonciative des textes, et celle du rôle de la dimension énonciative dans l'argumentation. L'exploitation double, marquée dans son style, de l'effacement et du marquage de la personne d'une part, et de la nuance et du renforcement de l'assertion d'autre part manifeste ce que l'on pourrait envisager comme des coups de force dans l'assertion, dont nous reparlerons dans notre dernier chapitre.

¹ Les modaux au présent et à l'infinitif sont également sur-représentés dans ses textes, cf. le descriptif des styles de chaque auteur en annexes.

Les caractéristiques idiolectales au niveau morpho-syntaxique du genre signalent des rapports spécifiques au lecteur (complexité syntaxique, ponctuation énonciative, balises méta-textuelles), des modes variés d'inscription du scripteur dans des textes plus ou moins auto-centrés (prise en charge énonciative), des spécificités quant aux relations logiques entre propositions (explicitation des relations à l'aide de connecteurs *versus* usage des deux-points) et au caractère plus ou moins polémique de l'argumentation (avec les marques de personne et de non-personne, les marques d'emphase, les marques polyphoniques comme le conditionnel, la négation, certains connecteurs d'opposition ou de concession, etc.). La spécificité des pratiques scripturales montre ainsi l'intérêt des aspects énonciatifs dans la variabilité du genre. Elle pourrait être liée par ailleurs à des pratiques spécifiques de la recherche, en termes d'identités disciplinaires, selon par exemple que l'objectif visé intègre ou non un enjeu d'opposition dans le champ de recherches où s'inscrit l'article, et avec lui l'auteur.

D'autres spécificités scripturales restent plus difficilement interprétables. L'inventaire donné de chaque variable marquée dans l'usage qu'un même auteur fait du genre en atteste, avec des styles se caractérisant par une longueur spécifique des mots, la préférence pour des termes au pluriel ou au singulier, un rapport spécifique entre mots-outils et mots signifiants, certaines constructions syntaxiques (par ex. l'attribut) plutôt que d'autres, etc. Ces traits suggèrent la proximité du style en sciences avec ce qu'il est possible d'observer dans les textes littéraires, mais leur fonction dans les textes reste peu évidente à cerner ; ils montrent que le style est une affaire de choix scripturaux, ou plutôt d'actualisation de possibilités diverses de la langue car la question reste posée de savoir en quoi ces « choix » sont ou non conscientisés.

6.4.4. Styles et lexique transdisciplinaire

Il nous reste à analyser en quoi les usages singuliers du genre manifestent des spécificités au niveau du lexique transdisciplinaire. On commencera par signaler que ce lexique transdisciplinaire est particulièrement productif pour la caractérisation des styles essentiellement en termes de caractéristiques positives : 99 termes sur les 100 entrent dans le style d'au moins un auteur, un seul n'entre que dans un style, 60 termes contribuent par exemple à la caractérisation de quatre styles et plus, et sur ces 60, 57 en caractérisent positivement le même nombre. C'est donc en termes de sur-représentation, plutôt que de sous-représentation, que les termes transdisciplinaires jouent un rôle discriminatoire.

Le cas le plus frappant est celui de A15, en raison du nombre de termes jugés spécifiques de son style, réunis dans le tableau ci-dessous ; c'est à ce niveau que son style se singularise le plus, davantage qu'au niveau morphosyntaxique qui s'avère globalement peu distinctif.

En positif :

théorique, méthodologique, examen, définir, constat, expérimental, problématique, problème, procéder, définition, catégorisation, représenter, méthodologie, réflexion, tableau, conclusion, conceptuel, décrire, étudier, démarche, conception, traiter, investigation, discussion.

En négatif :

indiquer

Figure 6.3. Liste des termes transdisciplinaires discriminant le style de A15.

Ce lexique transdisciplinaire peut poser un problème de biais, en raison de certains usages conceptuels des mêmes unités formelles¹, comme nous en avons discuté en décrivant l'usage de l'article en LET et SCL à partir de cette entrée (cf. chapitre 5).

Indépendamment de ces biais très isolés, une limite se manifeste dans la difficulté qu'il y a à les interpréter le rôle visiblement important joué par ces termes dans la diversité des usages du genre. On peut voir dans la préférence d'un terme ou d'un autre des spécificités idiolectales, pour des termes dont le statut dans la construction des savoirs est potentiellement équivalent (par ex. *problème* et *question*, *traiter* et *étudier*, etc.). Comme nous l'avons signalé dans la partie précédente à propos de la différence d'usages des termes chez les doctorants et non doctorants, ils peuvent éventuellement aussi signaler des types de textes ou des types d'études privilégiés par l'auteur. Certains termes semblent pouvoir indiquer une dimension empirique, d'autres une perspective plus théorique, qui intègre des éléments de discussion épistémologique (cf. les termes dans le style de A15 cités ci-dessus), mais la plupart des termes n'offrent en eux-mêmes que peu de prise sur de tels profils de textes.

Nous nous contenterons de quelques observations prudentes, en notant par exemple que l'association des termes dans un style peut signaler que les textes tendent à rendre explicitement compte des étapes de la démarche scientifique : élaboration des hypothèses (*présupposé*), construction des observables (*expérimentation*, *matériau*), analyse et conclusions (*démontrer*, *impliquer*) dans les textes de A9. Dans le style de A7, les termes marqués, dédiés à la discussion théorique (*question*, *notion*, *idée*, *discussion*), semblent aller dans le sens des autres traits mis en évidence, et pourraient se rallier à l'hypothèse de textes prospectifs (futur, conditionnel) et argumentatifs (négation, connecteurs).

6.4.5. Bilan sur les traits stylistiques

En guise de bilan sur la caractérisation différentielle des styles et la validation des traits discriminant les textes de chaque auteur par rapport au corpus « genre », on retiendra d'abord le nombre très important de traits textuels susceptibles de varier avec l'auteur. A titre illustratif, seules 17 variables sur les 237 comprises dans les jeux Cordial et TnT ne dépendent pas de cette variable². Ces 17 variables comprennent les pronoms possessifs, les pronoms disjoints de seconde personne, les pronoms personnels de 2^{ème} personne du singulier, ou encore le subjonctif imparfait et les symboles linguistiques.

Cette analyse complémentaire corrobore l'extrême variabilité des usages du genre et la pertinence de la notion de styles personnels. Les variables qui ne sont pas soumises à un effet du style suggèrent au contraire une stabilité du genre ; on remarque que ce sont principalement des variables marginales dans l'article qui sont concernées : ce sont sur les traits qui caractérisent le genre par leur absence ou leur rareté que les textes présentent la plus grande homogénéité. Autrement dit, la stabilité du genre dans l'usage que chaque auteur en fait, ici restreinte à ce qui est commun aux textes, pourrait tenir aux éléments qui y sont prohibés, plutôt qu'à ceux qui y sont en usage.

Cette stabilité provient aussi de configurations variées, mais se distribuant de manière régulière dans le genre, comme c'est le cas avec les marques personnelles dont nous avons discuté, et sur lesquelles nous reviendrons au chapitre 7 pour mettre en évidence des profils de

¹ Ainsi en est-il du terme de « commentaire » qui caractérise positivement le style de A1. Or, l'auteur traite dans nombre de ses textes de commentaires méta-énonciatifs, et c'est donc la spécificité de l'objet qui se trouve mise en évidence par le caractère marqué de ce trait dans ses textes.

² I.e. elles se comportent de la même manière quel que soit l'auteur du texte analysé : il peut arriver qu'elles entrent dans la caractérisation de l'un ou l'autre des styles, mais ce caractère discriminant reste trop partiel et en cela hasardeux pour conclure à un effet de l'auteur du texte sur leur usage.

textes en fonction des configurations de marques. La variable « symboles linguistiques » n'est pas non plus une variable marginale : son comportement stable au sein du corpus est à comprendre par l'unité disciplinaire de ce dernier, qui en fait une caractéristique attendue du genre.

Si l'on s'intéresse aux variables discriminantes, qui entrent dans la caractérisation des styles, la productivité des variables retenues est vraie non seulement au niveau morpho-syntaxique avec les jeux TnT et Cordial, mais aussi avec le lexique et le lexique transdisciplinaire. Les traits les plus soumis à la singularité des usages du genre demandent d'abord à mettre de côté le biais résultant des caractéristiques marquées du matériau de certains textes (ceux qui prennent des exemples relevant de l'oral). Ils montrent cependant le poids du sujet de l'article, c'est-à-dire, d'une part, son objet, avec le caractère discriminant de certaines thématiques marginales (politique, sémiotique) et du traitement linguistique spécifique de l'objet dans les textes historiques. D'autre part, on voit que le sujet de l'article tient doublement à son corpus et à son objet (par ex., *vache folle*).

Plus largement, les caractéristiques stylistiques signalent des types d'approches propres à un auteur : plus historiques (cf. A3, A4 ou A9), plus descriptives et plus exemplifiées (cf. A2, A6, A8, A11, A13), marquées par la discussion théorique et le positionnement doctrinal (A15), ou encore par une argumentation forte, voire tendanciellement polémique (A4, A7, A10), bien que ces quelques pôles ne soient sans doute pas mutuellement exclusifs, et qu'il reste délicat d'associer systématiquement traits textuels et types d'approches.

Un rapport spécifique au lecteur et à l'objet se manifeste dans des pratiques idiomatiques du genre, qui peuvent cependant s'expliquer par des effets de formation et par des identités disciplinaires plurielles. L'existence dans l'écriture de l'article de styles d'auteur au sens traditionnel de spécificités idiolectales est principalement liée aux aspects syntaxiques et énonciatifs, mais montre plus largement que la singularité des usages du genre et la singularité d'un auteur en sciences tiennent indissociablement à des spécificités scripturales et à des pratiques spécifiques de la recherche au sein du champ considéré.

Pour avancer dans l'analyse des styles, et dans celle de la variabilité du genre que l'usage singulier de chaque auteur permet de mieux cerner, il nous faut à présent nous pencher sur le fait que les styles ne sont sans doute pas tous singuliers au même degré.

6.5. À quel degré un style se distingue-t-il ? L'homogénéité des textes d'un même auteur et leur singularité au sein du corpus.

La mise en évidence des spécificités stylistiques des articles d'un auteur permet de voir sur quelles variables les styles s'appuient, et quels sont les lieux-pivots privilégiés de la variation stylistique. Le fait qu'un auteur se distingue par l'usage abondant qu'il fait d'un concept, ou par l'usage très restreint de tel ou tel type de pronoms dans son texte est à comprendre de manière structurale, par rapport à l'opposition des auteurs entre eux. Le style est donc abordé en termes de traits qui distinguent les textes d'un même auteur en raison de leur sur- ou de leur sous-représentation par rapport à l'ensemble du corpus.

Or, la caractérisation différentielle des styles au sein du corpus « auteurs » et par comparaison avec le corpus « genre » suggère qu'un nombre variable de traits définit la spécificité des textes d'un auteur par rapport à un autre : certains styles se caractérisent par un nombre très important de traits quand les textes d'autres auteurs, plus restrictivement définis, semblent davantage se fondre dans un modèle commun du genre.

En recourant à des méthodes de classification, nous avons donc cherché à approfondir ce constat de l'existence de degrés de distinctivité des styles. Il s'agit d'évaluer dans quelle

mesure un style est homogène et dans quelle mesure un style est singulier, selon que ses textes présentent des caractéristiques plus ou moins unifiées, et plus ou moins propres à l'auteur ou partagées par d'autres.

La méthodologie adoptée doit permettre de classer les textes en fonction de critères linguistiques de différenciation et d'analyser alors le positionnement des textes d'un même auteur par rapport aux classes constituées par les textes du corpus en fonction de ces critères. Deux démarches complémentaires ont été utilisées : l'une, menée sur des jeux de variables différents (variables Cordial, TnT, etc.), mobilise une classification ascendante hiérarchique couplée à une analyse en composantes principales ; elle produit des classes de textes en fonction de deux premiers facteurs, déterminés comme tels en tant qu'ils épuisent les plus importants principes de structuration de la diversité des textes. La seconde n'est réalisée que sur un jeu de variables (TnT) et consiste en une méthode de validation des classes produites : les visualisations provenant de la première analyse demandent à ce que se trouve précisée la confiance que l'on peut accorder à la position de chaque point (Lebart, 2004).

C'est précisément ce que permet la technique de « Bootstrap », technique assez récente rendue possible par le développement de la puissance de calcul des ordinateurs, et dont on doit la formalisation à l'ouvrage de référence de B. Efron et R.J. Tibishirani (1993). Les auteurs expliquent que le terme provient de l'expression anglaise « to pull oneself up by one's Bootstrap », littéralement « se soulever en tirant sur les languettes de ses bottes ». Si la traduction reste difficile, la notion de « Bootstrap » renvoie à l'idée de « se tirer d'un mauvais pas » (Buvat, 2000), par soi-même ou, de manière figurée, par « la force du poignet » (Palm, 2002) ; elle constitue une alternative aux méthodes classiques d'inférence statistique, en élargissant leurs conditions d'application et/ou en neutralisant les éventuels biais qu'elles peuvent induire (c'est dans cette seconde optique que nous y recourons). Elle vise à définir l'intervalle de confiance par la méthode de l'erreur standard (« standard Bootstrap confidence interval ») et va ainsi permettre de tracer des zones de confiance (ellipses ou enveloppes convexes de réplique) (Efron et al., 1993, Lebart, 2004).

Le principe général de cette technique consiste à rééchantillonner l'échantillon initial, pour permettre une inférence statistique plus fine, basée sur les résultats des échantillons ainsi obtenus. On réalise un certain nombre d'échantillons – qualifiés d'échantillon Bootstrap – obtenus par tirage aléatoire d'observations de l'échantillon initial. Sur chacun des échantillons Bootstrap, on estime les différents paramètres du modèle pour voir si la distribution de la suite de paramètres ainsi obtenus converge vers la réelle distribution du paramètre. La visualisation produite par la méthode du Bootstrap ne représente donc pas les textes eux-mêmes mais des ellipses de confiance relatives à leur position en regard des configurations de variables attendues sur les axes factoriels ; elle doit nous permettre ainsi de corroborer la significativité de la variable stylistique et le positionnement des textes de chaque auteur en termes de distinctivité.

Nous présenterons dans un premier temps les résultats obtenus à partir des classes de textes quant aux degrés de distinctivité des styles, plus précisément à l'homogénéité et la singularité des textes d'un même auteur par rapport aux autres textes entrant dans le corpus « auteurs », puis nous montreront en quoi ils sont validés et spécifiés en regard des ellipses de confiance produites par la méthode de Bootstrap telle qu'elle est implémentée dans DTM.

6.5.1. De l'homogénéité et de la singularité des styles d'auteur

La première méthode établie rappelons-le sur la base d'une CAH associée à une ACP, vise à tester le degré de spécificité des textes d'un auteur, tout en approfondissant cette notion de

spécificité stylistique. La classification produite fournit un positionnement des textes qui permet d'apprécier l'homogénéité des styles (les textes de l'auteur présentent des caractéristiques unifiées et sont rassemblés dans une même classe) et leur singularité (les textes de l'auteur sont isolés dans une classe qui tend à leur être exclusivement dédiée).

Nous avons procédé à une triple classification, à partir de trois systèmes de descripteurs : les textes sont décrits au moyen des variables TnT d'une part, au moyen des variables TnT et du lexique transdisciplinaire d'autre part, et enfin au moyen des variables Cordial et du lexique transdisciplinaire. Faire varier les données entrant dans la classification a pour but de spécifier les résultats obtenus en neutralisant les éventuels effets liés au type d'étiquetage morpho-syntaxique et en précisant le cas échéant l'homogénéité et la singularité des styles en fonction des descripteurs retenus dans l'élaboration de ces classes.

Un style a été considéré homogène si la (quasi) totalité des textes d'un même auteur étaient regroupés dans une même partition, et singulier si cette même classe s'avérait isolée. En assignant le score de 1 en cas d'homogénéité, et de 2 en cas d'homogénéité et de singularité, on obtient le tableau suivant :

Scores :
 Style homogène et singulier : 2
 Style homogène (et non singulier) : 1
 Style non marqué : 0

	TnT	TnT + épisté	Cordial + épisté	Moyenne des scores	Total des scores
A1	2	2	2	2	6
A2	2	1	0	1	3
A3	0	1	0	0,33	1
A4	2	2	2	2	6
A5	2	1	0	1	3
A6	1	1	1	1	3
A7	0	0	1	0,33	1
A8	1	0	1	0,67	2
A9	2	2	1	1,67	5
A10	1	0	0	0,33	1
A11	1	0	0	0,33	1
A12	2	2	2	2	6
A13	1	1	0	0,67	2
A14	0	1	0	0,33	1
A15	0	1	1	0,67	2

Figure 6.4. Scores d'homogénéité et de distinctivité des styles en fonction des jeux de variables entrant dans la classification.

Les résultats obtenus diffèrent sensiblement lorsque l'on considère les deux types d'annotation morphosyntaxique, TnT et Cordial ; au vu de la caractérisation de chacun des styles (cf. leur descriptif en annexes 4), les différences ne semblent pas tant tenir à l'annotation elle-même qu'aux variables prises en charge dans chacun des jeux. Ainsi, certains styles sont peu, voire non marqués avec TnT tandis qu'ils le sont avec Cordial (A7, A15). A l'inverse, certains qui semblaient homogènes, voire singuliers avec l'étiquetage TnT, sont non marqués avec Cordial (A2, A5 et A13).

En revanche, pour des auteurs dont le style ressort fortement (A1, A4, A12), ou pour d'autres dont le style est à l'inverse peu ou pas marqué (A10, A11, A14), ce degré de

distinctivité stylistique reste similaire, apparaissant ainsi comme indépendant du système de description pris en compte.

Si l'on neutralise ces différences en calculant la moyenne des scores obtenus par les auteurs avec les trois systèmes de variables, on obtient la représentation suivante, qui rend compte de chaque style selon son degré de distinctivité.

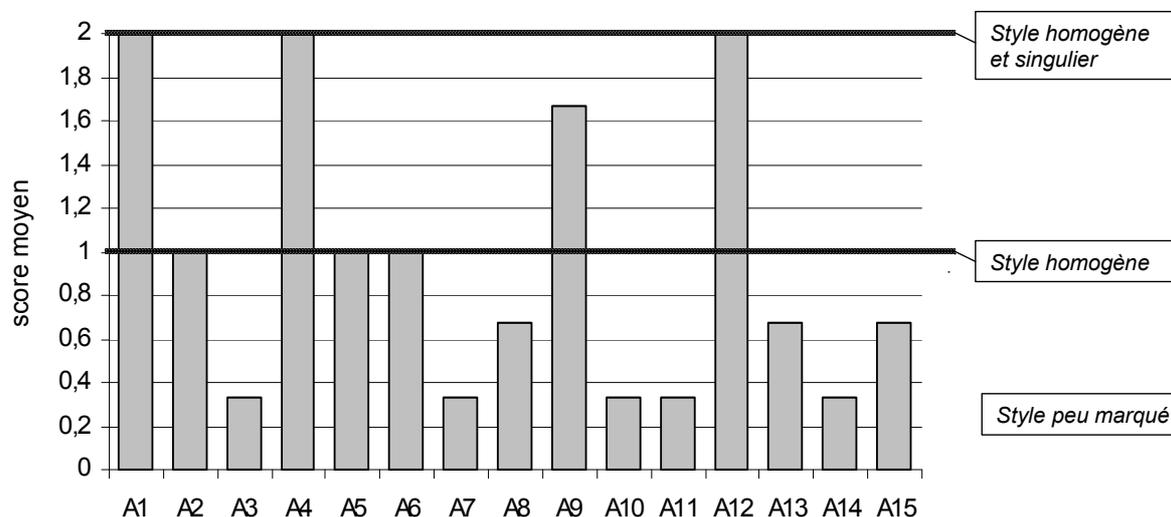


Figure 6.5. Singularité et homogénéité des styles d'auteur

Certains styles affichent un faible degré de distinctivité (A3, A7, A10, A11 et A14) ; on voit cependant que leurs textes présentent des caractéristiques suffisamment stables pour être rassemblés dans une même classe avec l'un des jeux de descripteurs, ce qui valide la notion de style même chez des auteurs dont les textes se conforment le plus aux tendances centrales du genre.

Au contraire, d'autres styles se distinguent très nettement (A1, A4 et A12) : les textes de l'auteur sont alors systématiquement regroupés dans la classification, ce qui atteste de la présence de caractéristiques stylistiques stables, et ils sont isolés des autres au sein d'une classe qui leur est quasi dédiée, ce qui dénote une singularité stylistique. Le style de A9 se distingue également très honorablement, mais de manière moindre si l'on considère la classification fondée sur l'étiquetage Cordial, qui ne prend en compte que les temps verbaux, et non les noms propres et les dates, fortement caractéristiques du style comme nous l'avons vu, en raison d'un mode historico-narratif dans ses textes, lié au domaine de spécialité (la linguistique historique).

Les textes de A6 sont également systématiquement réunis, mais au sein de classes plus importantes : le style de A6 serait ainsi identifiable en termes d'homogénéité de ses textes, mais les caractéristiques stylistiques de l'auteur seraient finalement moins saillantes, dans la mesure où elles seraient partagées par de nombreux textes. On notera que A2, A5, A8, A13 et A15 se trouvent dans un cas similaire, quoique de manière variable pour A8 et A13 en fonction des variables retenues.

La sortie de DTM qui suit illustre bien la spécificité de certains styles par rapport à d'autres, en termes de partitions stylistiques fournies à partir de l'association des jeux TnT et du lexique transdisciplinaire. On a coloré :

- en rouge les styles homogènes et singuliers (auteurs dont l'ensemble ou presque des textes est contenu dans une même partition isolée, i.e. constituée des textes de ce seul auteur),

- en bleu les styles homogènes (auteurs dont l'ensemble ou presque des textes est contenu dans une partition réunissant aussi d'autres textes, d'auteurs variés, ou de certains auteurs en particulier).

Se manifeste ainsi le caractère homogène et singulier de certains styles (A1, A4, A9, A12) et le caractère homogène de ceux d'A2, A3, A6, A13 et A15 avec ces descripteurs.

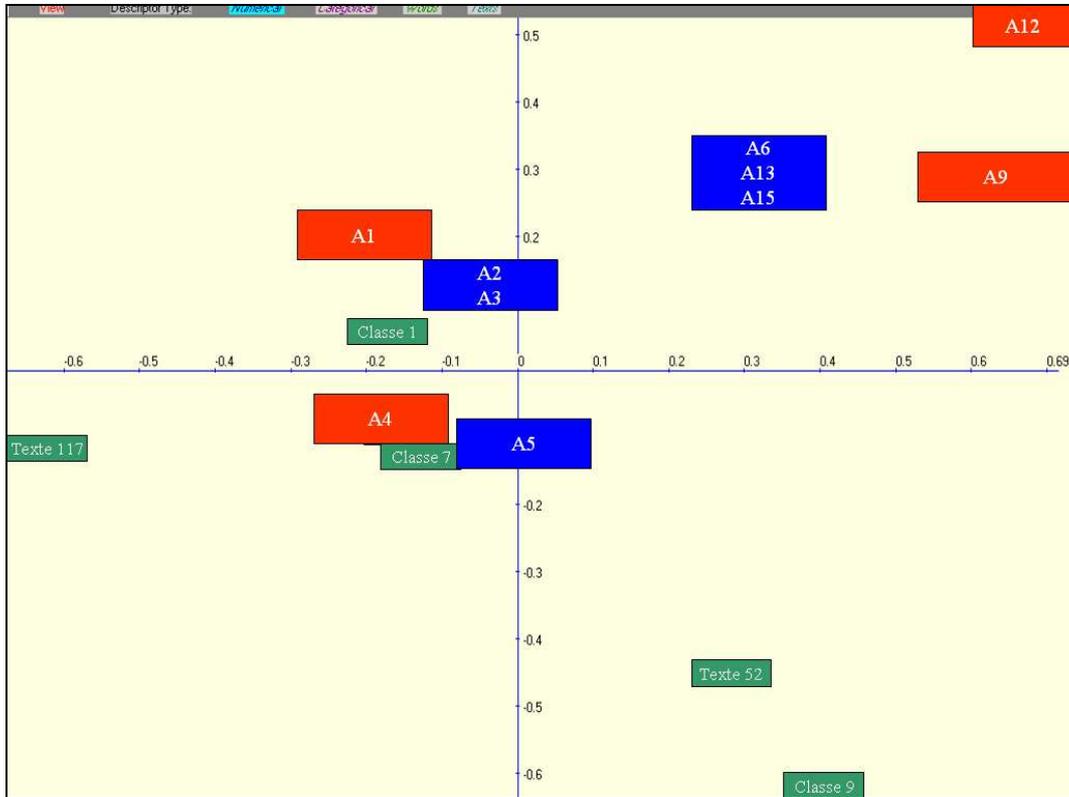


Figure 6.6. Classification des textes des différents auteurs dans DTM, à partir des variables TnT et du lexique transdisciplinaire.

Quoique tous les styles ne se singularisent pas au même degré, les textes d'un même auteur présentent des caractéristiques suffisamment stables pour être regroupés au sein d'une même classe avec au moins un des jeux de variables. Si la notion de style se trouve ainsi validée, l'usage singulier du genre que fait chaque auteur peut cependant tendre à être plus ou moins marqué par rapport aux tendances centrales du genre.

En conformité avec les résultats obtenus à partir de notre entreprise de caractérisation (cf. annexes) les descripteurs « responsables » de la variabilité du genre montrent le biais introduit par un matériau oral (cf. notamment chez A1, A2, A6). Ils confirment le caractère marqué d'un mode historico-narratif (spécifique des textes d'A9), d'une argumentation tendanciellement polémique (spécifique des textes d'A4), et le rôle discriminant de la structuration des textes (usage important de virgules, paragraphes, phrases et propositions de taille plus importantes, moins d'indices de structuration chez A12). On retrouve également dans les styles peu marqués des caractéristiques attestant du potentiel discriminant de la ponctuation et des temps verbaux notamment.

Reste à corroborer la significativité de la variable stylistique et la distinctivité des styles en validant les classes de textes ici obtenues par la méthode du Bootstrap.

6.5.2. Validation de la distinctivité des styles par la méthode Bootstrap

Afin de valider la significativité de la variable stylistique et la distinctivité des styles, nous avons appliqué *via* DTM la méthode du Bootstrap en prenant comme échantillon initial le corpus « auteurs » d'une part, d'autre part le corpus « auteurs » intégré au corpus « genre ». Au rééchantillonnage des textes et à la distribution des paramètres linguistiques produits par le Bootstrap s'associe un Tri Systématique de Signification (TSS) : il consiste à intégrer la variable indépendante « Auteurs », positionnée *a posteriori* sur les axes principaux au moyen d'une valeur-test exprimant « la signification statistique de la coordonnée de la variable sur l'axe » (Lebart, 2004 : 712).

Une zone de confiance resserrée et non recouverte tend à corroborer la distinctivité des styles : elle valide la position et la position relative des textes d'un auteur par rapport aux autres textes du corpus. On obtient ainsi pour le corpus « auteur » la représentation graphique suivante :

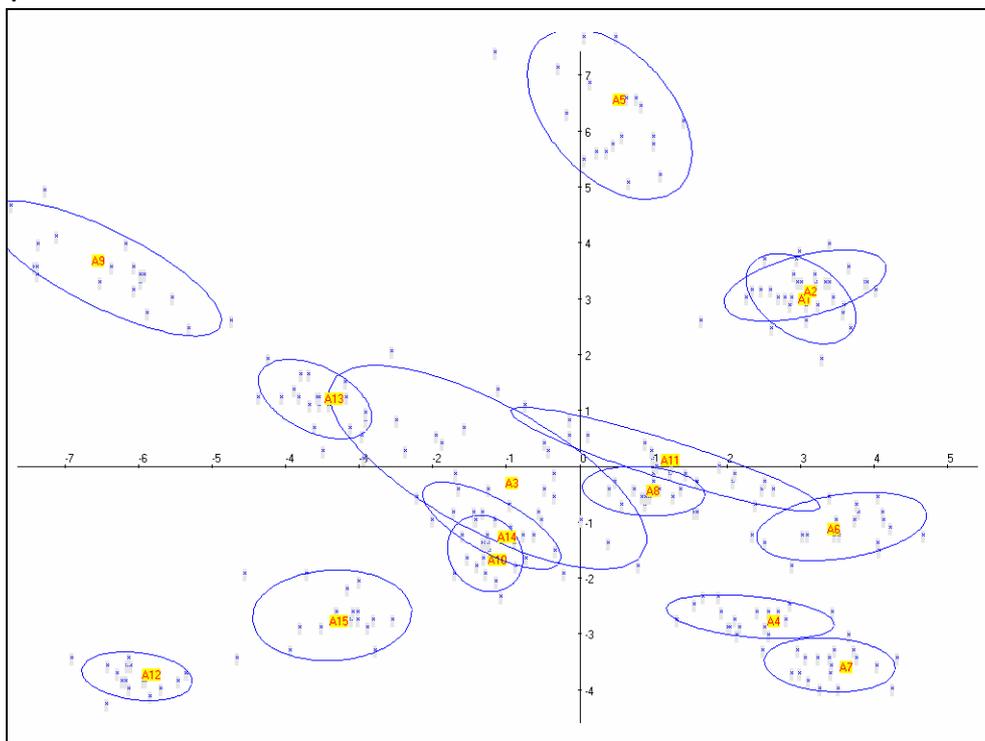


Figure 6.7. Ellipses de confiance autour des 15 auteurs observés sur le premier plan factoriel

D'emblée, on peut noter que les ellipses sont toutes relativement restreintes, ce qui atteste d'une significativité de la variable stylistique : les données observées dans le corpus permettent d'inférer que les articles d'un même auteur présentent des caractéristiques relativement homogènes d'une part, et assez spécifiques pour se distinguer de ceux des autres.

La taille et la position des ellipses manifestent cependant des différences de degrés de distinctivité déjà constatées par ailleurs. Six styles sur quinze se présentent ici comme particulièrement singuliers, dans le sens où les ellipses ne se recouvrent pas (A4, A5, A7, A9, A12 et A15). On retrouve ainsi les styles mis en évidence *supra* comme étant homogènes et singuliers (celui de A4, A9 et A12) ; les résultats diffèrent de ceux déjà obtenus en ce qui concerne A7 et dans une moindre mesure A15, dont les styles avaient été qualifiés de peu marqués. On peut comprendre cette différence en lien avec le jeu de variables utilisé (variables TnT seulement).

Il est intéressant de constater que la méthode du Bootstrap corrobore les pôles extrêmes d'un continuum relatif à la distinctivité des styles mis en évidence *supra*, en validant le caractère homogène et singulier de certains styles (A4, A19, A12), et le caractère peu marqué de certains autres (A3¹, 10, A11, A14). La méthode cependant apporte certaines précisions, quant à la proximité de certains styles (A1 et A2) ; on peut la comprendre par le fait que A1 et A2 ont en commun d'être des spécialistes du discours, qui travaillent plus précisément encore sur du matériel oral.

La méthode semble également spécifier la position et la position relative des textes d'auteurs dont les styles se situent de manière plus centrale sur ce continuum : A2, A5 et A6 seraient à ranger du côté des styles discriminants, bien que leur score ait permis de penser qu'ils ne l'étaient pas autant que d'autres. C'est vrai aussi de A15, qui présente un style plus isolé que A8 et A13, alors que tous trois obtenaient le même score précédemment. Cela étant, même si les résultats demeurent relativement convergents dans le sens où les trois forment des zones de confiance suffisamment restreintes pour attester et prédire la spécificité de leurs textes, on note des fluctuations assez difficilement interprétables dès lors que l'on se situe dans un entre-deux entre styles très marqués et styles peu marqués.

L'hypothèse que la distinctivité des styles demeure stable quelle que soit la méthode employée pour les styles les plus et les moins discriminants se trouve validée par une ultime analyse statistique, consistant à appliquer la méthode Bootstrap en combinant le corpus « auteurs » et le corpus « genre ». Il s'agit de mettre les styles à l'épreuve du genre, en confrontant les textes de chacun des auteurs à l'ensemble constitué par les textes du corpus « genre » ; pour procéder au tri systématique des significations, on intègre la variable « auteurs » en affectant aux textes du corpus « genre » la variante « autres » (i.e. autres auteurs que ceux du corpus « Auteurs ») et l'on obtient la représentation suivante :

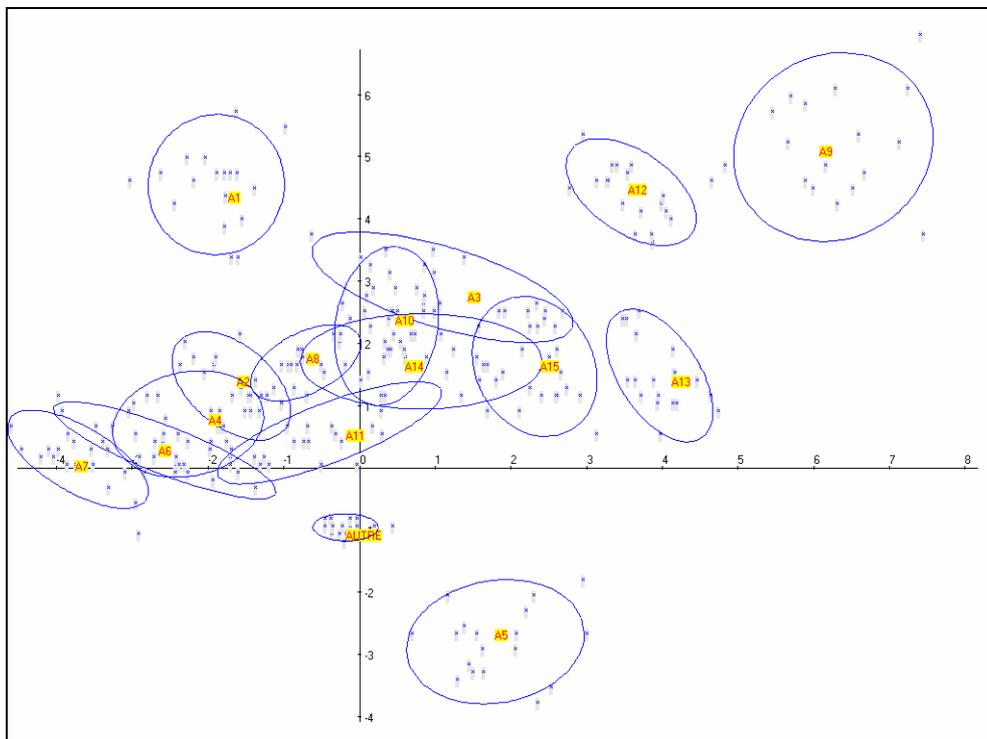


Figure 6.8. Ellipses de confiance autour des 15 auteurs observés et de la variante « autre » (tous auteurs du corpus « genre » confondus) sur le premier plan factoriel

¹ Le style de cet auteur est le moins distinct sur le plan factoriel : on note en effet 5 recouvrements stylistiques et l'ellipse est plus grande.

Si les styles d'auteur se distinguent relativement les uns des autres au sein de leur corpus, la procédure nous permet d'apprécier leur significativité une fois les textes intégrés au corpus « genre », dans lequel la dimension stylistique est neutralisée.

Les ellipses demeurent suffisamment resserrées pour attester de la présence de spécificités stylistiques relativement homogènes dans les textes d'un même auteur. Cependant, le recouvrement de la plupart d'entre elles suggère que ces caractéristiques sont souvent proches les unes des autres, la position attendue des textes d'un auteur pouvant aller jusqu'à s'inscrire dans l'ensemble caractéristique des textes d'un autre. Quatre styles sur quinze sont cependant particulièrement distincts : on retrouve les styles A5, A9 et A12 déjà isolés dans le corpus « auteurs ». En revanche, si les styles A1 et A2 étaient recouverts dans le premier corpus, on note que le style A1 est bien distinct sur le plan factoriel, tandis que A2 est en intersection avec trois autres styles, ce qui rejoint les résultats fournis au moyen des scores d'homogénéité et de distinctivité. Les styles jugés peu marqués dans ce cadre, en particulier ceux de A3, A10, A11, A14 se retrouvent ici encore en position centrale, en intersection avec de nombreux autres styles.

De nouveau donc, le caractère très marqué ou à l'inverse peu marqué de certains styles est validé, et c'est au niveau des styles intermédiaires que les résultats des différentes méthodes sont le moins stables, comme le montrent les positions de A7, A8, A13 ou encore A15 dans les trois visualisations (visualisation des scores et visualisations Bootstrap). On notera enfin qu'un style comme celui d'A4, jugé discriminant au sein du corpus « auteurs », le serait moins une fois ce corpus intégré à celui du genre, ce qui suggère que la caractérisation différentielle des styles reste dépendante des textes représentés dans le corpus « auteurs ». On voit ainsi confirmée l'importance lors de la caractérisation du style d'avoir procédé à une validation avec le corpus « genre » des effets jugés significatifs au sein du corpus « auteur ».

On peut donc conclure de ces diverses études sur la distinctivité des styles que la confiance à accorder aux résultats est très forte pour les auteurs dont les styles sont très ou très peu marqués. Cela étant, il reste délicat de prédire des caractéristiques des textes d'un même auteur et de leur degré de distinctivité, pour des styles jugés intermédiaires. L'ensemble des analyses corrobore toutefois la significativité de la variable stylistique en termes d'effets sur les textes : si la caractérisation différentielle des styles au sein du corpus « auteurs » est productive, la notion de style reste valable une fois les textes de chaque auteur confrontés aux normes du genre.

Par ailleurs, il nous reste à commenter la spécificité de l'ellipse « autre » dans le dernier graphique : représentant les positions attendues des textes du corpus « genre », le caractère extrêmement resserré de cette ellipse montre qu'il existe une unité forte du genre, attestée dans le corpus mais également prédictible pour d'autres textes, en dépit de l'hétérogénéité des pratiques que l'on vient d'évoquer. C'est à cette permanence du genre que nous nous intéresserons pour clore cette étude. Si l'analyse des styles valide la pertinence de cette notion y compris dans un genre *a priori* fortement contraint, elle ne remet pas cependant en question la stabilité du genre, et permet même de mieux cerner ce qui fait cette stabilité.

6.6. Retour au genre et à la question de sa stabilité

La stabilité du genre se manifeste d'abord dans ce qui est commun aux textes : nous l'avons montré *supra* (cf. bilan de la caractérisation des styles), seuls quelques rares traits sont communs aux textes quel qu'en soit l'auteur : ces traits immuables représentent des éléments marginaux dans le genre, et permettent de penser que ce qui est commun aux textes du même

genre tient à des éléments qui y sont tendanciellement prohibés plutôt qu'aux caractéristiques positives des textes.

Il reste à voir si cette stabilité du genre se manifeste jusque dans sa variabilité. La diversité des textes se structure-t-elle selon certaines tendances stables ? Quels traits textuels spécifiques sont responsables de tendances distinctes, qui caractérisent le genre dans sa variabilité ? Par ailleurs, cette diversité interne du genre se structure-t-elle de manière similaire ou différente avec ou sans neutralisation de la variable stylistique ? Autrement dit, reste-t-elle stable dans le corpus « genre », où les auteurs représentés sont extrêmement diversifiés, et dans le corpus « auteurs » ? Nous nous intéresserons dans cette optique aux axes de contraste internes du genre, mis en évidence à partir d'une analyse en composantes principales (ACP) entreprise au sein du corpus « genre », pour tester dans un second temps leur permanence au sein du corpus « auteur ».

Cette méthode factorielle permet de voir par la comparaison des textes comment se positionnent les uns par rapport aux autres leurs traits caractéristiques. On se restreint aux caractéristiques morpho-syntaxiques, telles qu'elles sont fournies par le jeu de variables TnT. La représentation graphique produite par l'ACP au moyen du logiciel DTM s'appuie sur les deux premiers axes factoriels, ainsi identifiés comme étant les plus significatifs pour rendre compte de la diversité interne de chacun des deux corpus¹. On a surligné sur chacune des représentations graphiques les variables qui nous paraissent intéressantes à commenter, en particulier parce que leur positionnement montre qu'elles tendent à fonctionner ensemble, comme indices de certaines configurations textuelles structurant la diversité interne du genre.

A partir du corpus « genre », on obtient ainsi la représentation suivante :

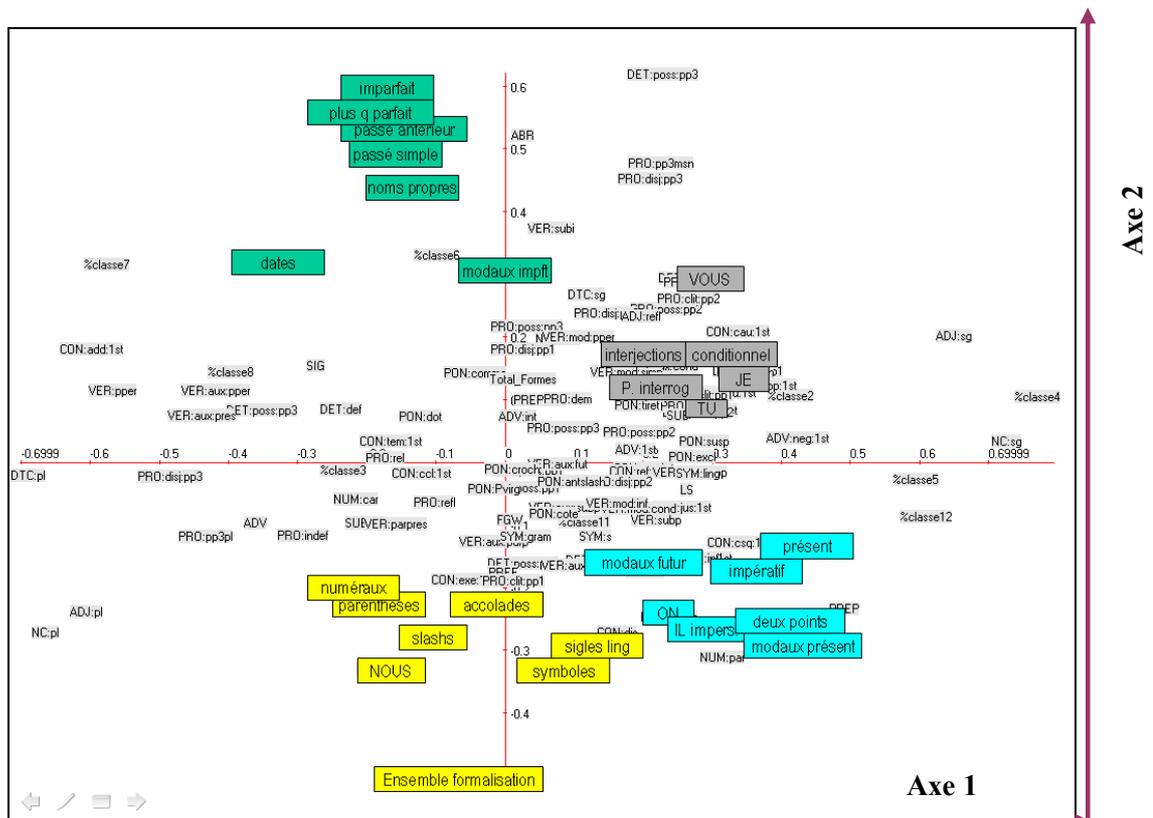


Figure 6.9. Positionnement des variables sur les deux premiers axes factoriels – corpus « genre ».

¹ Les valeurs propres des facteurs rendent compte de paliers décroissants, et les deux premiers facteurs extraient 12,41% d'inertie dans le premier corpus, 16,38% dans le second.

Sur les axes 1 et 2 s'opposent des groupes de caractéristiques qui reproduisent en partie les plans d'énonciation du discours et de l'histoire mis en évidence par E. Benveniste (1966), tout en signalant les limites. On remarque en effet un mode à visée atemporelle typique des textes scientifiques, et qui reproduit leurs caractéristiques attendues en se définissant par l'usage du temps le plus représenté, le présent, intercorrélé aussi au futur et à l'impératif, au *il* impersonnel et au *on* (en bleu).

Au pôle opposé se manifeste un mode narratif (Weinrich, 1973, Adam, 1999) dont nous avons discuté dans notre partie précédente : celui-ci ne se satisfait pas complètement de la définition du plan d'énonciation historique dans la bipartition benvenistienne, et qui se caractérise par l'emploi intercorrélé du passé simple et de l'imparfait, et de leurs homologues composés et modaux (en vert). Ces temps verbaux sont associés aux dates et aux noms propres, qui confirment la présence d'un mode narratif dans des textes comportant une dimension historique dans leur objet.

Le mode historico-narratif s'oppose sur le premier axe à un ensemble de descripteurs caractéristiques de l'oral (en gris : marques de première et de seconde personne, points d'interrogation et d'exclamation, interjections, etc.), qui rend compte du biais introduit par des articles très exemplifiés qui travaillent sur des corpus oraux, aux caractéristiques linguistiques marquées par rapport à celles de l'article. Sur le deuxième axe, le mode narratif s'oppose aux indices de structuration du texte et aux marqueurs de formalisation (en jaune : numéraux, symboles, formalisations linguistiques, slashes, parenthèses et accolades).

On observe ainsi quatre tendances, de nature toutefois très distincte, qui régulent la morpho-syntaxe du genre de l'article et reflètent un contraste entre les articles comportant un traitement narratif de leur objet, des textes que l'on pourra dire en première approche plus exemplifiés, dont les traits ressortent en raison du caractère marqué de ce qui est pris comme exemple (l'oral, en gris), ou tiennent à la formalisation nécessaire à l'analyse d'exemples (en jaune). La formalisation va de pair avec des textes plus structurés, et reste proche d'une quatrième tendance marquée par les caractéristiques attendues du discours scientifique (présent, *il* impersonnel, etc.).

La spécificité du mode narratif (en vert), très nette, se confirme lorsqu'on examine le positionnement des textes (et non plus des variables qui les caractérisent) sur les deux premiers facteurs : on retrouve en effet l'ensemble des textes appartenant aux numéros thématiques et aux revues plus historiques en positif sur l'axe 2. On pourrait donc voir dans ces textes historiques une forme de « sous-genre » de l'article de SCL, caractérisé à la fois par un objet historique, et par son traitement narratif. Cela étant, nous avons montré aussi dans notre partie précédente que ce mode narratif peut aussi bien caractériser des textes dans leur ensemble que certaines de leurs parties uniquement, quand un article comporte un aperçu historique sans que ce soit là son seul objet.

Par ailleurs, il est intéressant de constater que ce mode narratif, très présent en LET comme nous l'avons dit dans la partie précédente, s'oppose au sein des SCL à des articles plus structurés et plus formalisés (en jaune). Nous retrouvons en effet dans cette corrélation entre indices de structuration et formalismes ce que nous avons dit des modèles de textes dans notre première partie d'analyse, où ces deux traits se trouvaient caractéristiques d'un pôle représenté par certains articles de SCL, par contraste avec ceux de LET.

La même analyse factorielle des traits de différenciation des textes réalisée sur le corpus « auteur » fournit la représentation suivante ; nous y avons repositionné les groupements de descripteurs observés précédemment :

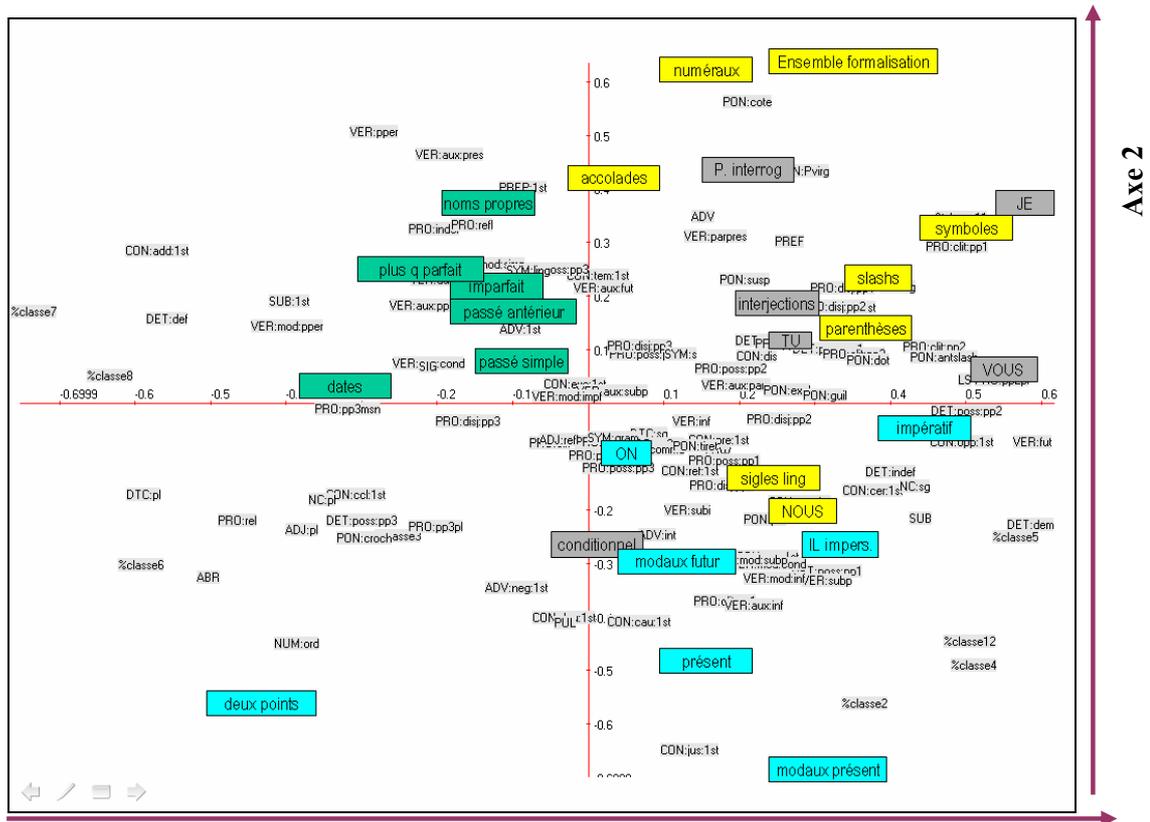


Figure 6.10. Positionnement des variables sur les deux premiers axes factoriels – corpus « auteurs ».

Le premier constat est que l'on relève des différences d'organisation du corpus « auteurs » par rapport au corpus « genre ». Cela étant se manifeste la présence persistante du groupement mode de narration historique (en vert) mis à jour dans la section précédente, qui s'oppose sur l'axe 2 à l'ensemble des caractéristiques les plus représentées du discours scientifique – nettement moins intercorrélées cependant que dans le graphique précédent (en bleu).

Les différences les plus nettes se situent au niveau des deux tendances à la formalisation et la structuration des textes d'une part (en jaune) et du caractère marqué de ce qui est pris comme exemple dans certains textes (l'oral, en gris). On obtient ainsi un pôle spécifique qui rassemble les deux types de caractéristiques, et correspondrait à des textes largement descriptifs, marqués par la place faite à leur matériau et à l'analyse de ce dernier.

L'association de la formalisation et des indices de structuration (numéraux), dont nous avons montré qu'elle renvoyait à un pôle de textes décrit au chapitre 4, se présente de nouveau comme une tendance marquée ; elle permet de corroborer le continuum des modèles de textes, car elle se distingue d'un profil central caractérisé par l'usage des traits attendus du discours scientifique et par des symboles linguistiques, également attendus dans le genre de l'article en SCL.

Notons par ailleurs que le pronom *nous* n'est plus associé au pôle formalisation mais au pronom *on* et aux *il* impersonnels, tandis que le conditionnel n'est plus corrélé aux descripteurs de l'oral, et rejoint l'ensemble des éléments traditionnellement considérés comme caractéristiques des textes scientifiques (présent, modaux au présent, deux-points, etc.).

Finalement, nous pouvons donc conclure, avec le corpus « auteur », à trois tendances structurant la diversité des textes. La première est constituée par le mode de narration historique, très nettement identifié par une corrélation stable de traits (en vert). Celui-ci s'oppose aux caractéristiques attendues de l'article en SCL : avec le présent, le conditionnel le futur et les symboles linguistiques en particulier, on retrouve en effet ce que l'analyse menée dans les précédentes parties nous a permis de poser comme des traits centraux et relativement figés du genre. L'analyse permet ici de préciser que ce modèle central du genre se caractérise par une pratique descriptive, qui à l'inverse des textes plus historiques doit faire place à l'exemplification. Enfin, les formalisations et les indices de structuration (numéraux) se présentent comme tendanciellement marqués, signalant donc que les usages peuvent se spécifier à ce niveau, comme nous l'avait suggéré l'analyse de modèles de textes dans notre première partie.

Malgré les changements observés, le corpus « Auteurs », pourtant plus restreint en termes de taille, et moins représentatif du domaine et du genre¹, porte bien les traces d'une organisation morphosyntaxique que l'on peut tenir pour générique. Les tendances structurant la diversité interne des deux corpus ne sont pas identiques, mais attestent de la permanence d'un noyau dur générique au sein du corpus « auteurs ».

La double description des axes de contraste interne, au sein du corpus « genre » et au sein du corpus « auteur » montre d'abord que le genre est loin de se limiter aux caractéristiques attendues du discours scientifique, puisque celles-ci vont jusqu'à constituer un pôle spécifique qui signale leur pertinence tout autant que la nécessité de ne pas restreindre la diversité des textes à ces seuls traits canoniques.

Le mode narratif est nettement marqué, et ce de manière très stable, si bien que les textes comportant une dimension historique peuvent apparaître comme un « sous-genre » spécifique de l'article de SCL ; celui-ci dépend vraisemblablement beaucoup des revues et des numéros thématiques représentés dans le corpus², et ne doit pas masquer la possibilité pour un texte de recourir à ce mode sans pouvoir être mis au rang des textes exclusivement dédiés à un objet historique, ni *a fortiori* au rang des textes de linguistique historique.

Quant aux autres tendances déterminant la diversité des textes, la confrontation des deux analyses montre qu'elles sont moins clairement délimitées et moins stables. L'une résulte du caractère marqué du matériau oral de certains articles, l'autre d'indices de structuration et de formalisation contribuant aux modèles de texte traités dans la première partie. Les deux tendances prises conjointement peuvent indiquer le caractère descriptif de certains textes, mais il faut souligner que cette spécificité se manifeste en tant que telle parce qu'elle s'oppose au mode narratif, et qu'elle semble étroitement liée à des caractéristiques attendues du genre de l'article en SCL.

Au final, cette ultime étude permet-elle d'un point de vue méthodologique de valider la pertinence générique de chacun des deux corpus, et la permanence relative du genre en dépit de sa variation stylistique. Les différences de résultats suggèrent cependant l'impact du corpus, au niveau du nombre de textes comme de la diversité des thématiques représentées ; elles signalent aussi l'intérêt qu'il y aurait à comparer le genre de l'article de revue avec d'autres genres représentés, quoique de manière minoritaire dans le corpus « auteurs », à savoir l'article publié en tant que contribution à un ouvrage collectif, et l'acte de colloque.

¹ Un nombre nettement plus restreint de domaines de la linguistique y est en effet représenté.

² Ainsi, la revue *Histoire Epistémologie Langages*, bien représentée dans le corpus de C. Poudat, se caractérise-t-elle par ce mode narratif et montre un positionnement spécifique dans la classification des textes, avec d'autres numéros et d'autres articles.

6.7. Conclusion sur la variation stylistique du genre de l'article

L'analyse de la variation stylistique d'un genre *a priori* fortement contraint permet de mieux cerner la diversité des textes, entre la régulation des usages au niveau de la communauté discursive et la spécificité des productions d'un même auteur.

Si tous les niveaux de la textualité varient avec le genre (Rastier, 2001), l'étude montre que dans la limite des descripteurs considérés, c'est vrai en grande partie du style aussi : quasiment tous sont soumis à un effet significatif de la variable « auteur ». On voit se manifester là l'extrême hétérogénéité des textes, y compris au sein d'un même genre. Cela étant, l'étude montre la pertinence des deux notions de genre et de style comme principes de structuration de cette diversité des textes : des traits stylistiques émergent bien de l'écriture de l'article, et restent relativement constants que l'on compare les textes d'un même auteur à ceux du corpus « auteurs » ou à ceux du corpus « Genre ».

Les traits qui ne sont pas soumis à la variation stylistique indiquent que le genre doit sa stabilité principalement aux éléments qui y sont marginaux voire prohibés, ou à des possibles diversement exploités, mais se distribuant de manière relativement régulière. La stabilité du genre implique une forte prédictibilité quant aux caractéristiques des textes (cf. l'ellipse « autre » sur la figure 6.8). Elle se vérifie aussi au niveau de sa diversité interne : le genre se structure en axes de contraste relativement permanents dans l'un et l'autre des corpus, c'est-à-dire quels que soient le nombre d'auteurs et le nombre de textes par auteur représentés dans le corpus.

Si l'étude valide l'hypothèse qu'il existe des usages du genre propres à un auteur, elle montre aussi que la variation stylistique de l'article de SCL dépend largement des auteurs considérés : tandis que les textes de certains se distinguent par des propriétés particulièrement homogènes et singulières, d'autres présentent des caractéristiques peu marquées ; on peut l'expliquer par la variabilité de leurs pratiques, et par la proximité de leurs textes avec les tendances centrales du genre. Aux deux pôles constitués par les styles peu marqués et les styles homogènes et singuliers, les caractéristiques des textes d'un même auteur sont relativement prédictibles, mais une zone intermédiaire se manifeste, qui empêche de dire de tous les auteurs si leurs textes sont plutôt du côté d'un modèle central du genre ou d'un usage singularisé.

La validation des deux notions demanderait cependant à être complétée d'une analyse de leurs relations réalisée au moyen d'une entreprise de classification menée sur un corpus plus vaste, intégrant de multiples genres et permettant d'analyser des styles dans l'usage de ces genres. Le genre de l'article et les styles d'auteur émergeraient-ils d'un tel corpus ? Comme on s'appuie sur une définition des styles relative à celle des genres, retrouve-t-on, dans le cas de l'article de recherche, les résultats mis en évidence pour les textes littéraires (Muller et Brunet, 1988), qui indiquent que les genres l'emportent sur les styles en matière de classification ?

L'étude réalisée neutralise cette question en analysant les styles au sein du genre, comme facteur de leur variabilité. Elle permet ainsi de montrer que la diversité du genre, abordée à travers la variation stylistique, ou en neutralisant cette variable, tient pour beaucoup au sujet des articles (leur objet et leur matériau), à des types d'approches, elles-mêmes liées aux objets traités et à la manière de les traiter (par exemple, le mode narratif, la dimension argumentative, l'analyse d'exemples ou les textes plus théoriques). Des modèles pluriels de textes se manifestent, qui montrent la possible existence de sous-genres, tout autant que l'irréductible diversité des pratiques scripturales et des pratiques de recherche au sein du champ considéré.

Au terme de ce parcours, la question qui se pose est de savoir ce qu'est le style en sciences¹. Les descripteurs retenus ont par hypothèse le statut d'éléments caractéristiques du discours scientifique et du genre de l'article ; s'ils s'avèrent effectivement discriminants pour rendre compte de l'existence de styles dans l'écriture de l'article, les styles qui se singulariseraient par des traits non pris en compte ici seront tenus pour conformes aux caractéristiques du genre. Inversement, notre étude ne permet pas de dire que les auteurs que notre analyse a distingués ont un style qu'on tiendra pour particulièrement original ou interpellant à la lecture.

On peut citer à ce titre le cas de A7. Il s'agit d'un linguiste connu, qui a un style bien à lui, assez aisément reconnaissable dans la communauté des linguistes : tout en respectant parfaitement pourrait-on dire les normes du genre, l'auteur va jusqu'à s'en jouer en les explicitant, en les commentant, ou même en usant d'images et d'expressions idiomatiques inattendues et complètement marginales. Que l'on en juge par la conclusion de l'un de ses articles, qui se limite aux quelques mots reproduits ci-dessous :

(21) Conclusion

C'est fini. Du moins, pour le moment. [A7-060]

Or, la présente étude ne fait pas émerger ces spécificités, mais met en avant au contraire l'inscription de ses textes dans les régularités du genre en montrant qu'A7 n'a pas un style très distinctif. Elle permet d'avancer que les particularités du style de A7 ne se situent pas aux niveaux retenus² (descripteurs morphosyntaxiques et lexique transdisciplinaire de ses textes), niveaux dont il intègre parfaitement les normes. Les quelques caractéristiques qui le distinguent sont loin de rendre compte de ce qui frappe à la lecture des textes de l'auteur ; et pour cause, humour et dérision ne se manifestent pas aux niveaux pris en compte, et demanderaient donc à faire l'objet d'un balisage spécifique, supposant lui-même de s'interroger sur ce qui en termes de marques est porteur de ce sentiment de décalage et de jeu avec les normes du genre.

¹ La même question est posée G. Molinié et P. Cahné (1994), mais à propos du discours littéraire.

² On voit cependant ici que le recours à une phrase nominale en conclusion est un trait morpho-syntaxique marqué, qui est cependant en lui-même trop isolé pour ressortir comme caractéristique dans l'analyse menée.

Chapitre 7.

La subjectivité et la polyphonie de l'article. Pour une approche énonciative de la rhétorique scientifique.

Les caractéristiques supposées du discours scientifique en font le lieu de l'objectivité, de la neutralité et de l'impersonnalité. La présence dans les textes de marques personnelles représente alors une entrée privilégiée pour dépasser une telle analyse, jugée normative et idéalisée. C'est ainsi que nombre d'études s'emploient à démontrer que le texte scientifique comporte bien un « ethos » ou une « persona », et en prennent pour preuve ces *je* ou ces *nous* par lesquels le scripteur laisse sa trace, différemment selon les langues et les cultures.

Une fois admis que le texte scientifique comporte, en français, des *je* ou des *nous* académiques, comme nous l'avons à notre tour vérifié dans notre corpus, il faut convenir qu'un tel constat ne permet pas complètement d'échapper à l'opposition entre subjectivité et objectivité ; or, comme le souligne C. Kerbrat-Orecchioni (1980), c'est cette opposition qui est à la base de l'idée d'objectivité du texte scientifique, et qui empêche de mieux cerner les deux catégories. Les marques personnelles ne suffisent pas qui plus est à mieux cerner ce qui fait l'« ethos », la « persona » ou cette forme de subjectivité propre au texte scientifique, et dont elles corroborent l'existence.

Analysées depuis E. Benveniste (1966) en tant que marques des « véritables personnes de l'énonciation », les formes de *je* en particulier signalent dans l'énoncé une réflexivité de sa propre énonciation, réflexivité que les approches indicielles en linguistique de l'énonciation ont prise pour objet. L'idée d'une réalité pré-existante, codée en langue, ne vaut pas pour le langage, et il en va en conséquence de même pour l'« énonciation représentée » dans le sens de l'énoncé. Il ne s'agit pas de nier qu'un sujet est là qui produit un énoncé, mais d'étudier l'énonciation représentée en tant que telle, au moyen de l'ensemble des procédés linguistiques qui servent cette représentation. Une telle approche, résolument ancrée dans les textes, pourrait peut-être éclairer alors ce que l'évidence de l'énonciation effective¹ (un locuteur s'adressant à un autre et référant au monde au moyen d'un code) tend à masquer.

¹ Cette évidence, calquée sur le modèle de la conversation ordinaire, nie la spécificité de l'« ordre scriptural » (J. Peytard, 1970, repris aussi dans M. Dabène, 1987, 1996), et celle des genres.

7.1. Introduction : l'article comme type énonciatif de texte

Les typologies énonciatives des textes proposées par E. Benveniste (1966, 1974), J. Simonin-Grumbach (1975) et J.P. Bronckart (1985) permettent d'aborder la spécificité de l'énonciation représentée dans le genre de l'article. Les caractéristiques énonciatives de l'article peuvent en effet être décrites sous l'angle du « discours théorique » (Bronckart, 1985) dont il est avec d'autres genres un des représentants.

L'analyse menée par J.P. Bronckart des types énonciatifs de textes en tant que types de mondes discursifs (ou de mises en discours) repose sur l'idée de fonder une typologie sur la relation entre les unités linguistiques et ce qu'il appelle le contexte ou les paramètres externes de l'action langagière (op. cité). Il met ainsi au jour des « architypes de discours », dont le discours théorique est un cas, en les décrivant à partir de deux catégories d'opérations psychologiques constitutives de la mise en discours. La première concerne les opérations qui déterminent le rapport entre les coordonnées du monde ordinaire (monde dans lequel est produit le texte) et les coordonnées du monde que produit le texte (les coordonnées organisant le contenu thématique du texte). Le rapport peut être de l'ordre de la conjonction (mode de l'exposer : le monde du texte se veut en adéquation avec le monde ordinaire) ou de la disjonction (mode du raconter). La seconde catégorie d'opérations rejoint les travaux antérieurs sur les types énonciatifs de texte : elles concernent le rapport entre le texte et la situation physique de sa production, et déterminent si le texte entretient un rapport d'implication ou d'autonomie par rapport aux paramètres de cette situation.

L'énonciation telle qu'elle est représentée dans l'article obéirait donc à deux principes caractéristiques du discours théorique. D'abord, en termes de rapport au monde, le texte de l'article – parce qu'il relève du mode de l'exposer – pourrait être défini comme un texte visant la production de savoir sur le monde. Ainsi la rhétorique parle-t-elle de « logos » pour rendre compte du fait que le discours se fonde sur la raison ; en d'autres termes, il se fonde sur une logique du vrai, et c'est ce qui explique que l'idée qu'il existe une rhétorique de la science ait dû faire l'objet de revendications récurrentes¹. Par ailleurs, le texte scientifique a bien une situation d'énonciation comme tout texte, mais cette situation n'est pas (ou que peu) impliquée (ou représentée) dans le texte : il tend à s'autonomiser de cette situation, de sorte que l'interprétation peut se faire indépendamment d'hypothèses sur la situation où le texte a été produit.

Les caractéristiques supposées du discours scientifique, celles qu'on lui attribue spontanément, ne paraissent donc pas complètement infondées si l'on se réfère aux propositions de J.P. Bronckart. La transparence et l'objectivité sous l'angle desquelles on l'envisage peuvent être comprises en regard de ces caractéristiques « architypiques ». Elles permettent par ailleurs de fournir des pistes pour l'analyse d'un « ethos », d'une « persona », d'une forme de subjectivité à l'œuvre dans le texte, ou encore, ici, d'une figure de l'auteur en tant que principe guidant notre investigation des aspects énonciatifs de l'article.

L'analyse proposée par E. Benveniste (1966) sous l'angle de l'opposition entre « discours » et « récit » se trouve donc complétée. Comme le montre G. Philippe (2002), l'effacement énonciatif du « discours théorique » (discours historique dans la perspective d'E. Benveniste) ne se caractérise pas uniquement en terme d'exclusions et d'absence (autrement dit par l'idée de « gommer » ou « masquer » toute trace de locuteur) ; il correspond à une possibilité qu'a le discours d'autonomiser son objet en semblant le couper de la situation *hic*

¹ On le voit dans l'article de P.N. Campbell (1975) qui insiste largement sur cette idée que le texte scientifique comporte une dimension rhétorique et souligne qu'elle a du mal à s'imposer. Comme l'explique J. Gardes-Tamines (1996 :14), citant les écrits d'Aristote (1991), la rhétorique circonscrit initialement son objet au vraisemblable et en exclut le vrai : celui-ci ne se discute pas, il repose sur la démonstration et non sur les procédés rhétoriques de l'argumentation.

et nunc de l'échange. Ainsi la présence de marques personnelles ne permet pas à elle seule de contredire l'effacement énonciatif caractéristique de l'article¹. Ce dernier est à comprendre comme une tendance (la non-implication tendancielle dans le texte des paramètres physiques de sa situation de production *hic et nunc*). Elle relève selon R. Vion (2001) d'un « simulacre énonciatif », autrement dit d'un régime énonciatif spécifique, car « la subjectivité est partout » (Kerbrat-Orecchioni, op. cité), si bien que se manifeste nécessairement dans le texte la présence d'un locuteur.

La distinction entre une subjectivité déictique, qui concerne le sujet *hic et nunc* à l'origine du texte, et une subjectivité modale, en tant que repère des opérations modales² doit permettre alors d'avancer dans l'analyse d'une forme de subjectivité représentée dans le texte scientifique. Elle offre la possibilité d'envisager de manière graduelle l'effacement énonciatif, en fonction du caractère plus ou moins embrayé ou déembrayé de l'énonciation représentée (sa dimension déictique), de la présence ou non de « subjectivèmes »³, et de la prise en charge énonciative des points de vue, i.e. la question des « énonciateurs ». Cette dernière se complique de la question de la polyphonie (ou du dédoublement des points de vue)⁴ et demande à être envisagée en regard de la « cohérence polyphonique »⁵ requise par le genre de texte concerné : il s'agit d'envisager en quoi les différentes instances à l'origine des points de vue peuvent être indexées au locuteur, tout en les maintenant dans leur irréductible hétérogénéité.

Notre analyse de la subjectivité et de la polyphonie de l'article prend comme point de départ des phénomènes énonciatifs (marques personnelles, subjectivèmes, références à d'autres discours et tours polyphoniques) pour montrer comment se construit le point de vue de l'auteur dans le texte, et comment se constitue dans le texte une figure de l'auteur caractéristique du genre, irréductible au scripteur à l'origine du texte. Elle voudrait permettre en particulier de concilier l'idée que le texte scientifique obéit à un régime tel qu'il est pris en charge par un énonciateur universel (les faits parlent d'eux-mêmes), et l'idée que des marques personnelles peuvent intervenir, impliquant dans le texte celui qui en est à l'origine.

Une telle étude doit également pouvoir décrire en quoi, alors même que le texte scientifique vise prioritairement une fonction explicative, il comporte nécessairement une dimension argumentative (Amossy, 2000) ; dimension de tout discours y compris quand il n'est pas argumentatif à proprement parler, celle-ci est définie par R. Amossy en des termes proches de la notion d'argumentation chez J.B. Grize (1990), en tant que démarche visant à modifier les représentations d'autrui, par les moyens du discours.

L'idée est que les caractéristiques énonciatives de l'article (l'effacement énonciatif, les marques personnelles, les subjectivèmes, le dédoublement et la hiérarchisation des points de vue) doivent contribuer à servir cette dimension : la construction du savoir repose sur des procédés textuels tels que le dit se trouve légitimé, et se présente en cela comme plus ou moins convaincant pour le lecteur. Il ne s'agit donc pas d'envisager une quelconque intentionnalité du scripteur à recourir à des procédés rhétoriques ; notre objectif est de décrire énonciativement le genre, de voir en quoi la discipline ou le statut de l'auteur peuvent influencer sur les caractéristiques énonciatives, et de suggérer d'éventuelles pistes pour une analyse pragmatique et rhétorique fondée sur l'énonciation représentée dans les textes.

¹ Elle ne contredit que l'idée que le locuteur est masqué, et que toute trace en est gommée.

² Cf. notamment A. Rabatel (2004) ; la distinction repose sur celle de l'actualisation déictique et de l'actualisation modale comme deux composantes de l'actualisation de la langue par sa mise en fonctionnement dans l'acte d'énonciation.

³ Ou toute marque lexicale de la subjectivité (C. Kerbrat-Orecchioni, 1980).

⁴ Cf. dans le chapitre 2, les développements sur l'analyse polyphonique dans la théorie de la ScaPoLine et chez A. Rabatel.

⁵ K. Fløttum (2002 :342), et cf. la note précédente.

-
- Nous nous intéresserons successivement aux quatre aspects suivants :
- l'énonciation personnelle, avec les marques personnelles *je* et *nous*, que nous avons étudiées d'un point de vue quantitatif, et que nous observons ici qualitativement, dans la diversité de leurs emplois. L'irréductible polyphonie de ces marques montre que la « personne » qu'elles impliquent dans l'article se présente comme une instance complexe, qui se caractérise doublement par son inscription dans un collectif et par sa singularité.
 - l'effacement énonciatif et la prise en charge personnelle, que nous traitons d'abord d'un point de vue phraséologique, avec le *il* impersonnel, et ensuite en termes de profils de texte ; l'étude suggère comment, plutôt que d'être deux caractéristiques contradictoires, l'effacement énonciatif et la prise en charge personnelle s'associent pour déterminer le régime énonciatif spécifique de l'article ; la subjectivité du locuteur représenté dans le texte repose à la fois sur une énonciation personnelle, et sur une énonciation universalisante (tours impersonnels, délocuté).
 - les marques de l'autre, et notamment les renvois aux autres acteurs du champ et aux points de vue attestés dans ce champ. A l'étude des citations et du discours rapporté, nous préférons une étude qui prend comme entrée les références à des auteurs ou ouvrages, et les points de vue attestés, pour nous intéresser à la représentation de la « circulation des discours » dans l'article ; cette représentation pose notamment les questions de la traçabilité des références et de la construction de l'univers de référence.
 - les objectifs annoncés dans l'article, et le lexique axiologique utilisé pour évaluer l'intérêt ou les limites d'un point de vue. Ces deux entrées permettent, à l'aide aussi des procédés précédemment décrits, de mieux cerner comment se construit dans les textes la légitimation de la recherche présentée ; elles fournissent ainsi des pistes pour une analyse du rôle des aspects énonciatifs dans la dimension argumentative des articles.

7.2. Les emplois de « *je* » et « *nous* » ou l'irréductible polyphonie de la première personne linguistique.

Nous avons étudié dans notre analyse morpho-syntaxique de l'article des marques de personne d'un point de vue quantitatif. Il convient cependant de se pencher sur les emplois de ces marques à travers une analyse plus qualitative. La linguistique de l'énonciation a bien décrit la diversité des emplois des marques de la première personne, en montrant en particulier qu'elles peuvent se présenter dans des emplois génériques, où elles renvoient à tout être humain, à côté d'emplois attendus, où elles renvoient de manière plus transparente au locuteur qui les actualise¹.

Les emplois du *je* et du *nous* qui vont retenir ici notre attention représentent la plus évidente « trace de soi » (« traces of self », K. Fløttum, 2005) à rechercher dans les textes, et constituent en cela une première entrée intéressante pour l'analyse de la figure de l'auteur. Notre objectif est d'analyser la diversité des emplois. Cet objectif descriptif permet d'une part de s'intéresser à des « marques » dont le fonctionnement énonciatif est éminemment complexe, et montre l'irréductible polyphonie linguistique. Il permet d'autre part de voir ce que ces personnes linguistiques disent quant à la « véritable personne de l'énonciation », au sens de personne à l'origine de l'acte d'énonciation telle qu'elle est représentée dans l'énoncé. La spécificité francophone de *nous* demandera également à traiter du lecteur, tel qu'il est impliqué dans le texte, à l'instar de la représentation que donne le texte d'une figure de son auteur.

¹ Cf. notamment D. Maingueneau (1994).

Pour étudier plus en détail les emplois des marques de 1^{ère} personne, nous avons relevé :

- les pronoms personnels en fonction sujet (*je, nous*), en fonction complément (*me, nous*) et les formes disjointes (*moi, nous*),
- les déterminants possessifs (*mon, ma, mes, notre, etc.*),
- les pronoms possessifs (*le mien, etc.*).

Quand nous parlerons de formes de *je* ou de formes de *nous*, ce sera donc pour évoquer l'ensemble des variantes morphologiques de chacune des deux personnes linguistiques.

L'analyse est menée sur 300 occurrences de *je* et 300 de *nous* et vise à mettre en évidence les types d'emplois de ces premières personnes linguistiques dans l'article. Le premier constat que l'on peut faire en effet est que toutes ne concernent pas l'inscription de l'auteur et que les valeurs qu'elles peuvent prendre sont très variées ; elles affectent le statut du signe, et la notion de personne de l'énonciation.

7.2.1. L'usage autonymique des marques de personne

D'abord, comme nous l'avons dit à propos de la présence surprenante de marques de 2^{nde} personne à l'issue de l'analyse automatique, l'extraction des exemples et citations à laquelle nous avons procédé ne suffit pas à éliminer complètement le risque que de courts exemples ou de courtes citations parasitent le comptage des occurrences.

En particulier, les titres d'ouvrage cités dans les articles de LET posent ce problème (« *Daniel Sangsue, à propos du Voyage autour de ma chambre de Maistre, écrit...* », LET025). Dans le même sens, de nombreuses occurrences correspondent à des emplois autonymiques, en SCL et dans une moindre mesure en LET (« *le je du discours direct* », SCL030). L'analyse des formes personnelles rejoint ainsi ce que nous avons montré à partir des guillemets quant à l'omniprésence du matériau étudié et à son intrication dans les passages de l'article où est analysé ce matériau.

7.2.2. L'usage conceptuel du « je » et du « moi »

Hors des emplois strictement autonymiques, il est intéressant de constater que le *je* peut servir selon un procédé de substitution métonymique à désigner le locuteur en tant que catégorie conceptuelle. La personne linguistique prend un statut spécifique et associe une valeur de mention (autonyme) et d'usage : elle devient une non-personne au sens de E. Benveniste (1966), car elle relève du délocuté. Les deux extraits suivants attestent de cet usage spécifique de la personne linguistique ; dans le premier, l'auteur analyse le poème, si bien qu'on reste proche de la valeur de mention du signe :

maintenant il [le poète] parle pour mettre en scène un *je* et un *tu* ou un *vous* qui entrent en dialogue, c'est-à-dire en communication amoureuse, communication fondée sur le « sentiment » [LET044]

dans le second, la valeur d'usage l'emporte, mais se trouve spécifiée par les parenthèses :

Il n'était pas surprenant qu'un chercheur travaillant à une approche dialogique du discours s'intéresse aux textes de vulgarisation scientifique destinés au grand public, ainsi qu'à la place du 'je' (sujet parlant) dans cette configuration de médiation [SCL090]

Cet emploi spécifique de la personne linguistique pour désigner une non-personne (énonciativement parlant) se manifeste dans l'usage bien connu de *moi* comme concept philosophique. On l'observe de manière fréquente dans les articles de LET tout particulièrement :

Il semble donc que Montaigne amplifie le détachement du sujet de soi-même dans laquelle Burnyeat voit l'erreur principale du pyrrhonisme. Mais le pyrrhonien ne doit accepter aucune conception du **moi** et chez Montaigne c'est justement le **moi** qui est en question. Il parle de soi-même comme "autre moy-mesme" (II,2,805b), parce qu'il change chaque heure et chaque minute et il ne peut pas dire ce qu'est le **moi**, dont la peinture est l'objet des Essais [LET006]

Le statut spécifique de délocuté est de nouveau signalé par la présence d'un déterminant ; l'absence ici de marque typographique indiquant le statut spécifique du signe rend très net le caractère lexicalisé de cet emploi conceptuel de *moi*.

7.2.3. Le « je » comme marque personnelle en usage générique

Intéressons-nous à présent aux marques personnelles qu'il est possible d'analyser comme telles : une personne énonciative est désignée par son usage du *je*. Parmi ces emplois, certains peuvent rejoindre en partie l'emploi conceptuel que nous venons de signaler. La prise en charge de l'assertion est le fait d'un *je* générique que l'on pourrait substituer par « quiconque » : *je* fait intervenir un sujet universel, que tout locuteur empirique doit ainsi pouvoir incarner.

Cette valeur se manifeste dans des contextes divers. Il est des cas où le *je* représente un locuteur en tant que producteur d'un énoncé bien particulier ; dans d'autres, on a plutôt affaire à un *je* philosophique, qui représente un sujet parlant, qui pense ou agit. Les deux énoncés pris comme exemples ci-dessous rendent compte de ces deux pôles du sujet que met en scène le *je* dans son emploi générique :

Si par exemple j'énonce à l'adresse de Marie *Tu es vraiment la reine de la tarte*, il se pourrait qu'elle rosisse de plaisir, éventualité pratiquement exclue si **je** lui déclarais en revanche *Tu es vraiment la reine des tartes*. [SCL100]

C'est un mouvement bien étrange de son être que de préférer l'assurance à l'incertitude (...) : j'ai beau savoir que **mon** sentiment de certitude répond non à la qualité intrinsèque de la proposition à laquelle j'adhère, mais seulement à **mon** désir éperdu de tenir la vérité, en **moi** une puissance mystérieuse persévère et règne, qui n'est pas proprement **moi**, et qui **me** domine pourtant; (...) Comment se produit une telle méconnaissance, comment se fait-il que **je m'**obstine aussi continûment à refuser ce dont **je** fais tous les jours **en moi-même** l'évident constat ? [LET011]

Dans le premier extrait, le *je* peut renvoyer à tout locuteur, mais tout locuteur *hic et nunc* des deux énoncés adressés à Marie et pris pour exemple par l'auteur ; dans le second, les formes de *je* sont nombreuses et complexes, et peuvent être mises au compte d'un *je* philosophique qui rejoint la catégorie conceptuelle du « moi » : les formes de *je* tendent à renvoyer à un sujet transcendantal au sens kantien.

Ce sujet transcendantal a pour caractéristique son unité et sa permanence au-delà de la production d'énoncés particuliers. Cette unification du sujet (*moi*) qui s'opère dans la diversité de ses représentations (*je*) peut être analysée à partir de l'extrait suivant :

Un même sujet parlant peut donc faire partie de plusieurs communautés linguistiques à des moments différents où il est locuteur. **Je** ne **me** présente pas comme membre de la même communauté linguistique selon que **je** parle avec **mon** patron ou avec **mon** garagiste habituel. [SCL100]

le « *je* » qui parle avec son garagiste n'est déjà plus le *je* d'un énoncé particulier (comme celui qui énonce « *tu es la reine de la tarte* » *supra*). Il diffère du « *je* » qui parle avec son patron, car, c'est ce que dit l'auteur de l'article, le « *je* » « *ne [se] présente pas* » de la même manière. Pourtant, ces différentes représentations sont indexables à « *un même sujet parlant* ».

Outre la continuité des valeurs de *je*, on voit aussi comment les marqueurs temporels dans cet extrait comme dans le précédent contribuent à signaler la permanence du sujet par delà les « *moments différents* » où il est locuteur (« *habituel* » dans « *mon garagiste habituel* », SCL100, « *continûment* » et « *tous les jours* » dans LET011).

Les deux pôles mis en évidence ne vont pas sans rappeler la distinction dont nous avons discuté en traitant de l'analyse de la polyphonie chez O. Ducrot et dans la théorie scandinave de la polyphonie linguistique (cf. chapitre 2) : tous deux envisagent dans le locuteur (tel qu'il est mis en scène dans l'énoncé) deux images possibles, le locuteur en tant que tel et le locuteur être du monde. La théorie scandinave propose de parler de locuteur de l'énoncé et de locuteur du texte en signalant toutefois l'équivalence entre les deux.

Le premier extrait cité se prête bien à exemplifier le « locuteur en tant que tel » de l'énoncé particulier « *tu es la reine de la tarte* ». Cependant, dans le second et le troisième extraits, la diversité des formes de *je* ne se satisfait pas complètement de cette bipartition. H. Nølke (2001, cf. chapitre 2) montre que dans « *je me demande si* », on a d'une part le « *je* » à l'origine de l'acte de parole, et d'autre part le « *me* » qui renvoie à une image du locuteur qui est indépendante de l'énoncé particulier, i.e. l'être du monde dans les termes d'O. Ducrot. Ici, le « *je* » se prête moins bien semble-t-il à une analyse en termes de « locuteur du texte » qu'en termes de « être du monde » car nous avons affaire à des *je* génériques. Par ailleurs, c'est précisément sur les frontières entre les catégories de locuteur en tant que tel et en tant qu'être du monde que se construit linguistiquement la réflexion sur le « moi » chez Montaigne dans le second extrait (LET011), et sur le sujet parlant et les communautés linguistiques dans le troisième (SCL100).

Ces emplois génériques de *je* manifestent la complexité du sujet tel qu'il se constitue dans l'usage du *je*. La distinction entre le locuteur en tant que tel et le locuteur être du monde est opératoire pour rendre compte de deux pôles identifiables dans la diversité de ces emplois. Elle permet en particulier de montrer qu'en LET, c'est principalement le pôle du locuteur être du monde qui est représenté dans l'usage d'un *je* philosophique. Inversement, les articles de SCL mobilisent davantage le pôle du « locuteur en tant que tel » par la pratique de l'exemple. Cependant, dans les deux disciplines, les deux pôles sont représentés, et les formes de *je* attestent de la continuité existante entre les deux ; le co-texte permet éventuellement de saturer l'interprétation en spécifiant de quel côté on se situe, comme nous l'avons dit des marques temporelles. C'est aussi ce que montre l'emploi de « *quel qu'il soit* » ci-dessous ; il confère au *je* de l'énoncé particulier une permanence caractéristique du *moi* :

Il y a sans doute une relation entre les privilèges du Commencement et ceux du Sujet. Si **je** donne tant d'importance au moment où **je** jette sur le papier les premiers mots d'un discours **quel qu'il soit**, c'est parce que, en ce moment encore, ils **m'**appartiennent totalement. Ils sortent de **mon** silence, ils sont bien l'expression de mon "**moi**". [LET010]

Finalement, nous retiendrons donc que l'usage générique du « je » dans les textes manifeste une continuité entre le *je* à l'origine d'un énoncé particulier et le *moi* permanent par-delà la diversité des énoncés qu'il produit ; cette continuité¹ est même exploitée par les articles dans l'analyse du langage avec la pratique de l'exemple, et dans l'analyse du texte littéraire et des réflexions philosophiques qu'il livre sur le *moi* de l'être humain.

¹ On pourrait aussi parler d'unification du sujet.

7.2.4. « Je » et « nous » comme marques personnelles qui renvoient à l'auteur

De même que *je*, les formes de *nous* se présentent parfois dans des emplois génériques, où elles renvoient à un collectif du type « tout le monde », comme nous le verrons plus loin. Mais abordons à présent ces emplois en sens inverse si l'on peut dire, et centrons-nous sur ceux, plus attendus, où les marques de personnes renvoient à l'auteur.

Sur 300 occurrences de *je* et ses dérivés, et 300 occurrences de *nous* et ses dérivés, ce sont les emplois les plus fréquents. Les marques personnelles se satisfont ainsi de l'analyse de E. Benveniste (1966) qui propose d'y voir les véritables personnes de l'énonciation. En l'occurrence, il s'agit du locuteur à l'origine du texte ; il s'auto-désigne au moyen de la première personne linguistique, soit au moyen d'un *je*, soit d'un *nous* dit académique, caractéristique des écrits de recherche francophones.

A partir de l'analyse des contextes dans lesquels ces marques personnelles interviennent, il semble possible d'identifier quelques emplois récurrents ; relativement proches, ils apportent cependant des spécifications quant à la « véritable personne de l'énonciation » désignée par la personne linguistique dans la perspective de E. Benveniste, c'est-à-dire, dans la nôtre, quant au locuteur représenté dans l'énoncé par l'usage de la personne linguistique.

Le *je* et le *nous* de l'auteur peuvent d'abord être employés dans le cadre de balises méta-textuelles – dont nous avons traité dans l'analyse linguistique du genre au chapitre 5. et dans celle des styles d'auteur au chapitre 6. En association avec le futur, avec des termes métadiscursifs, ou encore avec des déictiques textuels (*ici, maintenant*)¹, le *je* et le *nous* renvoient à l'auteur en tant qu'instance qui prend en charge l'organisation des contenus du texte et l'indique à l'attention du lecteur.

On relève ainsi l'intervention de *je* ou *nous* lorsqu'un extrait de passage cité est mis en gras par l'auteur (« *nous soulignons* » LET043) ; c'est vrai aussi concernant le texte de manière plus large (« *je conclurai cet article en indiquant...* » LET006, « *je cite la lettre à M. de Franquières* » LET003, « *je vais commencer par fixer un certain nombre de concepts de base, puis je me focaliserai sur...* » SCL104, « *nous l'avons souligné dans les exemples précédents* » SCL042, etc.). Un « locuteur du texte » est ainsi mis en scène, qui commente ses propres énoncés.

De manière proche de ce premier emploi, les marques personnelles peuvent également faire intervenir l'auteur en tant que sujet à l'origine du cadre d'analyse et de sa construction conceptuelle. Il peut s'agir de signaler une proposition de dénomination, de définition ou de taxinomie : « *Montaigne s'intéresse aux comportements que nous appellerons extrêmes soit parce qu'ils présentent...* », LET001, « *exemple magnifique de ce que j'appellais à l'instant un changement de pied* », LET011, « *je parlerai de séquence dialogale* », SCL006, « *nous entendons 'stratégie' au sens de Fayol* », SCL003, « *pour le classement des exemples nous distinguons entre les SN indéfinis...* », SCL044, « *nous prenons comme critère...* », SCL044, etc.

Les marques personnelles sont utilisées dans le même sens pour définir l'objet traité dans le texte. L'énoncé peut avoir pour but de spécifier des exclusions (« *nous n'évoquons pas ici les fonctionnements taxinomiques des dénominations des quartiers* », SCL063, « *je ne vais pas reprendre ici des débats anciens en sciences humaines* », SCL007, « *je laisse cette tâche pour de futurs travaux* », SCL052), ou inversement, de circonscrire l'objet en termes

¹ La notion de déictique textuel est expliquée notamment dans D. Maingueneau (1991). Les unités formelles sont identiques aux véritables déictiques spatiaux et temporels, mais renvoient au *hic et nunc* du texte et non au *hic et nunc* de la situation où il est produit.

positifs (« *notre dernière hypothèse sera que* », LET069, « *notre but dans cette étude* », LET089, SCL096, « *je me donnerai pour tâche, plus précisément, de relever le défi de...* » LET020, « *nous avons essayé de montrer que* », LET040, « *nous souhaitons mettre l'accent sur...* » LET089, « *l'approche que je tente d'élaborer dans cette étude* », SCL029, « *c'est plutôt cette seconde voie que j'emprunterai* », SCL034, « *je me propose de regarder de près* », SCL052, « *à la différence de la lexicologie préoccupée des usages et de l'histoire dans laquelle nous nous inscrivons* », SCL063, etc.) ou encore, de manière mixte, « *nous écarterons directement les cas où (...) et nous garderons le modèle* » SCL059.

Les marques personnelles s'associent de manière privilégiée à un lexique rendant compte des procès caractéristiques de l'activité de recherche. Ainsi désigné, l'auteur se présente comme initiant des choix interprétatifs. Il est ce sujet à l'origine de points de vue, qui sont plus ou moins nuancés, et introduits comme des conclusions nécessaires, ou en tant qu'avis propre au chercheur : « *il semble que cette définition nous donne le moyen d'aborder la question qui nous occupe* », LET003, « *il nous semble en effet que ces figures d'artistes secondaires* », LIT089, « *cet enchevêtrement (...) nous autorise à proposer une nouvelle interprétation des gestes* », LET040, « *ceci nous conforte dans l'idée* », SCL059, « *ceci nous incite à conclure* », SCL060.

Ces avis sont discutés : ils sont présentés comme le fait d'une « personne » (« *Bien que conscient des difficultés théoriques que pose une telle partition, je la maintiendrai cependant en raison* », SCL034), mais tout concourt à présenter les choix de cette dernière comme scientifiquement réfléchis (au double sens de réflexion et de réflexivité par rapport à la construction du savoir).

Ainsi en va-t-il également dans l'extrait suivant : « *mon opinion (non ma doctrine...)* » et dans des expressions figées caractéristiques de la phraséologie du genre (« *à notre connaissance* », « *à notre avis* », « *me semble-t-il* », « *selon nous* ») ; ces éléments fonctionnent comme des commentaires méta-énonciatifs (J. Authier-Revuz, 1995), ou, dans la perspective de R. Vion (2004), comme des modalisations par lesquelles l'assertion se trouve dédoublée et commentée ; ici, la modalisation restreint la valeur de certitude de l'assertion, en référence à l'auteur en tant que sujet épistémique, i.e. à l'origine de la construction du savoir. La nuance introduite peut au final donner à l'assertion un caractère convaincant, en la présentant comme réfléchie et relativisée.

Cette fonction est très nette avec le commentaire apporté par les parenthèses dans « *mon opinion (non ma doctrine...)* », où les marques de polyphonie se multiplient ; en jouant, linguistiquement parlant, sur le statut épistémique d'« *opinion* », elles assoient la qualité proprement scientifique de l'assertion ainsi introduite¹.

Les emplois des marques personnelles peuvent affecter le statut épistémique d'une assertion, et au niveau global du texte, l'analyse de son objet et la définition des opérations cognitives et langagières de la recherche. Elles impliquent dans le texte l'auteur en tant que chercheur. Avec les formes d'auto-citation, c'est même la prise en charge d'un processus continué de la recherche qui caractérise cet auteur². Instigateur d'une démarche d'analyse, il peut par ailleurs être présenté dans le processus même de la recherche.

En association avec le passé composé, les formes de *je* et de *nous* se prêtent bien ainsi à l'exposé de la méthodologie adoptée, exposé caractéristique des articles de SCL par rapport à ceux de LET : « *nous avons travaillé sur des traces...* » SCL002, « *dans les contextes que nous avons pu repérer* » SCL044, « *dans un corpus que nous avons dépouillé*

¹ Le phénomène est très proche de l'extrait d'un article de SCL que nous avons examiné à propos des termes transdisciplinaires, où le terme « supposer » introduisant une assertion est alors commenté en ces termes dans une parenthèse (« *hypothèse à vérifier d'une manière systématique* ») (SCL025).

² Par exemple « [cela] confirme nos travaux antérieurs » (SCL090).

exhaustivement » SCL044, « *sur le plan de la syntaxe, nous avons relevé que les SN...* » SCL052, « *nous avons ainsi relevé cette variation à Pontoise* » SCL062, « *nous avons effectué d'une part une enquête sur...* » SCL064, « *j'ai décidé de prendre un maximum de verbes en compte pour l'analyse de mon corpus* » SCL033¹.

Intervient alors aussi la mise en scène du chercheur sur le terrain. L'usage du *je* ou du *nous* renvoient au chercheur en tant qu'enquêteur (comme en attestent certains des exemples ci-dessus), mais également en tant qu'acteur social impliqué dans différents champs d'activité liés à sa fonction de chercheur. On le voit par exemple dans l'article cité ci-dessous, où l'auteur est mis en scène à la fois en tant qu'enseignant et en tant que chercheur, qui se fonde sur son activité d'enseignement et la constitue en objet de recherche à des fins praxéologiques :

Je voudrais commencer par une anecdote : travaillant auprès d'élèves en grande difficulté, **je m'**entretenais avec Mourad, réputé mauvais lecteur. (...) Suite à cette « rencontre », partant du postulat qu'on ne lisait jamais seul et que toute lecture devait être socialement partagée, **nous avons essayé, avec les enseignants** des classes qui accueillent ces élèves en difficulté, de travailler sur une mise à plat et un partage des stratégies de lecture. **Je** vais donc présenter maintenant deux situations qui permettent d'entrevoir la pluralité des lectures. Il **nous** a fallu recueillir des traces de cette compréhension. **Nous** avons travaillé sur des traces « ordinaires » et des traces plus « extra-ordinaires ». **Je** commencerai par présenter une situation assez habituelle d'échanges oraux (...) [SCL002]

Le *je* dans ces extraits renvoie alternativement à l'enseignant et à l'auteur du texte, et le passage entre *je* et *nous* rend compte de ce passage entre une fonction de prise en charge du texte et de son objet, et le rôle de l'enseignant travaillant en collaboration avec ses collègues. L'auteur peut également être mis en scène en tant qu'acteur social pour signaler le rôle politique du chercheur, comme nous l'avons vu en traitant des temps verbaux au chapitre 5. (avec l'article SCL011, traitant des réformes de l'orthographe, et où l'auteur dit avoir été appelé par le ministère à prendre en charge ce dossier politique).

En résumé, les emplois du *je* et du *nous* académique en tant que marques de personne servent dans l'article la représentation d'une instance auctoriale qui se caractérise par la prise en charge du processus de la recherche, de la construction du savoir et du texte qui en rend compte. L'acteur scientifique est ainsi impliqué dans le texte non seulement comme l'auteur à l'origine des assertions, mais aussi comme chercheur « en action », dans la démarche de découverte, et parfois aussi en tant qu'acteur social.

Parce qu'elles sont privilégiées dès lors qu'il s'agit de montrer la recherche dans son processus, les marques personnelles doivent pouvoir intéresser l'analyse des formes de mise en intrigue de la découverte que nous avons évoquées à propos des temps verbaux et des ponctuations. On en voit un nouvel exemple ici : « *à première vue ces analogies pourraient venir nous aider dans notre tâche de compréhension, mais force est de constater (...)* » (LET067). L'usage du *nous* semble particulièrement concerné et il reste alors à discuter des emplois non strictement académiques de *nous*, et en particulier de l'implication du lecteur dans cette personne linguistique.

7.2.5. La question de l'inclusion du lecteur dans le « nous »

Avec le *nous*, la question se pose dans nombre d'occurrences de savoir si elles sont plutôt « writer-oriented » ou doublement « writer- » et « reader-oriented », autrement dit s'il s'agit

¹ On est très proche ici de l'emploi précédemment décrit consistant à rendre compte de la définition de l'objet.

d'un *nous* académique, représentant dans le texte son auteur, ou d'un *nous* qui désigne un collectif au sein duquel sont inclus l'auteur et possiblement son lecteur¹.

Dans les balises méta-textuelles, nous avons vu avec les temps verbaux que le sémantisme du verbe joue pour beaucoup dans la désambiguïsation des formes de *nous*. Ces balises ont pragmatiquement un rôle de guidage du lecteur. Linguistiquement, cet effet pragmatique peut se construire de diverses manières ; nous l'avons évoqué ci-dessus, le *je* est également concerné. Quoique de manière diverse en fonction de la configuration linguistique, l'énoncé tend semble-t-il à se présenter comme adressé : l'auteur et le lecteur se trouvent tous deux impliqués à titre d'images dans le jeu énonciatif.

Dans « *nous renvoyons le lecteur à l'article de...* », ou dans « *nous tenons à rappeler que...* », le *nous* se présente clairement comme un *nous* académique qui exclut la représentation du lecteur de la prise en charge énonciative de l'assertion. Le lecteur est quant à lui désigné sur le mode du délocuté dans le premier exemple et, dans le second, il est virtuellement présent d'un point de vue syntagmatique avec la possibilité d'un objet indirect pour le verbe « *rappeler* ».

Dans des tours tels que « *nous allons voir* », « *nous pouvons observer que* », « *nous pouvons examiner* », « *il nous reste à voir si* », « *tourbons-nous à présent vers* », « *il nous faut remarquer qu'il se pose plusieurs problèmes* », rien ne permet de dire que l'assertion est le fait d'une énonciation personnelle – dans l'optique où l'on envisage cette dernière comme le fait d'une personne prise dans son unicité. Le *nous* revêt une valeur potentiellement collective et peut être interprété comme une représentation qui associe l'auteur et le lecteur. L'instruction au lecteur est très nette : il se trouve mis en scène dans la réalisation du procès, selon des formes variées (par ex. l'impératif ou la modalisation avec « pouvoir ») ; dans le cas des balises méta-textuelles, le procès peut concerner l'organisation du texte (« *tourbons-nous* »), mais il implique aussi le lecteur dans la construction de l'objet (« *nous pouvons observer que* »).

La question de l'inclusion ou non du lecteur dans le *nous* est complexe. De toute évidence, elle ne tient pas qu'au *nous* lui-même. Elle se constitue dans des expressions relativement figées, caractéristiques de la phraséologie du discours scientifique et dont les propriétés polyphoniques fondent les effets pragmatiques. La rhétorique de la science, dès lors qu'elle introduit l'acteur scientifique dans son texte, met en face de lui son lecteur²; elle l'implique dans la construction du texte et du savoir au moyen de ces expressions routinisées.

C'est ainsi que l'on pourrait expliquer l'affinité observable entre les questions rhétoriques et la personne linguistique du *nous*, comme on en voit un exemple ci-dessous :

N'avons-**nous** pas là une voie qui nous permette de rencontrer Montaigne et de discerner quelque tendance matérialiste dans sa pensée ? Nous ne le croyons pas, pour deux raisons [LET003]

Associé à un verbe de jugement, le second *nous* (« *nous ne croyons pas* ») exclut pour sa part l'interprétation collective laissée ouverte par le premier (« *n'avons-nous pas là... ?* »).

Dans ce second extrait tiré du même texte, le *nous* et l'interrogative accompagnent des verbes à l'impératif :

Montaigne imagine, d'une part, Platon introduisant plus de subtilité dans sa conception, et d'autre part, la raison humaine lui répliquant: "Il faudrait luy dire de la pat de la raison humaine..." (p. 518). Or,

¹ Il existe bien entendu des cas où le collectif indiqué par *nous* n'inclut pas le lecteur, en particulier lorsque le chercheur désigne son équipe de recherche, ou une équipe sur son terrain de recherche (cf. *supra*, une forme de *nous* désignant l'auteur au sein d'un groupe d'enseignants).

² C'est ce que dit E. Benveniste des véritables personnes de l'énonciation (1966), avec l'idée que le locuteur, énonçant, implante l'autre en face de lui dans son énoncé.

quelle est cette "raison humaine"? **Allons, deux pages plus loin, à la fin de cette prosopopée de la raison: que lisons-nous?** [LET003]

Dans ces marques linguistiques : *nous*, interrogative¹, impératif², l'implication du lecteur se multiplie ; ce constat suggère que l'analyse polyphonique pourrait permettre de mieux définir linguistiquement la mise en intrigue de la découverte que l'on a évoquée. Nous avons signalé quelques-uns des procédés qui contribuent à cet effet, et qui sont liés à la mise en scène du chercheur et du lecteur dans le processus de la recherche. Par ailleurs, nous avons vu dans les styles d'auteur avec la question rhétorique par exemple que ces procédés sont pour partie à mettre au compte de pratiques idiomatiques, affectionnées par certains auteurs plutôt que d'autres.

En LET, ces emplois du *nous* qui incluent le lecteur en l'associant à l'auteur font l'objet d'un emploi spécifique à ce champ. Ces emplois font intervenir *nous* pour parler du matériau de la recherche. On observe en effet que l'auteur étudié peut être désigné sous une forme possessive : « *notre auteur* » LET004, « *nos deux auteurs* » LET001. Plus largement, l'analyse de ce que livre le texte étudié affiche une prédilection dans le recours à cette marque : « *Montaigne nous dit* » LET003, « *Machaut nous montre* » LET040.

Dans cet emploi du « nous », une interaction est mise en scène entre l'auteur étudié et son lectorat ; l'auteur de l'article se présente ainsi prioritairement comme un lecteur du texte étudié, et il en va de même pour le lecteur de l'article :

Telle serait, en définitive, parmi d'autres, une des voies royales où choisit de **nous** emmener le Voyage de Montaigne [LET004]

Tout lecteur des *Essais* sait à quel point Montaigne aime à narrer. Son enfance apparaît à travers de nombreux exemples. Pourtant là où **nous** sommes gorgés de confidences, **nous** pourrions rester sur notre faim. C'est que Montaigne se livre au récit du moi, de l'homme: le contraire d'un enseignement (...) [LET002]

La figure de l'auteur et du lecteur de l'article se trouvent ainsi conjointes et réunies dans le collectif des lecteurs de l'œuvre étudiée. L'auteur de l'article est un lecteur parmi d'autres, mais en même temps, il est ce lecteur capable de décoder l'interaction du texte et du lectorat ; il apparaît ainsi comme un interlocuteur privilégié du texte étudié. C'est ce que l'on voit également avec les formes de discours indirect libre dans les commentaires faits sur le texte étudié :

Cependant, cet attachement raisonné qu[e Montesquieu] porte à la monarchie française n'affecte pas son jugement sur la démocratie. L'étude des deux ensembles est, à ce moment-là, encore disjointe. (...) Tout change lorsqu'on passe au présent. L'observation objective cède le pas à des positions beaucoup plus polémiques. **Il faut imposer l'idée de** la possibilité d'une monarchie sagement tempérée par le maintien des privilèges et distinctions traditionnels, mais accueillante aux formes économiques nouvelles et respectueuse de la vie privée et de l'individualité de chacun. **Il faut, pour ancrer cette idée**, agiter l'épouvantail du despotisme, certes localisé dans un Orient imaginaire, mais dont les germes se développent sans arrêt dans la monarchie française, continuellement menacée par ses tendances au nivellement, oublieuse des hiérarchies consacrées dont la disparition menacerait tout le monde et pas seulement les privilégiés. Ainsi, suivant qu'il s'intéresse au passé ou au présent l'attitude de Montesquieu changerait du tout au tout. [LET050]

L'auteur de l'article décode ce que le texte dit à son lecteur, et va jusqu'à adopter momentanément le point de vue de l'auteur étudié pour mieux faire parler le texte. Des

¹ Cf. la partie consacrée aux ponctuations.

² Cf. la partie consacrée aux temps verbaux.

figures multiples de lecteur sont impliquées, et telles que le lecteur de Montesquieu et le lecteur de l'article ne font qu'un.

En LET, cet emploi du *nous* pour désigner le lectorat du texte étudié est très proche de celui qui consiste à inclure le lecteur dans l'article (« *tourbons-nous vers...* »). Cependant il peut être considéré comme plus large, à l'image de la gradation, discutée à propos de *je*, entre un locuteur d'un énoncé et un locuteur être du monde. Tel quel, nous ne l'avons observé qu'une fois en SCL, dans un article sur la traduction, où sa valeur se trouve cependant explicitée : « *nous* » désigne « *les lecteurs ultérieurs* » du roman étudié (« *nous, les lecteurs ultérieurs, risquons de ne pas 'reconnaître' le DIL* », SCL024).

Si cet emploi est caractéristiques des articles de LET, un emploi au fonctionnement énonciatif similaire intervient en SCL, quoique de manière bien plus rare que son homologue en LET : *nous* renvoie à l'auteur et son lecteur, tous deux impliqués en tant que chercheurs du domaine dans un collectif plus large :

Nous souhaitons montrer dans cet article en quoi la synthèse de la parole éclaire la problématique du rythme de la parole et nous confronte aux limites de **nos modèles actuels**. [SCL090]

Les articles de LET font donc usage d'un *nous* qui conjoint la figure de l'auteur et celle du lecteur de l'article dans une même personne linguistique au service de l'analyse du texte étudié, tandis qu'en SCL, le procédé concerne le champ de recherches. Dans les deux cas cependant, l'auteur se trouve ainsi inclus parmi ses pairs. Cela étant – on le voit bien dans l'extrait ci-dessus – l'auteur est également individualisé (« *nous souhaitons montrer* »), et l'assertion se présente comme le fait d'une instance surplombante, qui prend la parole au nom de la communauté de chercheurs. Bien que spécifiques aux enjeux de l'article (son matériau, le champ de recherches), ces occurrences de *nous* sont toutefois à rattacher à l'ensemble plus large où *nous* renvoie à un collectif.

7.2.6. Le « *nous* » comme collectif : le sujet épistémique et culturel

Mettons momentanément de côté les usages de *nous* dont nous venons de traiter, et revenons à l'idée plus générale que les formes de *nous* peuvent prendre une valeur non académique ou collective ; nous avons dit qu'elles sont en cela susceptibles d'impliquer une figure du lecteur dans le raisonnement et la logique de la découverte.

On en voit en particulier un exemple avec les circonstancielles « si... alors » à valeur générique, qui servent à rendre compte de manière relativement codifiée du raisonnement et sont notamment caractéristiques des démonstrations mathématiques :

si l'argument est sélectionné par le nom, **nous** avons affaire à un possessif ; s'il est sélectionné par le verbe, **nous** aurons affaire à un pronom datif ou à un autre type de complément [SCL059]

La personne linguistique prend ici une valeur possiblement générique : le *nous* est supposé désigner tout locuteur, et la valeur collective confère à l'assertion un caractère universel. Avec ce type d'emploi, non seulement le rapport au lecteur mais aussi le statut du savoir se trouvent affectés par la valeur générique de la personne linguistique : le *nous* tend à diluer la personne de l'auteur dans un collectif par lequel son propre point de vue bénéficie de l'appui de l'autorité scientifique. La « véritable personne de l'énonciation » n'est plus un individu empirique, mais une catégorie générique du « sujet épistémique ». Elle est telle qu'elle peut être instanciée par l'auteur comme par tout autre locuteur virtuel ainsi impliqué dans le texte par le *nous* générique.

Ces emplois génériques restent très proche de ceux que nous avons précédemment évoqués, où *nous* désigne le lectorat du texte étudié en LET, et l'ensemble des chercheurs du champ en SCL. Dans les deux cas, *nous* met en scène dans le texte un sujet épistémique, acteur de la construction du savoir.

Dans d'autres emplois génériques, la catégorie désignée par *nous* ne renvoie pas au sujet épistémique, mais à l'humain en tant que sujet que l'on pourrait dire historique et culturel. Ainsi en est-il dans l'extrait suivant,

Le quartier prototypique est « une portion d'un fruit de forme sphérique, facile à découper en quatre parties égales », ce qui renvoie à l'usage que **nous** faisons des fruits [SCL063]

et dans ce second exemple, qui se caractérise par un glissement entre le *nous* générique du sujet épistémique (« *nous prenons une série* », « *nous constatons* ») et le *nous* générique qui renvoie à l'être humain (« *nos attentes* », « *nous imposent* ») :

Si **nous** prenons une série de noms comme la suivante (...), **nous** constatons que **nos** attentes par rapport au monde **nous** imposent dans un degré différent le besoin de placer les entités désignées dans un rapport de contiguïté avec d'autres entités [SCL059]

L'usage du *nous* générique pour désigner tout être humain est fréquent en LET (« *ce monde matériel existe sans nous ou en soi* », « *nous sommes immédiatement au monde* » LET003, « *la colère [...] loin de nous pousser à la quête du plaisir ou de notre profit* » LET045), ce qui converge avec ce que nous avons dit du concept du « moi », en tant que concept très présent dans l'analyse du texte littéraire.

On voit cependant que la valeur générique « tout être humain » peut faire l'objet de spécifications historico-culturelles (« *ce que nous appelons l'âge adulte* » LET002, « *il s'agit de reconnaître la distance qui nous sépare de l'Antiquité* » LET001, « *notre physique contemporaine* » LET062) ou encore, dans ces deux articles consacrés au dictionnaire de Furetière :

Dans un écho singulier avec **notre** propre pratique du dictionnaire, Furetière définit ainsi une des fonctions de son ouvrage [LET068]

Si spirituel s'oppose de façon simple à corporel, moral est plus ambigu. **Notre** familiarité avec le couple conceptuel du concret et de l'abstrait, absent sous la plume de Furetière, pourrait fausser la compréhension de l'expression « en morale », qui désigne très souvent dans le dictionnaire tout autre chose que le domaine éthique. [LET058]

De tels emplois relativisent nettement la propriété d'autonomie par rapport à la situation d'énonciation et la portée universalisante du texte scientifique, en montrant que le lectorat tel qu'il est impliqué dans le texte par l'usage du *nous* est un lectorat historiquement et culturellement défini. On voit aussi la difficulté qu'il y a à distinguer un emploi du *nous* en tant que collectif socio-historiquement spécifié (les contemporains de l'auteur), comme par exemple dans « *notre physique contemporaine* », de l'emploi plus nettement circonscrit au lectorat de l'article en tant que lectorat aussi du texte étudié (« *Montaigne nous apparaît comme un pur produit de la civilisation latine* », LET003).

A partir de la diversité des emplois de *nous*, plusieurs possibilités d'analyse sont offertes : la première consiste à identifier des emplois bien distincts, et à voir dans le pronom *nous* soit le seul auteur, soit l'auteur et son équipe de recherche, soit l'auteur et le lecteur, soit

encore un collectif (*nous* générique)¹. Il semble cependant plus intéressant de tenter de rendre compte du passage des uns aux autres. Mais la question se pose de savoir si dans cette perspective, il est pertinent de prendre comme point de départ une valeur « transparente », où la marque de personne renvoie à une personne énonciative prise dans son individualité (*nous* académique), ou aux deux personnes énonciatives que sont le locuteur et l'interlocuteur.

De tels exemples nous semblent signaler l'intérêt qu'il y a à penser prioritairement la valeur collective et potentiellement générique de *nous* ; elle se trouve spécifiée par le co-texte, dans des emplois qui se caractérisent par une exclusion graduelle : à partir de la valeur collective, les instructions du co-texte sont telles qu'elles permettent de spécifier le référent jusqu'aux cas où *nous* désigne un seul individu empirique à l'origine de l'acte d'énonciation (l'auteur, ou l'auteur cité dans le cas du discours direct qui oblige à exclure l'interprétation « auteur du discours citant »).

En tout état de cause, la diversité des emplois au sein d'un seul texte oblige à envisager les marques de personne sous l'angle de la continuité des valeurs ; si une perspective normative pourrait inciter à demander à l'auteur scientifique d'éviter d'utiliser ces marques sinon d'en faire un usage non ambigu, le parti pris descriptif qui est le nôtre permet de convenir que l'article n'échappe pas à l'ambiguïté et la polysémie caractéristiques du langage. Plus même, la gradation des valeurs prises par la personne linguistique a sa part de fonctionnalité dans la construction du savoir et de sa validation dans le texte. Les emplois de *nous* signalent une dilution linguistique de la personne énonciative prise en tant qu'individualité, et en cela une dilution de la responsabilité (H. Nølke, 1994) énonciative des assertions. Elle montre comment le sujet se constitue dans la possibilité qu'il a d'être le représentant d'un collectif.

7.2.7. Bilan : la multiplicité interprétative de « je » et de « nous »

A l'instar de ce que suggérait M. Foucault (1969b) à propos du texte mathématique², les valeurs du seul *je* dans l'article sont multiples. Il en va de même pour le *nous*, qui pose par ailleurs la délicate question de l'inclusion du lecteur ; le *nous* montre, plus encore que le *je*, la difficulté à établir des frontières nettes entre ses emplois.

Les emplois génériques du *je* permettent d'instancier en SCL un sujet parlant, à l'origine d'un énoncé particulier ou dans sa qualité d'être de langage, notamment dans la pratique des exemples. En LET, *je* et *moi* en emploi générique renvoient au sujet philosophique, et servent à discuter le texte littéraire sous l'angle de la conception du sujet qui s'y dessine. Les emplois qui se circonscrivent à l'auteur de l'article n'en sont pas moins diversifiés, puisque le *je* sert la mise en scène de l'auteur d'un texte, d'une recherche (dans la définition de l'objet, les choix d'analyse), de points de vue et de discussions quant à son propre point de vue. Le chercheur est aussi celui dont la pratique de recherche se prolonge en amont et en aval du texte seul (avec l'auto-citation notamment), celui qui se met en scène dans sa démarche de recherche en SCL (en corpus ou sur le terrain), et enfin parfois aussi un acteur social, hors du seul champ scientifique.

Les emplois du *nous* quant à eux sont loin de se limiter au *nous* académique. Ils peuvent aller jusqu'à renvoyer à tout être humain, mais cette valeur générique peut se trouver restreinte par des précisions socio-culturelles. La valeur prise est également liée au statut épistémique de l'assertion, quand le sujet collectif mis en scène est un sujet abstrait, défini par des opérations langagières et cognitives propres à l'activité de recherche. Le domaine peut en outre se trouver spécifié, le *nous* renvoyant alors au lectorat de l'article, défini comme le

¹ C'est le parti pris dans l'analyse de D. Maingueneau (1994), ce qui peut s'expliquer par la vocation didactique de cet ouvrage consacré à l'énonciation en linguistique française.

² Cf. au chapitre 2.

champ des SCL ou, en LET, comme l'ensemble des chercheurs, et des lecteurs du texte étudié. Enfin, *nous* inclut doublement l'auteur et son lecteur dans la cohésion textuelle (balises méta-textuelles) et dans la mise en intrigue de la découverte.

L'énonciation telle qu'elle est représentée dans le texte fait fonctionner l'irréductible polyphonie de la personne linguistique. Cette représentation ne se calque pas sur les propriétés de la situation *hic et nunc* définie par les trois coordonnées du temps, du lieu et des acteurs. Elle se caractérise au contraire par des glissements tels qu'on observe une dilution de la personne énonciative. De l'auteur du texte à l'auteur en tant qu'être humain, le sujet s'implante dans le texte comme résultant d'une intersubjectivité. Il se présente prioritairement comme un sujet épistémique, acteur de la construction du savoir ; la dilution de la responsabilité énonciative est telle que la figure du lecteur et celle de l'auteur ne font qu'une et dépendent d'une figure plus large, la science ou l'herméneutique littéraire.

Cela étant, d'autres emplois des marques personnelles permettent en même temps une individualisation (dans les formes de *je*, et celles de *nous* académiques, qui renvoient au seul auteur) et la mise en scène d'une identité d'auteur-chercheur au-delà du texte (par exemple le chercheur dans d'autres textes avant et après, dans l'auto-citation, ou le chercheur acteur social).

Au final, on voit donc que la prise en charge énonciative, abordée ici sous l'angle des seules marques de personne *je* et *nous*, repose sur l'implication dans le texte d'une figure auctoriale complexe, qui se caractérise à la fois comme une figure personnelle et comme une figure irréductible à la personne empirique à l'origine de la production du texte. C'est cette question de la prise en charge énonciative dans l'article que nous allons explorer à présent, en proposant de la caractériser par une tension entre effacement énonciatif et prise en charge personnelle.

7.3. Entre effacement énonciatif et prise en charge personnelle

La présence de marques de première personne montre que l'article de recherche n'obéit pas à une propriété attribuée au discours scientifique selon laquelle celui-ci se caractériserait par un effacement énonciatif maximal. L'examen des occurrences montre également que le *je* et le *nous*, pour mettre en scène dans le texte celui qui en est à l'origine, concourent à une représentation complexe de ce sujet : il ne se limite pas à l'idée de « personne » d'une situation *hic et nunc* de l'acte d'énonciation, mais se présente sous différentes facettes, en tant que sujet du texte, sujet épistémique et acteur de la démarche scientifique. Si l'auctorialité se constitue dans la subjectivité du chercheur ainsi mise en scène, elle se définit aussi par la dilution de la responsabilité énonciative que certains emplois signalent. La présence de marques personnelles ne remet pas en cause l'« effacement énonciatif » en tant que régime énonciatif vers lequel tend le texte, et la question se pose de savoir dans quelle mesure et comment effacement énonciatif et marques personnelles se combinent.

Dans un premier temps, nous aborderons cette question d'un point de vue phraséologique, en traitant des constructions récurrentes du *il* impersonnel ; ces constructions nous semblent indiquer la tension qui existe entre effacement énonciatif et inscription d'une subjectivité.

Dans un second temps, c'est d'un point de vue textuel que nous aborderons cette double caractéristique de l'effacement énonciatif et du marquage d'une subjectivité : nous proposons de définir des profils d'articles, selon l'usage que chaque texte fait de *je* et *nous*, du *on*, et du *il* impersonnel. Existe-t-il des articles qui se situent plutôt du côté de l'effacement, et d'autres plutôt du côté du marquage de la personne ?

L'étude est menée sur un sous-corpus de 20 articles par discipline et se base dans les deux cas sur une quantification des occurrences (relayée dans le second par une analyse statistique, dont les résultats sont détaillés en annexes 5.1.).

7.3.1. Le « *il* » impersonnel

Nous avons traité de la 3^{ème} personne linguistique avec l'analyse quantitative des marques de personne au niveau des verbes et des pronoms (cf. au chapitre 5, la partie 5.1.4.). Cependant, les résultats fournis par le logiciel d'analyse ne permettent pas comme nous l'avons vu de spécifier de quoi est faite cette 3^{ème} personne linguistique, qui correspond, d'un point de vue énonciatif, au délocuté, i.e. à une non-personne dans les termes d'E. Benveniste (1966). Nous avons montré avec cette analyse que la 3^{ème} personne du singulier, observée au niveau des flexions du verbe, discrimine les articles des deux disciplines (ceux de LET en font un usage plus abondant que ceux de SCL) et les articles des deux groupes de scripteurs (les non-doctorants y recourent davantage que les doctorants).

Aussi la 3^{ème} personne du singulier nous semblait-elle mériter d'être examinée de plus près, en particulier en termes énonciatifs. Le *il* impersonnel représente dans cette perspective une entrée particulièrement intéressante. Il doit permettre de mieux cerner la tendance à l'effacement énonciatif caractéristique de l'écrit scientifique, et dont il est, avec le passif, un des emblèmes majeurs en français. Une étude syntaxique et énonciative de ses occurrences peut ainsi fournir des données sur l'importance prise par ces tours impersonnels et sur la phraséologie de l'article, dans les deux champs considérés. Quand utilise-t-on un *il* impersonnel ? Et d'abord, comme nous avons vu que la 3^{ème} personne du singulier faisait l'objet d'un usage distinct dans les deux disciplines, est-ce vrai aussi du *il* impersonnel en particulier ?

Les occurrences de « *il* » et le taux de « *il* » impersonnel

Nous avons procédé à un relevé exhaustif des *il* dans les 20 articles de chaque discipline, et distingué les *il* impersonnels des *il* anaphoriques¹. Cette analyse fait état de différences importantes entre les deux disciplines. En valeurs absolues, le corpus d'articles de SCL contient beaucoup moins de *il* que celui de lettres, mais un taux de *il* impersonnels qui est près de trois fois plus élevé :

	<i>Il</i>	<i>Il</i> impersonnel	Taux de <i>il</i> impersonnel
LET	1016	259	25,49%
SCL	604	435	72,02%

Figure 7.1. *Il* et *Il* impersonnel dans les deux disciplines.

En SCL, le nombre plus faible de *il*, anaphoriques et impersonnels confondus est difficilement interprétable : les valeurs absolues dépendent de la taille de l'article, mais ce nombre plus faible de *il* dans les articles de SCL par rapport à ceux de LET doit également pouvoir être rapporté à la gestion de la continuité thématique. Cependant, comme le nombre de *ils*, *elles* et *elle* est au contraire équivalent, il manifesterait une différence au niveau des reprises

¹ On peut également avoir affaire à un *il* cataphorique, lorsque le référent du pronom est situé en aval dans le texte, mais nous parlerons de *il* anaphorique par souci d'économie. Dans les deux cas, le *il* anaphorique et le *il* cataphorique sont des *il* référentiels, et s'opposent en cela au *il* dit impersonnel.

anaphoriques seulement dans le cas des reprises de termes masculin singulier. On aurait alors en SCL une préférence pour des reprises anaphoriques pronominales de type démonstratives (*celui-ci...*) ou pour des reprises nominales. Une autre piste interprétative possible serait d'envisager en LET un goût particulier pour le recours à des clitiques dans des formes d'inversion du sujet (toujours dans le cas de groupes sujet au masculin singulier), notamment dans des incises de type « *écrit-il* » (LET026), dans des constructions avec *aussi* (« *Aussi son fils Alcacer exprime-t-il son souhait de mettre sa cape, une fois de plus, entre le taureau et lui* », LET103), ou *peut-être* (« *Peut-être Thadée est-il...* », LET024), ou encore dans des interrogatives (« *Pourquoi l'auteur fait-il une déclaration sur le sujet de Ferragus si tardivement ?* », LET025). Il resterait alors à comparer l'usage de chacun de ces emplois du *il* anaphorique dans les deux disciplines.

Le taux fort de *il* impersonnel par rapport à l'ensemble des *il* en SCL est quant à lui indiciel d'un mode de prise en charge énonciative spécifique. Alors même que la forme *il* est plus faiblement représentée dans les articles de SCL, elle renvoie dans près de trois cas sur quatre à un tour impersonnel ; dans les articles LET, un *il* sur quatre seulement correspond à ce type de tour impersonnel.

Cette tendance générale, pointée à partir du nombre total de *il* impersonnel dans les 20 articles de chaque discipline se trouve corroborée par l'analyse statistique. Mené sur le nombre de *il* impersonnel en fréquence relative¹, le comptage atteste la présence significativement plus forte de tours impersonnels dans les articles de SCL par rapport à ceux de LET².

En valeurs absolues, un article de SCL totalise ainsi 63 *il* impersonnels, mais ce cas reste cependant marginal et s'explique en partie par la taille de l'article. Plus de la moitié du sous-corpus de SCL (11 articles) comptent plus de 20 *il* impersonnels. Seuls 3 articles de SCL en comptent moins de 10, alors que c'est le cas de 9 articles de LET, où l'on observe même qu'un article ne contient absolument aucun *il* impersonnel, contredisant ainsi les caractéristiques attendues du discours scientifique.

Avec les valeurs en fréquence relative, qui neutralisent la longueur des textes, en moyenne, les articles de LET comprennent 14,2 *il* impersonnels, pour un minimum de 0 et un maximum de 29,7. Les articles de SCL présentent une moyenne beaucoup plus importante de 21,75 *il* impersonnels, avec un minimum de 8,7 et un maximum de 41,95.

Les constructions les plus fréquentes avec le « il » impersonnel

L'importance de ces tours impersonnels au regard notamment de la présence de marques personnelles dans l'article invite à s'y attarder : quand emploie-t-on un *il* impersonnel ? Il nous a semblé intéressant de nous pencher sur les caractéristiques syntaxiques et énonciatives de ces tours ; l'objectif est de repérer des constructions phrastiques privilégiées et d'observer dans quelle mesure le *il* impersonnel s'associe à des formes d'inscription du locuteur, dont le point de vue se construirait ainsi dans une tension entre l'effacement énonciatif des constructions impersonnelles et ces formes de manifestation d'une subjectivité.

Pour ce faire, nous avons procédé à l'aide d'un concordancier (Wconcord) à une analyse des occurrences du *il* impersonnel ; la quantification des mots du cotexte droite et gauche (3 mots à gauche, 4 mots à droite) et l'examen des occurrences en contexte nous a permis de faire émerger les constructions les plus fréquentes, et de voir en quoi le *il*

¹ On procède à un ratio sur le nombre de mots du texte, et on le rapporte à un article moyen de 5000 mots.

² *Anova*, $F=6,002$, $p=0,019$ (Cf. Annexes 5.1.)

impersonnel s'associe à des marques de subjectivité ou de dédoublement des points de vue (par exemple, des marques de première personne comme dans « *il me semble* », des épistémiques comme *sembler*, *vrai*, des axiologiques comme *difficile*, ou encore le phénomène polyphonique de la négation).

Les constructions les plus fréquentes que l'on peut identifier en contexte droit sont les suivantes¹ (occurrences en valeurs absolues, à envisager en regard du nombre total de il impersonnel, i.e. 259 en LET et 435 en SCL) :

Constructions	Exemples	Occurrences	
		LET	SCL
Il + y + avoir	<i>il y aurait d'autres exemples</i>	65	73
Il + être/devenir + adjectif	<i>il est vrai, il devient difficile</i>	39	92
Il + s'agir + de (Détn/Vinf) ²	<i>il s'agit de quatre types de</i>	46	93
Il + falloir +Vinf /Détn/queP	<i>comme s'il fallait aérer le texte</i>	53	44
Il + semble + queP	<i>il semble ainsi que Ferragus</i>	28	47

Figure 7.2. Les constructions les plus fréquentes du *il* impersonnel en contexte droit.

Et, de manière moindre toutefois :

		LET	SCL
Il + convenir ³ +de	<i>il convient de préciser</i>	9	17
Il + exister + Détn	<i>il existe bien un modèle de</i>	4	21
Il en est ainsi ; il en est/ va de même	<i>quoiqu'il en soit, ainsi en est-il des...</i>	9	13
Il + passif	<i>ce qu'il est convenu d'appeler</i>	4	16
Il est + N/Détn	<i>il est question, il est peu de gens</i>	7	11
Il + rester/demeurer + queP	<i>il n'en demeure pas moins</i>	2	9
Il + s'ensuit +DetN/queP	<i>il s'ensuit des méprises</i>	2	7
Il + apparaît + queP	<i>il apparaît ainsi que</i>	3	3
Il + être + à + Vinf	<i>il est à noter</i>	2	3
Autres verbes en construction impersonnelle	Notamment : <i>ressort, suffit, s'opère, s'avère, résulte, paraître, arriver</i>		

Figure 7.3. Les autres constructions fréquentes du *il* impersonnel en contexte droit.

Si ces structures sont les plus fréquentes dans les deux disciplines, on observe cependant des préférences spécifiques à l'une et à l'autre⁴. Ainsi, ce sont les constructions de type « il y a » et le verbe falloir qui dominent les emplois du *il* impersonnel en LET (65 et 53 occurrences), tandis qu'en SCL ce sont les constructions « il + s'agir » (*il s'agit de*) (93 occurrences) et « il + être + adjectif » (*il est vrai*) (92 occurrences) bien que les formes « il y a » soient bien représentées avec 73 occurrences. Outre ces constructions et l'emploi fort du verbe *sembler*, on observe en outre dans les deux disciplines un emploi relativement important de *exister* et *convenir*, et dans une moindre mesure de *s'ensuivre* et *apparaître*. D'autres constructions se

¹ Liste des abréviations utilisées dans les deux tableaux : Détn = déterminant, N= nom, V = verbe, Vinf = verbe à l'infinitif, queP = proposition subordonnée complétive.

² Ces structures mériteraient d'être traitées séparément dans la mesure où « il s'agit de + verbe infinitif » prend un sens déontique que n'a pas « il s'agit de + SN ».

³ Il faut distinguer sémantiquement l'usage impersonnel « *il convient de* », qui revêt une valeur déontique, des autres constructions où *convenir* est synonyme d'*admettre*, avec une valeur nettement polyphonique (« *il faut convenir que...* », « *nous conviendrons de...* »).

⁴ A nuancer toutefois par le fait qu'on s'en tient là à des occurrences prises en valeurs absolues, l'intérêt étant prioritairement de faire émerger des structures récurrentes, et de comparer entre elles leurs occurrences.

présentent comme des expressions lexicalement figées et intuitivement caractéristiques du discours scientifique : *il en est ainsi*, *il en va de même*, ou encore *il n'en demeure pas moins*, et *il est question*.

Notons que les constructions où le *il* impersonnel initie la phrase sont très fréquentes (25,8% en LET, 31,5% en SCL). Concernant le contexte gauche, en LET, la construction *il y a* se trouve fréquemment faire l'objet d'une focalisation (« ce qu'il y a de... »), alors que ce n'est que deux fois le cas en SCL.

On retrouve par ailleurs les constructions mises en évidence en contexte droit, notamment les verbes modaux apparaissant ici avec une inversion du *il* impersonnel, dans des tours non moins caractéristiques de la phraséologie du discours scientifique : *semble-t-il* (9 occurrences dans chaque discipline), *faut/faudrait-il* (6 occurrences en SCL, 4 en LET dont 3 sous la forme « *encore faut(drai)t-il* », et également, en SCL, la formule « *ainsi en est-il* », qui apparaît 7 fois.

Les autres éléments intervenant en contexte gauche sont des pronoms relatifs (*où*, *dont*, *que*) et secondairement d'autres subordonnants (*comme* et *lorsque* en SCL, *lorsque* en emploi égal avec *quand* en LET, *si* en LET principalement, ou encore *que*, notamment dans les focalisations évoquées en LET, et dans *qu'il s'agisse* en SCL, ou l'on observe en outre 5 occurrences de *quoiqu'il en soit*, absent en LET). Enfin, précédant le *il* impersonnel, on relève des connecteurs, notamment *mais* (17 occurrences en LET, 11 en SCL où il est relayé par *cependant*, *pourtant* et *toutefois* (8, 4 et 4 occurrences), *ainsi* (5 occurrences en LET, 10 en SCL) et *aussi* (4 en LET, 5 en SCL). En LET, *et* est relativement fréquent (11 occurrences), davantage qu'en SCL (6 occurrences).

Quant aux verbes en contexte gauche, il s'agit, outre les verbes modaux et les constructions figées qu'on a évoquées, de verbes caractéristiques de balises méta-textuelles en SCL (*signaler*, *souligner*, *noter*, *cerner*, *constater*) outre les verbes *prendre* et *faire*. En LET, de manière plus difficilement interprétable, on trouve *accorder*, *donner*, *assigner*, *diviser*.

Ces constructions syntaxiques dans lesquelles le *il* impersonnel entre de manière privilégiée permettent de dégager quelques éléments figés caractéristiques du discours scientifique (par ex. *semble-t-il*, *il convient de*), ainsi que des habitus peut-être plus proprement disciplinaires (*il existe* et *quoiqu'il en soit* seraient ainsi un trait des articles de SCL peu représentatif en LET ; inversement, la focalisation associée à *il y a*, dans *ce qu'il y a de...* serait quant à elle caractéristique des articles de LET par contraste avec ceux de SCL). Si les occurrences sont trop peu élevées ici pour conclure à des associations collocationnelles d'une part, et pour obtenir un panel plus diversifié de ces tours impersonnels d'autre part, on voit qu'une telle étude menée sur un plus grand corpus pourrait s'avérer utile à la mise en évidence de cette phraséologie scientifique et disciplinaire dans une perspective d'aide à la rédaction de l'article.

Les traits phraséologiques liés à l'emploi du *il* impersonnel méritent d'être abordés également d'un point de vue énonciatif : dans les locutions verbales impersonnelles comme dans les structures attributives, on voit comment l'impersonnel peut servir des formes de modalités aléthiques, ayant trait à ce qui est (ainsi en est-il du présentatif *il y a*, et des locutions impersonnelles *il existe*, *il s'agit de*). Il s'associe à des « subjectivèmes » (C. Kerbrat-Orecchioni) pour conférer à l'assertion ainsi introduite un caractère modal de type déontique, (comme dans *il faut*, *il convient de*), ou épistémique (comme dans *il semble*, *il est possible*, etc.). Le rôle que joue le *il* impersonnel dans la construction du point de vue dans l'article semble ainsi pouvoir être décrit à partir des phénomènes de manifestation de la subjectivité et de polyphonie qui caractérisent ses emplois.

Sur 259 *il* impersonnels en LET et 435 en SCL, on relève un certain nombre de termes récurrents et significatifs quant à la construction du point de vue du locuteur dans ces tours impersonnels :

-en termes d'opposition des points de vue : des négations (52 tours impersonnels concernés en LET, 63 en SCL) et des marqueurs d'opposition (*mais, toutefois, pourtant, cependant*) (18 en LET, 26 en SCL)

-en termes d'inscription de la subjectivité du locuteur au moyen des marques personnelles (*je, nous, on*), en SCL principalement (31 vs 4 en LET), et des adjectifs entrant dans les constructions attributives du type II + être + Adjectif, en particulier les axiologiques positifs ou négatifs (*intéressant, difficile, rare, inutile, frappant, etc.*).

-en termes de nuance de l'assertion introduite par l'impersonnel : des verbes épistémiques comme *sembler* (très fréquent, dans *il semble que* ou dans *semble-t-il*, cf. *supra*), et *pouvoir* (7 occurrences épistémiques concernées en LET, 11 en SCL¹), et les adjectifs épistémiques entrant dans les constructions impersonnelles attributives, en particulier *vrai* (6 tours concernés en LET, 5 en SCL) et *possible* et *probable*.

-en termes de renforcement de l'assertion : les adjectifs entrant dans les constructions attributives et conférant un caractère déontique à l'assertion (*nécessaire, indispensable, obligatoire*), et les adverbes *bien* (13 occurrences en LET, 11 en SCL) et, quoique rare, *évidemment* (2 LET, 1 SCL), le renforcement de l'assertion reposant avec ces adverbes sur le dédoublement énonciatif qu'ils introduisent.

On voit donc clairement comment, alors même que l'impersonnel manifeste un effacement énonciatif, les structures dans lesquelles il entre se caractérisent par une dimension subjective et polyphonique, telle que l'impersonnel joue un rôle argumentatif fort. Les modalités déontiques et épistémiques sont tout particulièrement concernées, et c'est la complexité de la construction linguistique de la dimension modale des assertions qui se manifeste, et avec elle celle de la validation des points de vue assertés dans le texte. Ainsi les injonctions (« *Il conviendrait assurément d'étudier de plus près...* » LET004, « *il faut sans doute aller plus loin et s'interroger sur la nécessaire relativisation...* » SCL048) se présentent comme soumises à la discussion quant à leur certitude et à l'épreuve de la vérité du fait de la confrontation des points de vue qu'introduisent les marqueurs polyphoniques et/ou épistémiques. Ceux-ci affectent la prise en charge énonciative de l'assertion, et le statut « évidentiel » du savoir (Dendale et Tasmowski, 2001).

Ces tours impersonnels sont plus fréquents dans les articles de SCL que dans ceux de LET, ce qui invite à se pencher sur les profils de texte en fonction des modes de prise en charge qu'ils mobilisent. C'est ce que nous allons faire à présent. Un tel constat demande cependant à tenir compte du fait qu'en SCL, les *il* impersonnels entrent pour beaucoup dans des structures présentatives de type *il y a*, ou dans les structures prédicatives *il + exister* et *il s'agir de + SN* ; en LET, il faut toutefois ajouter à la présence moindre de *il* impersonnels le fait qu'un certain nombre d'entre eux interviennent sous la forme de discours indirect libre, dans le cadre de gloses du texte étudié dont nous avons parlé à propos des occurrences de *nous* (« *Il faut imposer l'idée de la possibilité d'une monarchie...* », LET050, « *en temps de crise, il faut savoir reconnaître et écouter les voix prophétiques – telle est aussi l'opinion du Peuple...* », LET042).

Le fait que les tours impersonnels soient plus fréquents en SCL, de même que les marques personnelles *je* et *nous* suggère que s'il existe des profils d'articles variés selon un

¹ Nous parlons d'occurrences épistémiques de *pouvoir* car la même unité formelle peut être utilisée sans cette valeur modale (par ex. avec le sens « être capable de »). Sur ce point, nous pouvons renvoyer aux travaux de E. Thue Vold (2005, 2006,) menés dans le cadre du projet KIAP sur la modalité épistémique dans l'article de recherche.

degré d'effacement ou de marquage personnel, ceux-ci ne se limitent peut-être pas à l'idée qu'on aurait d'un côté des textes caractérisés plutôt par un effacement énonciatif et de l'autre des textes caractérisés plutôt par un marquage personnel.

7.3.2. Des profils d'articles en termes d'effacement et de marquage de la personne linguistique

Nous avons vu avec l'analyse des tours impersonnels la complexité énonciative de l'article : à l'encontre de sa désignation, un tour « impersonnel », caractérisé par l'effacement de la personne énonciative se révèle le lieu privilégié de l'inscription d'une subjectivité modale, et montre ainsi comment l'assertion bénéficie de ces marques subjectives et polyphoniques, et de l'effacement énonciatif qui lui confère un caractère universel.

C'est au niveau du texte à présent que nous allons aborder la double caractéristique de l'effacement énonciatif et du marquage de la personne. Plutôt que de retenir l'ensemble des marques de subjectivité et de polyphonie comme nous l'avons fait avec les constructions du *il* impersonnel, nous restreignons ici les phénomènes pris en compte, en prenant comme point de départ les formes de *je* et de *nous*, et les occurrences du *on*, du *il* impersonnel et du passif dont nous analysons les occurrences dans le sous-corpus de 20 articles par discipline.

Nous avons analysé les valeurs plurielles prises par *je* et *nous* dans l'article, mais il reste à voir comment ces deux personnes linguistiques se distribuent l'une par rapport à l'autre dans les textes : laquelle est choisie pour référer à l'auteur lui-même ? Peut-on parler de choix entre l'une et l'autre, ou des usages mixtes sont-ils attestés ? Et d'abord, tous les textes font-ils usage de ces marques personnelles ? L'usage de *je* et *nous* pose plus largement la question de savoir comment les articles se situent dans des degrés de marquage ou d'effacement de la personne énonciative. Notre objectif est donc d'étudier des configurations d'emplois de marques personnelles et de tours impersonnels dans les textes, et plus précisément des formes de *je* et des formes de *nous* (*je*, *nous* en position sujet et leurs dérivés *me*, *nos*, etc.), du *on*, du *il* impersonnel et du passif ; nous faisons l'hypothèse qu'il existe des profils divers d'article qu'il nous revient de mettre en évidence.

En annexes (annexes 5.1.) sont présentées pour chaque article les données ici traitées relatives aux occurrences de *je*, de *nous*, de *on*, de *il* impersonnel et de passif, ainsi que celles concernant l'effet de la discipline et du statut de l'auteur sur la distribution de ces variables linguistiques.

Nous nous centrerons dans un premier temps sur l'usage des marques personnelles, pour voir combien d'articles sont concernés par les formes d'auto-désignation de l'auteur dans son texte, et pour mieux étudier la question du choix de l'une ou l'autre des marques et de leur possible co-présence. Ensuite, nous essaierons d'envisager des profils d'articles selon qu'ils attestent d'une tendance à l'abondance des marques personnelles, d'une restriction voire d'une absence de ces marques, de la préférence pour des tours impersonnels ou d'une dilution de la personne énonciative dans un *on* polyphonique.

« Je » ou « nous » : quels modes d'auto-désignation de l'auteur dans son texte ?

Un examen des occurrences de *je* et de *nous* en contexte permet de mettre de côté celles qui correspondent à des autonymes, et à des extraits de citation ou du matériau étudié dans l'article ; l'objectif est en effet de se centrer sur les emplois où s'inscrit la voix de l'auteur, de manière diverse selon que *je* ou *nous* servent exclusivement l'auto-désignation, ou renvoient, au pôle opposé, à un collectif (tout locuteur, tout être humain).

Le premier constat que l'on peut faire est que l'usage du *je* (et ses dérivés) est assez peu fréquent, ce qui corrobore les analyses faites de ces marques de personnes dans l'écrit scientifique francophone, par contraste avec l'anglais notamment¹.

Sur 40 articles étudiés, soit 20 de chaque discipline, 6 articles de LET et 8 articles de SCL comportent des formes de *je*, soit approximativement un tiers des articles dans une discipline comme dans l'autre. En SCL, 2 des 8 articles concernés par l'usage de *je* attestent uniquement d'emplois génériques². Les textes concernés par un usage du *je* qui servent l'auto-désignation de l'auteur sont donc au nombre de 6 dans les deux disciplines. 14 articles des deux disciplines présentent ainsi un profil tel que l'auteur ne s'inscrit pas dans son texte au moyen du *je*, du moins ne s'auto-désigne pas en tant que chercheur-auteur.

L'auteur s'inscrit-il alors dans son texte au moyen du *nous* académique ? Hors cas de bruit dans l'analyse (quand *nous* présente un statut autonome, relève d'une citation ou d'un extrait du matériau étudié), 19 articles de SCL recourent aux formes de *nous*, contre 14 articles de LET, et l'analyse menée sur le sous-corpus confirme l'usage significativement plus abondant de *nous* en SCL, et chez les doctorants de LET par rapport aux non-doctorants³.

Si les formes de *je* peuvent renvoyer à l'auteur, elles peuvent aussi comme nous l'avons vu faire état d'usages génériques (tout locuteur, tout sujet). Les valeurs prises par *nous* se situent quant à elles sur un continuum entre les deux pôles de l'individu auteur (le *nous* académique) et un usage où l'auteur s'inclut dans un collectif (par ex. le lectorat du texte étudié dans les articles de LET), qui implique également le lecteur du texte, et peut tendre vers une valeur générique ou apporter une spécification d'ordre socio-culturelle⁴.

Face aux formes de *nous*, la question qui se pose est de déterminer si son usage exclut ou non celui du *je*. L'usage du *nous* est exclusif dans 11 articles de SCL et dans 8 articles de LET. Le *nous* renvoie alors à l'auteur, ou présente, de manière plus complexe, des emplois distincts ou plus diffus d'un *nous* réservé à l'auteur et d'un *nous* collectif voire générique.

Cependant, comme nous avons déjà eu l'occasion de le dire⁵, certains textes font un usage conjoint des formes de *je* et des formes de *nous*, et il importe alors de regarder ces occurrences de plus près. Vérifie-t-on une distinction nette des valeurs prises par *je* et par *nous* ? Ou l'article fait-il au contraire usage d'un mode mixte, recourant à la fois à *je* et *nous* quand il est question du seul auteur ?

8 articles de SCL et 6 articles de LET sont concernés par cet emploi double de *je* et de *nous* dans l'article. Dans la moitié des cas, une distinction nette est systématiquement respectée ; soit *je* présente des emplois générique et *nous* renvoie à l'auteur (SCL036, SCL099), soit, cas le plus fréquent, *je* renvoie à l'auteur et *nous* est utilisé pour sa valeur collective. C'est le cas dans SCL009, qui utilise massivement le *je* (20 occurrences) et recourt au *nous* à deux reprises : le *nous* à valeur collective sert une fois à désigner, sur le terrain, le chercheur inclus dans le groupe formé avec les jeunes locuteurs dont il analyse les pratiques via une démarche ethnographique (« à chaque pique-nique que nous avons fait ensemble »), et une fois à désigner l'équipe de recherche où est inclus le chercheur (« notre équipe »).

L'usage distinct du *je* pour l'auteur et du *nous* pour un collectif montre tout l'intérêt qu'il y a à ne pas envisager l'usage du *nous* dans l'article comme étant réservé à un *nous*

¹ Cf. par ex. I. Vassileva, 1998 et A. Mauranen, 2006.

² Comme dans « Si je ne puis appeler une chose par son nom que si la chose a été au préalable nommée ainsi, je puis désigner, référer à, renvoyer à une chose par une expression sans que cette chose ait été désignée auparavant ainsi » [SCL099] (cf. la partie 7.1. sur les différents emplois de *je* et de *nous*).

³ Les analyses statistiques donnent, pour la différence disciplinaire (*Mann Whitney*, $U=105$, $p=0,0102$), et pour la différence liée au statut de l'auteur dans l'article de LET (*Mann Whitney*, $U=25,5$, $p=0,0452$) (cf. annexes 5.1.). Ces résultats, qui excluent les formes de bruit, corroborent ceux mis en évidence pour l'ensemble du corpus (cf. chapitre 5)

⁴ Cf. la partie 7.1. sur les différents emplois de *je* et de *nous*.

⁵ Cf. l'analyse quantitative des occurrences de *je* et *nous* au chapitre 5 (partie 5.1.4.)

académique qui ne renverrait qu'à l'auteur. En LET tout particulièrement, une valeur nette de *nous* comme collectif incluant doublement l'auteur de l'article et son lecteur se manifeste, par contraste avec l'emploi du *je* lorsqu'il est question pour l'auteur de parler en son nom propre. Ainsi en est-il par exemple dans LET004, où l'on observe ce partage entre un *je* auctorial (« *Le mystère entoure ce que j'aimerais appeler le "côté de chez Louppes" de Montaigne* ») et un *nous* collectif qui renvoie au lectorat du texte étudié, et donc, doublement, à l'auteur et au lecteur de l'article (« *la relation que le Journal nous suggère d'établir entre les pouvoirs miraculeux de l'eau et l'utopie* »)¹.

Dans LET006 et LET107, la même distinction se manifeste entre le *je* individualisant de l'auteur et la valeur collective de *nous*. Les choses se compliquent cependant, car *nous* dans son emploi collectif peut avoir deux types de valeurs, la valeur de « lectorat du texte étudié » et la valeur générique du type « nous êtres humains » (avec une possible spécification d'ordre culturel quant à l'époque ou la société désignée par le *nous*). Ces trois valeurs sont nettement distinctes dans LET006 avec un *je* par lequel l'auteur s'auto-désigne (« *Je commencerai par l'exposé...* »), un *nous* pour le lectorat de Montaigne, (« *Montaigne ne nous donne aucune justification de cette distinction et aucun côté n'est préféré* »), et un *nous* générique (« *Montaigne accepte que les coutumes nous forment, il voit dans le pyrrhonisme la possibilité de se changer* »).

Il est cependant des cas où les formes de *nous* comme celles de *je* renvoient à la personne de l'auteur seule. On obtient donc un usage mixte des deux personnes linguistiques pour l'auto-désignation de l'auteur, i.e. pour la même personne énonciative. C'est le cas dans 3 articles de SCL et dans 3 articles de LET. Autrement dit, parmi les textes concernés par l'usage double de *je* et de *nous* (8 en SCL et 6 en LET), la moitié ou plus vérifie une distinction qui n'est pas établie par d'autres.

Il en va ainsi dans l'article SCL061 par exemple, où l'auteur s'auto-désigne alternativement par *je* et *nous* (« *J'appelle définie dans ce travail, une interprétation pour laquelle...* », « *Sans nous étendre davantage, nous proposons en (32a) une analyse syntaxique...* »). Dans SCL090, il est intéressant de constater que si l'auteur s'auto-désigne alternativement par *je* et *nous*, les formes de *je* se restreignent à des commentaires méta-énonciatifs sous la forme de l'expression figée « *me semble-t-il* ». La distinction de l'usage des deux marques ne concerne donc pas le partage de valeurs (dans les deux cas, il s'agit d'auto-désignation), mais repose sur un partage phraséologique.

Finalement, on obtient en LET des usages parfois très complexes de *je* et de *nous* à l'échelle du texte : c'est le cas par exemple dans LET012, où *je* renvoie à l'auteur dans l'introduction (« *Je pense bien sûr à [l'édition], admirable et courageuse, qu'André Tournon vient de donner* ») ; il est relayé dans la fin de l'introduction et dans le corps du texte par un *nous* qui inclut doublement l'auteur et le lecteur de l'article, renvoyant au lectorat du texte étudié (« *L'occasion nous est enfin donnée de recommencer notre lecture* », « *Nous sommes avertis : l'objet du chapitre n'est pas la conférence, mais la sottise de ses acteurs* ». Dans la conclusion de l'article enfin, le *nous* renvoie à l'auteur, et s'associe à l'implication du lecteur (« *Supposons que le lecteur nous ait suivi dans notre parcours, et qu'il en ait même approuvé les sinuosités : peut-être estime-t-il avec nous qu'en définitive d'un bout à l'autre de sa lecture il n'a pas vraiment progressé* »).

Plus largement, on voit que certains articles présentent un profil tel que l'usage de *je* ou de *nous* se circonscrit à quelques valeurs bien identifiables, tandis que d'autres au contraire se caractérisent par des emplois plus complexes : ils peuvent attester d'une alternative entre *je*

¹ Pour une analyse plus détaillée de cet emploi de *nous* spécifique aux articles de LET, voir la partie sur les différents emplois de *je* et de *nous* en 7.1.

et *nous* dans l'auto-désignation de l'auteur, et d'autres, qu'ils se situent ou non dans ce cas de choix mixte, montrent le caractère diffus des frontières entre les différents emplois de *nous*.

En résumé, si l'on s'en tient aux modes de désignation du seul auteur dans son texte, il est possible de récapituler les profils d'articles à l'aide du graphique suivant :

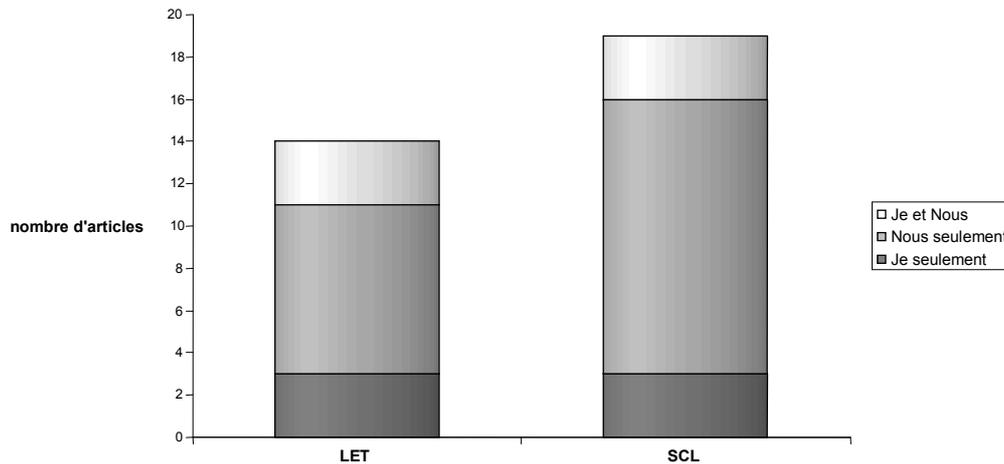


Figure 7.4. L'auto-désignation de l'auteur dans son texte

On voit que le *nous* académique concerne 13 articles en SCL et 8 en LET et domine donc largement par rapport au *je*. L'emploi mixte des deux formes reste rare mais attesté dans trois articles de chaque discipline, et ne semble pas lié au statut de l'auteur.

Un contraste net se manifeste dans le fait que tous les articles de SCL sauf un se caractérisent par l'auto-désignation de l'auteur dans son texte, tandis que 6 articles de LET ne sont pas concernés. Il nous faut alors approfondir ce constat de l'absence dans certains textes de formes d'auto-désignation et plus largement de marques personnelles. Nous proposons d'examiner des profils d'articles sous l'angle de tendances à la restriction *versus* à l'abondance des marques personnelles qui renvoient à l'auteur, et de mettre en perspective ces tendances avec l'usage du *on*, du *il* impersonnel et du passif.

Tendance à l'abondance *versus* tendance à la restriction de marques personnelles.

Nous venons de voir que 19 des 20 articles de SCL se caractérisent par l'auto-désignation de l'auteur dans son texte, que ce soit au moyen de *je*, *nous* ou les deux, contre 14 des 20 articles de LET. 1 article de SCL et 6 articles de LET excluent donc l'auto-désignation de l'auteur. Quant aux articles concernés par l'auto-désignation de l'auteur, le nombre très variable de marques personnelles montre qu'ils ne le sont pas tous au même degré. Se manifestent donc des profils distincts, selon que le texte tend à multiplier les formes d'auto-désignation personnelles, ou à les restreindre jusqu'à les éviter.

Une première différence se manifeste comme nous l'avons dit en fonction du nombre de marques servant l'auto-désignation de l'auteur. Pour les articles concernés par de telles formes d'auto-désignation, deux tendances nettes à la restriction d'une part *versus* à la

multiplication des marques personnelles se distinguent. 7 articles de LET sont ainsi concernés par les formes d'auto-désignation tout en les limitant à moins de 10 occurrences. En SCL, seuls 2 articles présentent ce profil et en contiennent 6 ou plus en valeurs absolues tandis que 4 articles de LET se limitent à 2 ou 3 formes d'auto-désignation de l'auteur.

Au contraire, le nombre d'occurrences des marques personnelles peut se présenter comme particulièrement élevé, et ce dans les deux disciplines. C'est toutefois davantage vrai des articles de SCL par rapport à ceux de LET, et des textes de doctorants par rapport à ceux des non-doctorants, ce qui confirme les tendances envisagées quantitativement et sans restriction quant aux valeurs prises par *je* et *nous*¹.

En valeurs absolues, le nombre maximal est de 28 *nous* en LET, et 6 articles se situent dans un profil d'usage abondant de *nous*, dont 5 articles de doctorants ; cependant, comme nous l'avons vu, le *nous*, tout particulièrement en LET, pose ce problème de frontière entre les emplois strictement réservés à l'auto-désignation et des emplois impliquant doublement l'auteur et le lecteur, et en particulier des emplois renvoyant au lectorat du matériau étudié. En SCL, on note un maximum très élevé de 86 *nous* (en valeurs absolues) et cet usage abondant se retrouve dans un article qui en comprend 74, un autre 51, un autre 44. Au final, 14 articles de SCL contre 6 de LET précédemment cités sont concernés par un usage de *je* ou de *nous* qui dépasse 20 occurrences ; plus encore qu'en LET avec les problèmes de frontières évoqués, ces occurrences tendent en SCL à être exclusivement dédiées à l'auto-désignation du chercheur dans son texte.

Le schéma ci-dessous résume ces deux tendances à la restriction et à l'abondance des marques de personne.

Formes de <i>je</i> et de <i>nous</i>	Tendance à la restriction		Tendance à l'abondance	
	0 occurrences ←	< 6	> 20 occurrences	→
Nombre d'articles LET	6	7	1	6
Nombre d'articles SCL	1	2	3	14

Figure 7.5. Tendances à la restriction et à l'abondance des marques de personne : les formes de *je* et de *nous*

Le schéma montre que les profils d'articles se distinguent dans les deux disciplines, avec une tendance plus marquée en LET à la restriction et une tendance plus marquée en SCL à l'abondance de formes de *je* et *nous* dédiées à l'auto-désignation. Des différences internes à chaque discipline se manifestent cependant, attestant de la diversité possible en la matière, comme nous l'a montrée l'étude des styles d'auteur. Il convient alors de s'intéresser de plus près au pôle caractérisé par la restriction des formes de *je* et de *nous* au service d'une auto-désignation de l'auteur, pour voir si les textes compensent ou non par l'usage du *on*, et en quoi les formes de *il* impersonnel et de passif permettent de parler d'une tendance nette à l'effacement énonciatif.

¹ Cf. chapitre 5.

Usage du « on »

Si l'on s'intéresse aux articles se caractérisant par l'absence de formes de je et de nous, on remarque que certains font en revanche usage de *on*, dont les valeurs, extrêmement complexes, se calquent sur les valeurs possibles de *nous* ; elles posent cependant plus encore le problème de frontières entre des emplois plus auto-centrés, et des emplois collectifs, impliquant doublement l'auteur et le lecteur, et pouvant aller jusqu'à prendre une valeur générique (« l'être humain »). Interviennent également des *on* exclusifs, i.e. où l'implication de l'auteur dans le *on* ne semble *a priori* pas permise, comme dans « *les groupes que l'on dénomme parfois 'constructions détachées' sont souvent traitées comme (...)* »¹ SCL048). Ceux-ci restent cependant très rares, la majorité des emplois de *on* pouvant être située sur un axe entre le *je* de la personne et le *on* générique, avec des valeurs indéfinies la plupart du temps².

Le problème posé par le *on* est très débattu dans le champ actuel des recherches sur la polyphonie linguistique, et tient aux glissements incessants entre les diverses valeurs ; comme pour le *nous*, la saturation de la multiplicité interprétative vient du contexte, mais c'est avant tout semble-t-il la possibilité de ces glissements qui caractérise énonciativement ce pronom.

On pourrait être tenté de voir un phénomène de cataphore dans la dernière phrase, dans une structure syntaxique qui correspond au schéma classique (...). **En réalité**, le référent est mentionné, de façon complète, au début de la séquence narrative [SCL048]

Ainsi, dans cet extrait, le *on*, possiblement interprétable en première approche comme incluant l'auteur sert finalement, avec le conditionnel et le verbe axiologiquement marqué « *tenter* », à introduire un point de vue envisageable (ou tentant) mais auquel l'auteur oppose la « *réalité* » des faits textuels. Le co-texte implique donc une réinterprétation du *on* avec une restriction apportée dans le sens d'une exclusion de l'auteur, et partant du lecteur.

Parmi les articles ne faisant usage ni de *je* ni de *nous*, l'un se distingue par un usage très abondant de *on* (40 occurrences dans LET053). Un autre fait état d'une intervention moindre (13 occurrences dans LET063). Viennent ensuite des articles dont les occurrences de *on* renforcent leur tendance à la restriction des personnes linguistiques (entre 2 et 4 occurrences dans LET032, et SCL086). Enfin, plus remarquable encore en termes d'effacement énonciatif, certains articles de LET ne comprennent aucune forme de *je*, de *nous*, ni de *on* : LET071, LET085 et LET091.

Ces résultats nous permettent d'affiner les profils précédemment décrits à partir des occurrences de *je* et de *nous* ; on voit en effet que les textes se caractérisant par l'absence de *je* ou *nous* peuvent présenter des profils variés selon qu'ils tendent à un recours abondant au *on*, signalant en cela la préférence pour une forme de dilution de la personne linguistique dans un pronom aux valeurs multiples et floues à l'auto-désignation stricte au moyen de *je* ou *nous*. A l'inverse, la restriction de l'inscription de l'auteur dans son texte peut être telle que l'article se caractérise par une absence totale de *je*, *nous* ou *on*.

¹ L'exemple est discutable : l'auteur tend à s'exclure du choix dénommatif qu'il évoque et du traitement fréquent du phénomène ainsi désigné, mais le *on* renvoie au champ de recherches (comme l'usage spécifique du *nous* en SCL traité dans la partie 7.1. consacrée à la diversité des emplois de *je* et de *nous*) et maintient en cela une ambiguïté quant à l'inclusion et la distinction de l'auteur par rapport à l'ensemble des chercheurs du champ.

² A partir des travaux sur le *on*, il est possible de dire que le *on* peut prendre une valeur inclusive *versus* exclusive (par rapport au locuteur), générique ou restreinte (l'être humain ou un groupe plus restrictif au sein de cette catégorie), et définie ou indéfinie dans le cas d'un collectif (un groupe de personne identifiées, par ex. une équipe de recherche, *versus* un collectif non identifié) (cf. par ex. D. Maingueneau, 1994, pour la plupart de ces valeurs).

« Il » impersonnel et passif

Tournons-nous alors vers les formes de *il* impersonnel et de passif, pour mieux cerner encore les profils de texte en termes de tendance à l'effacement énonciatif. Concernant les occurrences de *il* impersonnel, elles sont là encore extrêmement variables. Elles peuvent être absentes (un article de LET), rester faibles avec moins de 10 occurrences (10 articles de LET, 3 articles de SCL), ou à l'inverse prendre des proportions considérables (ainsi l'article SCL048, précédemment cité pour son usage du *on* se caractérise-t-il par 63 occurrences de *il* impersonnel). L'usage du passif est globalement plus homogène, mais des tendances adverses se manifestent dans quelques cas isolés d'usages très rares du passif (2 occurrences dans LET006 et 4 dans LET107), et d'usages au contraire très abondants (51 occurrences dans LET026, 54 occurrences dans SCL087). Comme le *nous*, le *il* impersonnel et le passif sont significativement plus utilisés en SCL qu'en LET¹.

Si l'on croise ces résultats avec ceux précédemment obtenus à propos des formes de *je*, de *nous* et l'usage du *on*, il est possible de mettre en évidence deux profils très spécifiques d'articles.

D'abord, sur le pôle des textes qui se caractérisent par la restriction, voire l'exclusion des formes de *je*, *nous* ou *on*, le *il* impersonnel peut également tendre à un usage faible voire absent. Le cas le plus frappant à cet égard est celui de LET071 : on n'y relève aucun *je*, aucun *nous*, aucun *on* et aucun *il* impersonnel. Il s'agit là d'un cas très particulier, mais il n'est pas moins étonnant de constater que cette tendance se vérifie, de manière moins radicale, dans d'autres textes, à savoir LET085 et LET095 (aucun *je*, *nous* et *on*, et moins de 7 *il* impersonnels, soit bien moins que la moyenne dans chaque discipline). Cette tendance se retrouve aussi dans un article de SCL, SCL084, où l'on ne relève aucun *je* ni aucun *nous*, 4 *on* et 5 *il* impersonnel, de même que dans LET032.

Le profil spécifique de ces textes permet de penser à des usages stylistiques du genre et la restriction est telle que ces pratiques semblent devoir être conscientisées. Comme ces profils sont davantage le fait d'articles de LET que de SCL, on peut se demander également dans quelle mesure la prescription négative que signalent ces profils (éviter l'usage des tours personnels et des tours impersonnels) représente un modèle de référence dans le champ disciplinaire.

Dans LET071, l'article le plus marqué par l'effacement énonciatif ainsi sondé, le texte obéit en termes de marques personnelles au plan d'énonciation historique tel que défini par E. Benveniste (1974). Cependant, alors que l'objet comporte une dimension historique puisqu'il porte sur l'ingénieur géographe Jomard et ses expéditions en Egypte, le texte n'intègre pas le traitement historico-narratif dont nous avons parlé à plusieurs reprises, mais se construit au contraire principalement au présent. On peut voir dans ce texte un modèle canonique du texte scientifique, objectif, essentiellement descriptif et exclusivement centré sur son objet. Une question rhétorique rend toutefois tangible une implication de l'auteur et du lecteur dans le texte. Mais avec l'effacement énonciatif qui définit le texte, l'absence de *il* impersonnel dont nous avons montré qu'il est souvent le lieu d'une argumentation forte atteste de la spécificité d'un tel profil d'article.

L'usage du *il* impersonnel et du passif est au contraire très important dans certains textes. On pourrait en première approche penser qu'ils constituent une alternative inverse à la prise en charge plus personnelle ; nous avons vu en effet avec le *on* qu'il pouvait servir à

¹ Les résultats des analyses statistiques menées sur le sous-corpus de 20 articles par discipline donnent, pour le *il* impersonnel (*Anova*, $F=6,002$, $p=0,019$), pour le passif (*Anova*, $F=4,776$, $p=0,0351$) (Cf. Annexes 5.1.).

compenser l'absence de *je* et de *nous*, tout particulièrement dans LET053 qui en fait un usage très abondant et ne contient ni *je* ni *nous*. Or, les résultats nous obligent à formuler une hypothèse inverse, celle d'un usage fort du passif et du *il* impersonnel dans des textes qui se caractérisent par un usage relativement important des marques personnelles et de formes d'auto-désignation (au moyen du *nous* académique ou du *je*). Le profil le plus frappant à ce niveau est l'article SCL048, qui contient 27 *nous*, 41 *on*, 63 *il* impersonnels et 34 passifs.

Nous avons déjà parlé de son auteur dans l'analyse des styles¹ ; comparés aux autres articles, ses textes affichent comme spécificité nette ce double recours fort aux marques personnelles et au *il* impersonnel.

On voit alors la complexité de cette question du marquage de soi dans les textes : loin de se résumer à une neutralité supposée, elle ne se satisfait pas plus d'une description en termes de degrés plus ou moins importants d'inscription de l'auteur dans son texte. Si certains textes se distinguent par deux tendances opposées à l'effacement et au marquage de la personne dans les textes, force est d'admettre, à un tel degré, qu'il ne s'agit pas d'une pratique contradictoire. Il importe alors de s'interroger sur ses effets, et si l'on se réfère à ce que nous avons dit de cet auteur dans l'étude de son style, elle semble avoir partie liée à l'argumentation, et à la dimension possiblement polémique de ses textes.

SCL048 n'est pas le seul texte concerné par ce double effet de marquage de la personne et de recours abondant au passif et au *il* impersonnel. On le retrouve, quoique à des degrés moindres, dans divers textes de SCL (par exemple, SCL087, avec 28 *nous*, 10 *on*, 22 *il* impersonnels et 54 passifs), mais également en LET (par exemple LET025, avec 25 *nous*, 34 *on*, 27 *il* impersonnels et 32 passifs). Il est difficile de dire combien de textes exactement sont concernés ; il faudrait pour cela fixer une limite, arbitraire, au-delà de laquelle décider que l'usage des marques personnelles et du *il* impersonnel ou du passif est à considérer comme particulièrement élevé.

Plutôt que de chercher à déterminer une catégorie de textes spécifiquement concernée, il nous semble préférable de poursuivre l'approche des textes sous l'angle d'un continuum, et de tendances spécifiques qui structurent leur diversité. On aurait donc là une tendance nette, caractéristique de certains profils, mais par hypothèse transversalement utilisée. Elle consiste en une tension entre effacement et marquage de la personne par laquelle se construit le point de vue de l'auteur. Nous retrouvons ainsi, en tant que caractéristique du texte pris dans son ensemble, ce que nous avons montré au niveau local avec les constructions du *il* impersonnel.

Au final, il semble possible de déterminer des profils d'article en fonction du marquage ou de l'effacement de la personne énonciative mais il importe de renoncer à envisager ces deux tendances comme deux pôles d'un continuum. Au contraire, la complexité des profils attestés fait état de tendances multiples :

1. une tendance à la multiplication des marques personnelles *je* et *nous* et à l'inscription forte de l'auteur dans son texte par ces formes d'auto-désignation ;
2. une tendance à la restriction des marques personnelles *je* et *nous*, éventuellement associée à une tendance à la restriction des formes de *on* et de *il* impersonnel, qui signale donc un penchant pour l'effacement énonciatif, qui poussé à son extrême pourrait indiquer la recherche d'un idéal de neutralité de l'écrit scientifique, alors même que cette notion-là ne saurait être fondée sur la seule question du marquage ou non de la personne, et que la dimension argumentative (Amossy, 2000) de l'article la contredit ;

¹ Il s'agit de l'auteur codé A4 (cf. chapitre 6).

-
3. une tendance à l'usage fort du *on*, et à la restriction des marques personnelles, que l'on interprétera comme une tendance à la dilution de la personne de l'auteur au profit d'une prise en charge collective, relativement indéfinie ;
 4. une tendance double à la multiplication des marques personnelles *je* et *nous*, et des tours impersonnels et passifs, que l'on interprétera comme une tension entre effacement énonciatif et marquage subjectif caractéristiques de la dimension argumentative des textes, voire d'une dimension possiblement polémique.

Nos interprétations de ces tendances resteront textuelles. Il n'y a pas lieu selon nous d'associer aux usages du *je*, du *nous*, du *on*, du *il* impersonnel et du passif des jugements de valeur, tels que la part d'objectivité et de subjectivité, et l'humilité, la modestie ou la fausse modestie du scripteur ou du locuteur représenté dans l'énoncé. Ces valeurs sont culturellement fondées : si l'anglophone voit dans le *je* un tour plus humble que le *nous* académique du français, ce dernier aura tôt fait de rétorquer le contraire, d'autant que le *nous* académique est également décrit comme un *nous* de modestie (Riegel et al., 1994). Les débats qui entourent la question n'en sont pas moins intéressants à relever : ils montrent l'importance attribuée à ces phénomènes et la diversité des modèles de référence en la matière, et posent alors la question de l'intentionnalité du scripteur dans de tels choix. *A priori* aisés à contrôler, sont-ils conscientisés ?

Là encore il est difficile de répondre. L'usage mixte du *je* et du *nous* pour l'auto-désignation du scripteur pourrait être perçu comme un défaut de conscientisation, et pourtant il est le fait de certains linguistes de l'énonciation¹. Est-ce à dire alors que l'auteur concerné par ce jeu de va-et-vient entre *je* et *nous* entend ainsi, tacitement, défendre la thèse de l'hétérogénéité énonciative, qui constitue l'objet même de l'article ? Cette proposition est bien entendu caricaturale, mais, de toute évidence, toutes les hypothèses sont permises dès lors que l'on se situe du côté des effets pragmatiques, et, *a fortiori*, du côté de l'intentionnalité du scripteur à produire de tels effets.

A notre sens, la prudence reste donc de mise face à ces faits textuels. On retiendra d'abord que seul un examen en contexte permet de les cerner : les études exclusivement quantitatives sur les usages du *je*, du *nous*, etc. posent le problème de la désignation de ces formes en tant que marques de personne ; or, leurs emplois génériques sont loin d'être négligeables, et se pose plus fondamentalement un problème de frontières entre leurs emplois. Par ailleurs, qu'ils impliquent le marquage de l'auteur dans son texte ou indiquent un effacement, les faits analysés entrent pour beaucoup dans des tours figés, caractéristiques de la phraséologie du genre ; nous avons vu que l'auto-désignation au moyen de l'alternance *je/nous* pouvait se justifier par ces choix phraséologiques, avec un *je* réservé à « me semble-t-il » un *nous* qui prend différentes valeurs (balises méta-textuelles, *nous* de l'auteur-chercheur à l'origine de choix quant à l'objet d'analyse et au traitement des données, etc.). C'est aussi à ce titre phraséologique que la question de la prise en charge énonciative est intéressante, en dehors de toute idée d'une intentionnalité et d'une éthique de l'exhibition et de la dissimulation dans l'écrit scientifique.

Nous allons à présent délaissier la question de la prise en charge énonciative (marquage de la personne et effacement énonciatif), pour nous tourner vers la représentation que donne l'article de la circulation des discours et des points de vue. Nous traitons des références à d'autres acteurs et d'autres textes du champ de recherches, et à la manière dont l'article rend compte de l'existant. Comme la prise en charge énonciative, cette entrée intéresse l'analyse de la subjectivité du locuteur représenté dans le texte, c'est-à-dire, dans le genre de l'article, la

¹ De même d'ailleurs que pour l'usage très abondant de marques personnelles et du *il* impersonnel, qui concerne cependant un autre auteur.

figure de son auteur. Elle permet en particulier de la décrire sous l'angle de la traçabilité des références, et sous l'angle de la construction dans l'article de son univers de référence.

7.4. La représentation de la « circulation des discours » dans l'article

Après avoir traité de l'instance auctoriale telle que représentée dans l'article sous l'angle des marques personnelles et de la prise en charge énonciative, nous abordons à présent les traces de l'autre dans le texte, et la représentation de la circulation des discours. Nous nous basons sur une conception polyphonique du texte, et renonçons donc à l'idée qu'on peut rendre compte de la circulation des discours à partir de seuls segments isolés de « discours autres » (la citation et plus largement le discours rapporté).

Nous nous intéressons ici aux points de vue présentés comme ayant été produits dans le champ social, auxquels est donc attribuée une source locutoire ; on aura ainsi affaire à des références dénominatives, qui renvoient à des acteurs du champ ou à des textes, et à des modes de désignations de l'existant qui permettent de voir comment celui-ci est cartographié dans l'article (à l'aide de noms de courants ou d'approches par exemple). Cette étude sur la gestion du renvoi aux sources et sur la construction d'un univers de référence dans l'article permet de mieux cerner comment le locuteur se positionne dans le texte, et comment son objet se trouve positionné par rapport au champ.

Nous commencerons l'analyse de l'article avec la traçabilité dans le renvoi aux sources, en prenant comme entrée les références dénominatives, dans toute leur diversité, et notamment les remerciements, rubrique secondaire qui montre cependant comment l'article et son auteur se présentent comme pris dans un réseau d'échanges.

Nous poursuivrons avec la multiplication des références (d'ordre livresque cette fois), et la complexité des dispositifs sémiographiques d'explicitation en discutant d'une forme de tentation d'exhaustivité dans le renvoi aux sources.

Nous centrant alors sur les « discours représentés », nous montrerons que leur gestion scientifique demande à identifier nettement les voix en présence.

Puis nous quitterons les seules références dénominatives aux auteurs et aux textes, en les considérant comme des modes parmi d'autres de désignation des points de vue attestés dans le champ, autrement dit de l'existant. Nous montrerons ce que ces modes de désignation suggèrent quant à la figure de l'auteur et mettrons en évidence des différences quant au statut de l'auteur.

7.4.1. Une quête de traçabilité dans le renvoi aux sources

Le premier constat que l'on peut faire n'a pas de quoi surprendre, l'article de recherche est un écrit où foisonnent les références. Celles-ci sont hétérogènes à double titre : d'abord, les formes qu'elles prennent sont variées ; le renvoi aux sources se construit selon des dispositifs sémiographiques complexes dont nous avons esquissé quelques traits à propos des notes et des bibliographies. En outre, les sources auxquelles renvoie l'auteur n'ont pas toutes le même statut.

Les références que l'on dira dénominatives (avec nom d'auteur ou renvoi à un texte) peuvent être le fait d'autres auteurs du champ disciplinaire. Mais les « discours autres » ainsi référencés relèvent aussi du matériau qu'étudie l'auteur, et qui, en LET comme en SCL est un matériau langagier : comme nous l'avons vu en traitant de cet aspect de l'article¹, on trouve, dans les articles de LET, des références à des textes littéraires ou historiques ; dans ceux de

¹ Cf. l'étude des exemples et des citations au chapitre 4.

SCL sont concernés les exemples dès lors qu'ils ne sont pas fabriqués par l'auteur lui-même mais se présentent comme attestés ou fabriqués par d'autres, ainsi que les textes (y compris les transcriptions d'interactions) ou encore les témoignages de locuteurs sur leurs pratiques langagières.

Nous avons déjà traité la diversité des matériaux utilisés dans l'article et montré comment analyser des types d'exemples en fonction précisément de la question du mode de renvoi aux sources ; ici, nous évoquerons parfois les exemples ou les citations du texte étudié, mais nous nous centrerons essentiellement sur les références aux auteurs du champ ; c'est ce que F. Grossmann (2003) propose de traiter en termes de « discours se référant à des auteurs », avec, comme il l'explique, un sens restrictif ainsi attribué à « auteur » pour désigner « ceux qui ont à faire avec la discussion scientifique menée dans l'article, à l'exclusion d'auteurs d'exemples, ou liés au corpus analysé » (op.cité : 256).

Les seules références aux auteurs du champ manifestent elles-mêmes une grande diversité. En témoignent les modes de désignation, qui feront l'objet spécifique de la partie suivante. Même si l'on s'en tient aux références dénominatives, la nature des discours autres ainsi mis en scène est extrêmement variée : le discours des auteurs du champ auquel l'article renvoie n'est pas qu'un discours compilé dans des textes publiés. Aux articles et ouvrages récents ou anciens, sous forme livresque ou numérique, s'ajoutent en effet des références à des séminaires, des conférences ou des tables-rondes qui n'auraient pas donné lieu à une publication. Il peut aussi s'agir d'échanges entre collègues, désignés comme tels ; comme nous allons le voir, les références y compris à des échanges participent de cette représentation que donne l'article du processus de sa constitution intertextuelle et interpersonnelle dans ses différentes strates.

Les références à des discussions entre collègues sont attestées à plusieurs reprises dans les articles des deux disciplines. Elles interviennent pour signaler en quoi l'échange, heuristique, a servi à initier une proposition, voire l'investigation menée, comme dans les deux extraits suivants :

Un collègue mathématicien m'ayant dit que l'ajout d'une ou plusieurs dimensions contribuait souvent à faciliter la résolution de problèmes, **il m'est venu à l'esprit** la comparaison suivante : vouloir intégrer les structures formelles des langues des signes dans le moule formel des langues orales serait un peu comme si, à l'heure actuelle, et malgré l'existence de l'avion, on persistait à faire entrer tout problème de déplacement fonctionnel d'un point à un autre dans le cadre structural de la géométrie plane (avec ses ponts, ses tunnels, ses lacets...). [SCL051]

L'étude qu'on va lire est la réponse à une double provocation. Je ne l'aurais sans doute jamais entreprise si je ne m'y étais pas senti poussé en quelque sorte par **une question que m'a posée James Supple au cours d'un échange épistolaire relatif aux Actes du colloque de Glasgow**. Notre très estimé **collègue britannique** se disait curieux de savoir si j'avais jamais songé à appliquer la méthode dite "philologique" à l'essai Des cochons. Est-ce qu'une telle lecture, demandait-il, donnerait des résultats? Est-ce qu'elle serait efficace? ("Would it work?"). Il faut dire que cette question de bon sens et de bonne foi suscitait en moi une réaction quasi viscérale, car elle ravivait des doutes qui m'avaient souvent assailli tant sur les limitations théoriques que sur les possibilités pratiques de cette façon quelque peu particulière d'aborder le texte de Montaigne. [LET020]

Ce second exemple nous semble particulièrement intéressant ; il est le fait d'une introduction d'article, et sert une présentation spécifique de la problématique de recherche : le rôle heuristique de la discussion scientifique est mis en avant, et ce rôle est traité en des termes psychologisants, dans un récit au passé qui intègre sous la forme de discours direct la parole supposée du chercheur cité ; l'authenticité de cette parole est mise en scène non seulement à l'aide de précisions relatives au cadre épistolaire des échanges, mais de manière tout à fait

théâtrale avec une incise « (« would it work ? ») » qui n'a d'autre fonction que de se présenter comme une réduplication fidèle de l'expression idiomatique du « collègue britannique »¹.

Dans d'autres cas, les références à des communications entre collègues semblent davantage se rapprocher des formes et des fonctions caractéristiques du renvoi aux autres auteurs dans le cas des sources livresques, comme on en voit deux extraits ci-dessous :

L'incise dans le discours citant apparaît comme une caractéristique des œuvres littéraires produites à l'étranger et traduisent, selon le traducteur littéraire Mistefa Aydogan*², [...]

* Communication personnelle [SCL023]

Ma camarade du groupe Ci-Dit Sophie Marnette me rappelle que le tour T5 est assez commun dans les prologues des chansons de geste, où il est question de présenter les événements à raconter comme étant déjà accomplis (voir Marnette 1998, 1999). **Dans ce cas, ajoute pertinemment ma collègue,** la formule de topicalisation "permet d'insister sur le thème premier du récit tout en présentant les événements comme faisant déjà partie de la mémoire collective, une formule que l'on devrait donc moins trouver dans les autres genres littéraires". [SCL021]

Ces références à des communications personnelles se présentent comme plus ou moins adossées au champ scientifique selon que le collègue concerné est ou non nommé, et plus ou moins conventionnelles au regard des renvois plus standardisés aux textes des auteurs du champ. Pour paraître anecdotiques, elles nous semblent avoir ceci d'intéressant que, comme les formes de renvois plus canoniques, elles concourent à une représentation spécifique de la construction des savoirs en tant que construction dialogique (doublement interdiscursive et interlocutive).

Elles radicalisent les attributs propres à l'ethos scientifique, en mettant en avant l'honnêteté intellectuelle en termes de dette³, et en donnant de l'auteur l'image d'un auteur scrupuleux quant au signalement et à l'authentification de ses sources. En outre, l'auteur se présente ainsi comme quelqu'un d'impliqué dans le champ social de ses recherches. L'auteur d'un article n'est pas seulement un lecteur, fréquentant assidûment des sources documentaires, il est aussi un acteur engagé dans son champ, nouant des conversations avec ceux qu'il appelle ses collègues et qu'il va jusqu'à présenter comme des « *camarades* » (SCL021). Pris dans un réseau de relations qui peuvent nourrir son travail de recherches, il est en contact avec des personnes susceptibles d'informer ses travaux (SCL023), et ouvert même aux discussions hors des frontières disciplinaires comme celle avec le collègue mathématicien cité dans le premier extrait.

Il faut bien entendu insister sur le fait qu'il s'agit là de représentations, qui prennent on le voit des formes diverses, et qui, du reste, ne sont pas nécessairement fidèles à l'élaboration effective du travail de recherche. Ainsi, dans le premier extrait, le nom du collègue mathématicien n'est pas utile à mentionner, de sorte qu'en contrepartie, rien ne permet de dire que ce collègue existe. Il n'en reste pas moins que cette seule mise en scène rend compte comme d'autres d'une construction – même reconstruite fictivement – du savoir basée sur le dialogisme, et sert d'argument d'autorité qui justifie la comparaison introduite par l'auteur.

La représentation donnée dans l'article de sa constitution intertextuelle se manifeste également dans les rubriques dédiées aux « remerciements » ; fréquents dans d'autres genres

¹ On pourrait dire aussi « so british ».

² Nous utilisons l'astérisque pour indiquer les appels de notes dans les extraits d'article que nous citons, et la note correspondante.

³ C'est un des traits dont parle P.N. Campbell (1975), à propos de la « persona » du discours scientifique, sans toutefois fonder son analyse sur de tels faits linguistiques.

de textes scientifiques tels que les ouvrages et les thèses, on les retrouve ici en fin d'article en annexes, ou, plus souvent, en notes, en général en début d'article (l'appel de notes peut se situer à la fin du titre, ou dans la partie introductive). Il s'agit pour l'auteur de signaler sa reconnaissance à d'autres¹ dans la construction de l'étude et de l'article, de manière relativement formelle² :

Un grand merci à Ignace. Merci également à C. Schnedecker dont les conseils avisés nous ont aidée à améliorer ce travail. [SCL107]

L'auteur remercie madame le Professeur C. Volpilhac-Auger pour ses lectures attentives et ses suggestions éclairantes. [LET051]

Je remercie George Hugo Tucker et Alexandre Tarrête de m'avoir suggéré depuis lors de nouvelles pistes de recherche, en direction respectivement de saint Augustin et de Platon. [LET009]

Merci au relecteur anonyme qui nous a signalé que le *prédicat de réalité* de Benveniste renvoyait déjà à la même chose [SCL099]

Les remerciements portent pour certains sur le financement de la recherche (« *Je remercie le Fonds FCAR pour sa contribution financière à l'ensemble de ces travaux* » SCL031). D'autres concernent le matériau de l'étude, et s'adressent aux acteurs du terrain (« *Que les collègues qui nous ont accueillis dans leur classe et qui se sont prêtés à l'expérience soient ici remerciés* » SCL001), et plus largement aux informateurs, mais aussi à ceux qui ont contribué à la mise en œuvre du protocole (techniciens du son, réalisation de transcriptions, etc.), ou encore aux auteurs du champ qui sont à l'origine des données exploitées dans l'article (« *Je remercie Claire Blanche-Benveniste et son équipe de m'avoir permis l'accès à leur corpus* » SCL033). Ces formes-là ne sont représentées que dans les articles de SCL.

D'autres remerciements, assez fréquents dans les deux disciplines, portent sur des sources documentaires conseillées ou fournies par des auteurs (« *je remercie Mme Liliane Dulac de m'avoir signalé ce dernier article* » LET044). Les remerciements les plus nombreux concernent la relecture de l'article, anonyme ou non, et l'aide ainsi apportée en termes de correction linguistique (« *Je remercie le Professeur Alain Bidaux pour son aide précieuse dans la rédaction de ce texte en français* » LET008), ou, le plus souvent, en termes de remarques sur le contenu.

De manière rare, on trouve également un cas de remerciements adressés par l'auteur aux responsables éditoriaux du numéro pour l'avoir sollicité à publier son texte : « *Je remercie Catherine Schnedecker et Anne Theissen qui m'ont proposé de publier cet article dans ce recueil* » (SCL110) ; l'auteur d'un article paru dans la revue d'un laboratoire remercie ce dernier de l'avoir accueilli, en présentant son texte comme une forme d'hommage à cette collaboration³. On relève enfin un cas particulier aussi, où l'auteur adresse ses remerciements pour les « *belles recensions* » dont a fait l'objet son ouvrage, et pour les encouragements que

¹ C'est là encore un trait de la « persona » du discours scientifique mentionné par P.N. Campbell (cf. note précédente).

² Le fait de ne mentionner que le prénom comme c'est le cas avec le « grand merci à Ignace » dans le premier extrait cité (SCL107) semble cependant pouvoir affecter le caractère intellectuel de la dette ; il permet peut-être de penser qu'on a affaire à un collègue non chercheur, ou à un proche.

³ « (...) Nous tenons à remercier tout particulièrement les membres de l'Institut de phonétique de l'Université Marc Bloch de Strasbourg, tant enseignants que personnel technique ainsi que ceux du département de linguistique générale de cette Université pour la qualité de leur accueil, chaleureux et amical au sein de cette université dans laquelle nous avons mené l'essentiel de nos travaux de recherche dans les domaines de l'acoustique et de l'articulatoire. Pour cela, avec plaisir et en sincère remerciement, nous espérons que ce travail trouvera une place dans les T.I.P.S. » [SCL083].

lui a faits un autre chercheur du champ¹, de sorte que le remerciement joue à la fois sur la reconnaissance que doit l'auteur aux autres et sur une image favorable donnée de ses productions.

Les remerciements ont manifestement leur importance dans le genre de l'article, ne serait-ce qu'en termes quantitatifs. Ils intéressent une analyse des textes qui décrit ces derniers au titre de pratiques sociales, sous l'angle des relations interpersonnelles qu'ils mettent en scène et de l'identité professionnelle qu'ils contribuent à définir.

En cela, il est à noter que les remerciements se présentent de manière spécifique dans les articles des doctorants, où ils sont particulièrement fréquents. Il s'agit de remercier leur directeur de recherche², voire l'ensemble du département et ces rubriques suggèrent l'impact du statut du scripteur sur l'importance de signaler ce type de dettes intellectuelles.

Par ailleurs, les remerciements signalent le rôle joué par l'autre dans la construction de l'article et/ou de la réflexion ; l'acte performatif implique souvent d'avancer les raisons du remerciement (*remercier X pour ses conseils, pour sa relecture, etc.*), et s'associe ainsi à un certain nombre d'axiologiques positifs : il est question de « *précieux commentaires* » (SCL082), de « *disponibilité et [de] nombreuses lectures soigneuses* » (SCL077), ou encore d'articles qu'un autre auteur a « *gentiment transmis* ». Si ces axiologiques donnent de l'auteur l'image de celui qui sait reconnaître la qualité humaine ou scientifique en l'autre, ils peuvent être plus ambivalents quant à cette image de soi du scripteur, comme dans ces remerciements adressés par un doctorant à son directeur :

J'adresse mes sincères remerciements à Monsieur François Wioland, mon directeur de recherche à l'Institut de Phonétique de Strasbourg, pour avoir relu cet article avec grande patience. [SCL079]

L'explicitation des bénéfices retirés par l'auteur manifeste dans le même sens des images différentes de lui, d'abord selon la nature des bénéfices thématiques. D'autre part, ces images données de l'auteur varient selon que les bénéfices retirés se présentent comme objectifs et indéniables (« *J'ai beaucoup profité des discussions des premières versions de cet article* » SCL020, « *nous la remercions pour ses nombreuses suggestions, dont nous avons tenu compte avec profit* » SCL085), ou comme un bénéfice personnel qui n'engage pas aussi nettement la qualité objective du travail (« *m'ont été très profitables* » SCL078) ; ils peuvent aller jusqu'à faire l'objet d'un dédoublement énonciatif dans une précaution oratoire signalant la possible non-adéquation entre le profit retiré par l'auteur de l'article et le profit objectif observable dans l'article (« *Merci à Renaud Dulong, François Leimdorfer, Nicole Ramognino et Irène Tamba-Mecz pour leurs commentaires qui ont, j'espère, permis d'améliorer ce travail* » SCL063).

Les remerciements intéressent la description de la représentation dans l'article de sa constitution intertextuelle, en thématiquant le caractère interpersonnel de la production et de la paternité du produit (« *Je tiens à remercier Pascale Beller pour la version française de ce*

¹ « Outre le compte rendu d'A. Tournon déjà signalé, j'ai été très sensible aux belles recensions dont j'ai eu connaissance, de la part de Barbara C. Bowen (*Sixteenth Century Journal*, XXXIII/3, 2002, p. 870-871), de Gérard Defaux (*Bibliothèque d'Humanisme et Renaissance*, LXIV, juillet 2002, p. 473-479) et de Colette Nativel (*Société Française d'Etudes du Seizième Siècle*, n° 54, novembre 2001, p. 18). Dans la Préface qu'il a bien voulu accorder à mon livre, Michael A. Screech m'avait déjà prodigué ses précieux encouragements, renouvelés lors d'une conférence donnée à la B.B.C. pour The Open University. Qu'il en soit ici vivement remercié. » [LET021].

² « Cet article est issu de notre mémoire de D.E.A. intitulé "Une semaine au quotidien, Analyse du Discours d'un journal politique", Nantes, 1999. Cette recherche a été dirigée par Mme Olga Galatanu à qui nous exprimons ici notre reconnaissance pour la lecture de cet article. » [SCL026].

texte qui, en quelque sorte, est aussi un peu le sien » (SCL077). S'ils pointent le rôle de l'autre dans le texte et la recherche, ils sont aussi le lieu d'un marquage spécifique de l'auteur en tant qu'acteur professionnel du champ de recherches et de la vie de laboratoire. Ils donnent des images variées de cet auteur dans le partage même de la paternité ; une remarque peut ainsi venir restreindre la portée du rôle joué par les personnes remerciées pour leur contribution, en rendant à l'auteur la responsabilité « *des imperfections qui demeurent* » (SCL007) : « *Bien entendu, aucune de ces personnes n'est responsable des faiblesses de ce travail* » (SCL056).

Enfin, comme les références aux échanges entre collègues, les remerciements se présentent comme un lieu intéressant pour analyser la figure de l'auteur dans le texte, en termes de rapport à son lecteur et aux autres auteurs du champ. Même en dehors de cas isolés où les remerciements semblent servir presque exclusivement de faire-valoir, les structures récurrentes des remerciements montrent l'intérêt en particulier des commentaires apportés, qui désignent le bénéfice retiré en termes de profit personnel, ou de profit objectif, et spécifient à propos de la paternité du texte et de l'étude menée ce que l'article doit à l'autre (par exemple, des matériaux, de précieux conseils) ou à son seul auteur (des faiblesses qui demeureraient).

Les remerciements et les conversations entre collègues peuvent paraître relativement secondaires au regard des sources citées dans l'article de manière plus canonique, lorsqu'une proposition est référée à une autorité extérieure. Cependant, ces phénomènes ont en commun de relever de la représentation que donne l'article de sa constitution intertextuelle et interpersonnelle.

L'hétérogénéité des formes de renvois aux sources qui se manifeste permet de dépasser les caractéristiques attendues (la citation, l'indication d'une référence au moyen du système « Nom d'auteur + date ») ; elle montre en quoi l'article met en scène sa constitution d'une manière qui se présente comme la plus fidèle possible, et qui entend rendre compte de chacune des étapes du travail, depuis l'accès aux données jusqu'aux ultimes remarques des relecteurs anonymes. C'est cette même quête de transparence, de traçabilité voire d'exhaustivité qui se manifeste dans les formes de renvois plus canoniques aux sources.

7.4.2. La multiplication des références : la tentation de l'exhaustivité ?

Les remerciements ou les références à des échanges entre collègues le montrent, les modes de renvois aux chaînons personnels et langagiers qui ont contribué à l'élaboration de l'article sont extrêmement variés. Même dans le cas des renvois aux sources documentaires du champ, la citation n'est qu'une des formes parmi d'autres, et la signalisation des sources passe plus largement par le biais de ce qu'on peut appeler des « identificateurs de référence », c'est-à-dire des noms d'auteurs et/ou d'ouvrages, qui ont pour fonction d'indexer un contenu propositionnel à une source textuelle déterminée.

De manière caractéristique, c'est le système « nom d'auteur + date » qui est utilisé, sous la forme d'une association des deux dans des parenthèses, ou de la seule précision de la date quand le nom est intégré à la syntaxe de la phrase. Mais c'est vrai des articles de SCL plutôt que de ceux de LET. En LET, la configuration est généralement la suivante : précédé ou non de l'indication du nom de l'auteur évoqué, un appel de notes renvoie directement à la référence exhaustive¹ ; parfois aussi, dans les deux disciplines, on recourt dans le corps du texte au titre de l'ouvrage avec le nom de son auteur.

¹ En vertu de la stratégie de renvoi aux sources majoritairement utilisée dans les articles de cette discipline, et dont nous avons traité à propos de la structure des articles et plus précisément de l'exploitation du périphrase Au chapitre 4.

L'indication de la page est loin d'être systématique, mais se présente en SCL en association avec le système « nom d'auteur + date » sous la forme « (Nom d'auteur, date : numéro de page) », tandis que dans les articles de LET, dans le corps du texte le nom de l'auteur seul se substitue au système « nom d'auteur + date », et s'associe à une indication du numéro de page précédée de *p.* (« *M. Bouvier, p. 101* »), LET103) ou de *pp.* quand une série de pages est désignée ; ces indications « *p.* » ou « *pp.* » se retrouvent dans le mode de référencement le plus utilisé, i.e. dans les références exhaustives directement indiquées en notes depuis l'appel de notes dans le corps du texte.

Dans l'article, comme dans d'autres genres de textes scientifiques, la profusion des références est telle qu'il n'est pas rare que celles-ci se trouvent juxtaposées. C'est vrai notamment des articles de SCL, où se manifeste une multiplication des noms d'auteurs,

Si l'existence du DIL dans les romans de Fielding est incontestable (Cohn 1978 : 112, Leech & Short 1981: 332, McDowell 1973: 84) [SCL024]

des dates de parution des ouvrages d'un même auteur,

Ensuite, c'est par l'intermédiaire des proverbes (1989, 1994, 1999a et 2000) et de l'articulation sens-référence-existence (1997 et 1999b) que nous avons retrouvé la problématique dénominateive¹. [SCL099]

ou encore des pages d'un même ouvrage comme le montrent ces trois extraits d'un article de doctorant :

Selon Bruno Poncharal (1998: 90-92, 181, 264-267) [...]

Selon Rolf-Jürgen Orf (1974 : 95-104, 169-170) [...]

Ce serait cependant un anachronisme de classer ces DD monophoniques comme des DDL (Mylne 1979: 53-54, Rosier 1999 : 266, 270, 296) [SCL024]

L'indexation des sources se présente en cela comme systématique, fût-ce au prix de longues précisions relatives à la multiplicité des auteurs ayant abordé une question, des ouvrages d'un même auteur et des différentes pages de l'ouvrage concerné, et on est alors en droit de se demander jusqu'où peut aller une telle multiplication, sans entrer en conflit avec les exigences de lisibilité du texte. En LET, un certain nombre de procédés, attestés également en SCL mais de manière moindre, sont ainsi utilisés pour alléger les notes² où sont indiquées les références aux autres auteurs du champ (notamment *op. cité* et *ibid.*, en plus des abréviations de titres d'ouvrage qui concernent le matériau étudié et non les travaux des auteurs du champ des études littéraires).

Finalement, les dispositifs sémiographiques déployés pour la gestion des références sont complexes, et manifestent à la fois une multiplication de précisions – une explicitation maximale – et la nécessité de procédés économiques. Le péri-texte est prioritairement concerné pour assurer cette gestion des références tout en répondant aux exigences de lisibilité du texte. Le système « intratextuel »³ de gestion des références intertextuelles qui se constitue par

¹ En l'occurrence, il s'agit de textes de l'auteur lui-même. La multiplication des ouvrages pour un même auteur est cependant attestée pour des références à d'autres, mais jamais avec un aussi grand nombre de références dans une même phrase que dans ce cas-là.

² Il faut alors affiner ce que nous avons dit de la gestion des références en LET : la référence exhaustive n'est pas toujours directement mentionnée à partir de l'appel de notes, puisque ce dernier peut renvoyer à une référence abrégée, qui demande alors à retrouver la note où la source textuelle est référencée de manière exhaustive.

³ Le problème de l'internalité du texte qui se manifeste avec ces cas de bibliographies communes aux différentes contributions au numéro explique notre réserve à utiliser le terme « intratextuel ».

l'usage du périphrase manifeste une différence disciplinaire forte comme nous l'avons vu dans notre analyse de la structure de l'article (chapitre 4), avec en LET, l'importance donnée aux notes, et en SCL, des systèmes plus variés, qui impliquent généralement la bibliographie et peuvent ou non passer par les notes.

Cela étant, la représentation que l'article donne de ses sources ne consiste pas simplement en une énumération de sources différentes ; il s'agit également de rendre compte de leurs interrelations, que l'on peut traiter, à l'instar du groupe de recherche international et interdisciplinaire sur le discours rapporté « Ci-Dit »¹, en termes de « circulation des discours ».

Un premier aspect bien connu de la représentation dans l'article de la circulation des discours concerne les cas où l'auteur se réfère à un autre en mentionnant un maillon intermédiaire, présenté comme la source qu'il aurait effectivement consultée (« *Jean Pic, cité par M. Bouvier* » LET103). Considérons l'exemple suivant, assez rare : l'auteur indique en notes, de manière presque caricaturale, deux maillons intermédiaires dans l'emprunt d'une citation.

Nølke va même jusqu'à suggérer qu'« il y a autant de conceptions des notions de thème, rhème, foyer, etc., qu'il y a de linguistes qui se sont occupés des structures thématiques et informationnelles de l'énoncé »*²

* Nous empruntons la citation à Siblot (1998: 40) qui l'emprunte lui-même à Wilmet (1997)

[SCL107]

Dans le même sens, la double indexation du matériau étudié permet de référencer les emprunts faits à un auteur d'un exemple ou d'un extrait du texte étudié:

Question frémissement sensuel, d'après Sandra, c'était un signe qui ne trompait pas, *ce geste*. (F. Seguin, *L'arme à gauche*, p. 83, 1990)*

*L'exemple est emprunté à Le Querler (à par.) [SCL107]

Ce même phénomène est attesté dans les deux disciplines lorsqu'est mentionnée l'édition originale d'un ouvrage dans les références à une édition ultérieure. C'est enfin un problème similaire que pose la circulation des concepts et des citations d'une langue à l'autre ; ils peuvent être traduits, mais en ce cas, la traduction s'accompagne presque systématiquement de la version originale. En témoigne l'exemple suivant, où la traduction elle-même est référée à son auteur :

Le médiatif* en bulgare [...]

*преизказано наклонение [...]. Le terme français appartient à Zl. Guentchéva [SCL027]

Dans cet écrit multi-référencé qu'est l'article, la gestion des références se présente comme scientifique en ce qu'elle rend tangible une quête de traçabilité : le renvoi aux sources, précisément documenté, va jusqu'à représenter la constitution de l'article dans ses différentes strates ainsi que la circulation des concepts et des discours.

¹ Cf. <http://www.ulb.ac.be/philoserlifra/ci-dit/index.html>

² Nous utilisons l'astérisque pour indiquer les appels de notes dans les extraits d'article que nous citons, et la note correspondante.

Enfin, qu'il s'agisse de rendre compte de la circulation des discours et des concepts, ou simplement de la multiplicité des approches qui ont traité une question, c'est une quête d'exhaustivité qui caractérise la représentation que donne l'article de ses sources, comme en atteste encore la note dans l'extrait suivant, tiré d'un article qui se présente comme une « *petite incursion dans la jungle topicale* »,

Il faut aussi souligner le(s) rapport(s) délicat(s) entre les dénominations thème et topique*

* Certains utilisent la forme *topic* [SCL107]

où les difficultés terminologiques et définitoires des notions de thématisation et de topicalisation conduisent l'auteur à juger utile de spécifier les différences linguistiques qui affectent les désignations.

L'exhaustivité est bien entendu factice ; alors même que certains procédés tendent à aller dans ce sens, nombre d'autres permettent à l'auteur de préciser qu'il n'entend pas épuiser le champ des travaux existants dans les quelques sources qu'il cite, de même qu'il ne prétend pas épuiser la question abordée. Ainsi en est-il des commentaires relativement fréquents comme « *à ma connaissance* », « *parmi d'autres études* », « *par exemple chez [tel auteur ou tel autre]* » etc.

C'est donc d'abord une quête de transparence et de traçabilité (rendre compte des sources qui ont été mobilisées dans l'élaboration de la réflexion menée dans l'article) qui nous semble constituer une norme guidant les pratiques d'écriture de l'article, et d'un point de vue praxéologique l'activité de recherche elle-même.

À côté de cette quête, un commun accord permet de reconnaître en l'exhaustivité son caractère utopique : l'article ne sera jamais le dernier maillon d'une chaîne linéaire, exhaustivement reconstituée comme telle, d'un savoir cumulatif. Pourtant ça et là peut affleurer l'idée d'un état des lieux intégral, ou encore celle de la source première : faire la somme des savoirs, ou remonter en arrière jusqu'à identifier l'origine d'une notion ou d'un débat sont des quêtes paradoxales, dont on se défie, mais sur lesquelles la rhétorique de l'article semble parfois jouer, en particulier dans la recension des travaux existants en introduction¹.

Si un tel rapport au savoir, encyclopédique, n'est sans doute pas le seul fait du texte et du champ scientifiques, il pourrait être utile de voir comment les étudiants notamment se positionnent par rapport à cette idée, et dans quelle mesure aussi les manuels méthodologiques thématisent comme une exigence, ou comme une utopie ce que le chercheur peut être convaincu d'avoir comme mission, en vertu de l'« *illusio* » spécifique du champ (P. Bourdieu, 1994).

Pour poursuivre la question du renvoi aux sources, dans les cas de références dénominatives (à un auteur et son ouvrage), nous nous intéresserons à la manière dont l'article gère sémiographiquement l'intégration des « discours autres », représentés comme tels, que sont les travaux de pairs ou le matériau étudié. Des stratégies spécifiques interviennent qui nous semblent pouvoir être rattachées à la transparence et la traçabilité qui caractérisent – à titre de quête – la gestion scientifique des « discours représentés ».

¹ Cf. par exemple ces premiers mots d'une introduction, où l'amorce se présente comme visant un panorama exhaustif des « *différentes théories de la lecture* », panorama restreint ensuite aux cinquante dernières années : « *Diverses conceptions fondent les différentes théories de la lecture. Leurs formulations jalonnent la recherche sur la communication écrite des cinquante dernières années. Tantôt elles concernent exclusivement la lecture de l'œuvre d'art littéraire, telles les théories de Jean-Paul Sartre, Lucien Goldmann, Hans Robert Jauss [...] Tantôt, [...]* » [SCL005].

7.4.3. Identifier les voix en présence : l'attribution énonciative des discours représentés

La proximité dans l'article de sources hétérogènes demande, outre le système de renvois intra-textuels favorisant l'identification des références, à gérer également l'attribution énonciative du dit. La représentation que donne l'article de ses sources va passer par des stratégies qui visent à distinguer, toujours dans le sens de l'explicitation maximale, ce qui relève du texte source, ce qui est à mettre au compte de l'auteur, et ce qui s'est passé lors de l'intégration d'un texte à l'autre. Il s'agit d'éviter tendanciellement la confusion entre les voix, en identifiant nettement chacune d'elles.

C'est le rôle de la citation que de maintenir distinguées deux voix, celle du discours citant et celle du discours cité, en marquant clairement le passage de l'une à l'autre ; mais un ensemble de procédés l'accompagnent, qui vont permettre à l'auteur d'explicitier les modifications apportées au texte source, ou à l'inverse, la fidélité de la représentation qu'il en donne.

Ainsi en est-il de l'usage des crochets à l'intérieur d'un segment citationnel, et des commentaires mis entre crochets, ou parfois entre parenthèses, ou encore en notes. Il s'agit de signaler des modifications linguistiques apportées au texte cité lors de son intégration dans l'article, ou de préciser si le marquage typographique (italiques, caractères gras) relève du texte original ou de cette intégration.

A côté des mentions bien connues comme « *je souligne* » (19 occurrences dans le corpus), et « *nous soulignons* » (44 occurrences), avec des variantes nombreuses (« *c'est nous qui soulignons* », etc.), citons deux extraits pour rendre compte de la mise en œuvre effective de ces procédés¹ :

"Je peux toujours m'évertuer [les italiques sont de Bres] à les dissimuler [les discours de l'autre] en ne les citant pas ouvertement (...)" [SCL021]

D'où l'hypothèse de B. Poncharal : **[L]a distance entre l'oral et l'écrit est plus importante en français qu'en anglais.** De même que **la distinction entre plan du récit et plan du discours est plus nette en français qu'en anglais.** (Poncharal 1998 : 180. Les caractères gras sont de Poncharal.) [SCL024]

Il n'est pas utile de s'attarder sur ces stratégies conventionnelles et routinisées, mais on peut toutefois se poser la question de la surcharge qu'apportent ces nouvelles précisions² ; les manuels d'aide à la rédaction scientifique que nous avons consultés les recommandent, mais proposent pour certains aussi d'éviter d'en faire un usage trop abondant. Interviennent alors des procédés d'économie, qui permettent au scripteur de signaler une fois pour toutes le traitement réservé aux extraits cités ; la première modification donne lieu à une note qui définit la convention adoptée pour l'ensemble du texte (« *Nous soulignons. Dorénavant, tous les italiques dans les citations sont nôtres* » LET049, « *C'est nous qui soulignons en italiques (ou en romain) les indices repérés dans les extraits des corpus que nous citons* » SCL090).

Plus surprenant peut-être est le fait que certains auteurs estiment nécessaire de transcrire fidèlement le texte source et le spécifient (par exemple, « *les italiques sont de Folkart* » ou « *les caractères gras sont de Poncharal* » dans SCL024), car la tendance générale caractéristique de ces commentaires consiste au contraire à signaler les modifications. Ce choix de la fidélité dans la représentation donnée du texte source, dans un article (SCL024) qui se caractérise par ailleurs par un ensemble de procédés d'explicitation de

¹ A notre tour de préciser que pour rendre compte de ces procédés, nous reproduisons ici l'exacte typographie des extraits cités.

² Qui peuvent aller jusqu'à spécifier ce qui dans la citation faite d'un auteur du champ est redevable à un autre.

ce type peut être interprété comme une spécificité auctoriale, d'ordre stylistique, ou comme une spécificité statutaire, l'article concerné étant le fait d'un scripteur doctorant.

D'autres stratégies participent des dispositifs sémiographiques complexes déployés pour pallier les risques d'ambiguïté liés à l'attribution énonciative du dire ; ainsi certaines visent-elles à répondre, partiellement au moins, à la polysémie des guillemets¹. C'est le cas dans un article où l'auteur propose de recourir aux crochets pour maintenir distinguées, dans l'interprétation qu'il fait du discours d'une revue politique (*Présent*), ses propres mots de ceux du discours analysé. Les éléments mis entre crochets renvoient donc systématiquement aux extraits de son matériau ; cela lui permet d'éviter une multiplication de guillemets qui renverraient à des voix différentes, car, comme l'explique l'auteur, les guillemets sont très utilisés dans le journal qu'il s'attache à décrire :

Pour le locuteur de *Présent*, les ["jeunes"] vivent dans des [zones de "non droit"], elles sont le lieu de la [barbarie ethnique], sorte de guerre continue. La ["culture"] désigne la culture des ["jeunes"] mais aussi le métissage culturel voire la culture des valeurs dominantes, considérées comme décadentes. Les valeurs dominantes et adverses du FN sont aussi désignées par l'expression ["politiquement correct"]. Le ["politiquement correct"] est véhiculé par ["la grosse presse"] qui désigne les grands médias. Il est aussi porté par les ["anti-racistes"], les adversaires politiques du FN, qui sont vu comme [anti-France, anti-Français et anti-nationalistes]. [SCL026]

Les guillemets posent en effet problème dans le passage du texte cité (texte d'un auteur du champ, ou texte étudié) à leur intégration dans l'article, et il n'est donc pas rare de remarquer l'utilisation différentielle de guillemets simples et doubles, de guillemets français et anglais, voire de l'italique et des guillemets, chacun représentant une attribution énonciative spécifique. C'est ce que l'on observe dans ces extraits de l'article SCL024, déjà évoqué pour ses nombreux procédés d'explicitation dans le renvoi aux sources :

Nous, les lecteurs ultérieurs risquons de 'repérer' ou de ne pas reconnaître le DIL selon notre "code de lecture" (Rosier 1999 : 77) moderne. [SCL024]

selon B. Poncharal (...) qui a fait une analyse contrastive du DIL (et du DIM *alias* les "énoncés 'hybrides'") [SCL024]

Il peut s'agir aussi, comme dans la note de cet article de LET, de distinguer non pas ce qui relève de l'auteur cité et ce qui relève de l'auteur de l'article, mais, au sein d'une citation, ce qui est à mettre au compte de son auteur, et des auteurs auxquels lui-même se réfère :

Bien métaphoriser, disait Aristote, c'est voir le semblable. "L'épiphore est ce coup d'oeil et ce coup de génie : l'inenseignant et l'imprenable. Mais il n'y a pas d'épiphore sans diaphore, pas d'intuition sans construction. En effet, le procès intuitif, rapprochant des choses éloignées, enveloppe un moment irréductiblement discursif [...] Il n'y a donc aucune contradiction à rendre compte de la métaphore successivement dans le langage de l'aperception, c'est-à-dire de la vision, et dans celui de la construction. Elle est à la fois « le don du génie » et l'habileté du géomètre qui s'y connaît dans « la raison des proportions » "*"

* P. Ricœur, op.cit p.248. Les expressions entre guillemets sont d'Aristote in *Poétique*, 1457 b 20.²

C'est de nouveau la circulation des discours qui est ici représentée, et la précision fournie en note permet d'identifier ce qui revient à chacune des trois voix en présence, celle de l'auteur

¹ Cf. les diverses instructions que peut fournir cette marque quant à l'interprétation du segment qu'elle encadre, dans la partie consacrée aux ponctuations dans l'article (5.2.)

² Article d'études littéraires hors corpus : Chauvin, D., (1996), « Des fleurs et des métaphores : les "prestiges de l'analogie et de la différence" dans les œuvres de Proust et de Jaccottet », *Œuvres et critiques*, Romanisches Seminar (Approches bachelardiennes), Peter Lang.

de l'article, celle d'Aristote qu'il commence par citer, et enfin celle de Paul Ricœur dans les quelques lignes citées, et qui cite lui-même Aristote.

Ne serait-ce que pour distinguer ses propres guillemets de ceux des auteurs que l'on cite avec des guillemets, l'exercice est périlleux : les signes utilisés ne font pas toujours système au sein de l'article et, si code il y a, celui-ci n'est pas toujours explicité à l'attention du lecteur. Par ailleurs, les procédés mis en œuvre pour assurer une gestion tendanciellement non ambiguë de l'attribution énonciative du dit ne sont pas le fait de tous les articles. Plus largement, l'article de recherche ne déroge pas à la règle : l'explicitation ne saurait être maximale, et plus, les effets de brouillage énonciatifs induits par la proximité de sources hétérogènes demeurent. F. Grossmann (2002) l'a bien montré dans son étude des modes de référence à autrui dans des articles de linguistique parus dans la revue *Langages* : on ne sait pas toujours exactement qui parle, et le mode de renvoi aux auteurs du champ caractéristique des articles de SCL, i.e. le système « nom d'auteur + date » est particulièrement concerné par ces ambiguïtés : postposé au contenu propositionnel, il ne permet pas en lui-même de déterminer jusqu'où en amont le segment est indexable à la source mentionnée, et ce qui du contenu est le fait du texte source ou de sa reformulation.

Par ailleurs, nous l'avons signalé, la « confusion » des voix peut être recherchée, au sens où elles se fondent en une seule selon le procédé de coénonciation décrit par A. Rabatel (2004a). C'est le cas dans certains articles de LET comme nous l'avons vu : dans la manière dont l'auteur de l'article rend compte du texte étudié, les assertions consistent en des gloses de ce texte, mais peuvent correspondre à du discours indirect libre. Aucune marque ne permet de distinguer les deux voix de l'auteur de l'article et du texte étudié. La coénonciation est cependant toujours provisoire, une assertion à mettre au compte du seul auteur intervenant nécessairement en amont et pour la clore.

De manière générale, la représentation donnée dans l'article de ses sources, tout en répondant à un principe d'économie, tend d'une part à maintenir distinctes la voix de l'auteur des autres voix, et témoigne d'autre part d'une traçabilité des références. Mais il ne s'agit que de tendances, qui sont loin de rendre compte de l'ensemble des pratiques effectives. Elles sont fortes cependant, si bien qu'on peut faire l'hypothèse qu'elles fonctionneraient comme des normes de nature prescriptives qui réguleraient les pratiques d'écriture de l'article. Cela étant, on voit aussi qu'elles jouent différemment selon les auteurs.

Par ailleurs, la gestion des sources semble indiquer une conscientisation des pratiques, en raison de la complexité des stratégies sémiographiques et linguistiques déployées et de l'explicitation par l'auteur des procédés auxquels il recourt. Il est vraisemblable qu'une forme de contrôle intervienne pour se conformer à cette prescription qui consiste à gérer scientifiquement ses sources. Mais, dans le cas d'ambiguïtés énonciatives¹, peut-on alors parler d'hétérogénéité « déniée »², dans le sens d'un usage conscient des effets de brouillages énonciatifs? Sans aller jusque-là, il convient à présent d'élargir la question du renvoi aux sources en l'ouvrant à d'autres modes de références que les seules références dénominatives (à l'auteur et/ou son ouvrage). Il s'agit de s'intéresser à la manière dont l'article cartographie son univers de référence, en examinant la diversité des modes de désignations des travaux du champ, ou de l'existant sur la question traitée dans l'article, et de mieux cerner leurs fonctions respectives et leurs effets.

¹ Que nous avons évoquées en références à l'étude de F. Grossmann (2002) sur les modes de référence à autrui chez les experts, avec l'exemple de la revue *Langages*.

² J. Authier-Revuz (1984) parle d'hétérogénéité « masquée » pour rendre compte du fait que l'hétérogénéité constitutive de tout texte ne se réduit pas à l'hétérogénéité montrée en tant que telle dans le texte.

7.4.4. Modes de désignation des points de vue attestés dans le champ, ou comment l'univers de référence est-il cartographié ?

Nous avons traité jusqu'ici des références dénominatives, qui mentionnent un nom d'auteur et/ou son ouvrage, qu'il s'agisse d'adresser des remerciements, de citer des extraits du matériau étudié, ou de renvoyer le lecteur aux travaux de pairs.

Plutôt que de nous centrer sur ces références explicites, nous prendrons ici comme point de départ les points de vue qui, dans l'article, se présentent comme des points de vue attestés dans le champ. *A priori* ce sont les points de vue existants dans le champ de recherches qui nous intéressent, mais il peut s'agir plus largement de points de vue existants sur l'objet et circulant dans le champ social, et non exclusivement dans le champ scientifique, la frontière entre les deux n'étant pas toujours nette.

Les références dénominatives sont une des manières de désigner la source de ces points de vue ; elle leur assigne une source locutoire, documentaire ou non, sous la forme d'une désignation rigide : la source locutoire est identifiée de manière non ambiguë, à l'aide d'un nom propre (celui de l'auteur, couplé à celui de l'ouvrage dans le cas des références exhaustives, que l'on peut dire canoniques dans le discours scientifique). Si la référence dénominative exhaustive est une forme attendue du discours scientifique, et caractéristique de l'article comme nous l'avons vu, d'autres points de vue sont impliqués dans le texte en tant que points de vue attestés dans le champ, sans faire l'objet cependant d'une désignation rigide.

Un point de vue présenté comme un point de vue attesté dans le champ scientifique (ou éventuellement social) demande pour être considéré comme tel à ce que sa prise en charge par une source locutoire soit impliquée dans le sens de l'énoncé. Ainsi, pour reprendre l'exemple emblématique des approches énonciatives, dans « ce mur n'est pas blanc », le point de vue « ce mur est blanc » auquel s'oppose un second point de vue (« pdv 1 n'est pas valide ») ne se présente pas comme un point de vue produit et attesté. Par contre, nous avons vu avec le conditionnel que le point de vue « Paul viendra demain », impliqué dans « Paul viendrait demain » n'est pas seulement impliqué en tant que point de vue, mais en tant que point de vue virtuellement pris en charge, qui peut être ou avoir été attesté ou attestable. Si l'on ajoute à cet énoncé au conditionnel une précision de type « selon Marie », « on dit que », etc., le pdv1 « Paul viendra demain » se présente comme un point de vue actualisé, c'est-à-dire attesté ou en d'autres termes effectivement pris en charge par une instance désignée « Marie », « on », etc.

C'est précisément cette question de la désignation d'une telle instance, présentée comme la source locutoire d'un point de vue qui va nous intéresser ici. Nous proposons d'examiner comment sont désignés dans l'article les travaux auxquels l'auteur se réfère, en montrant que ces modes de référencement sont plus ou moins explicites et remplissent des fonctions différentes.

Il s'agit donc d'abord de rendre compte de la diversité de ces modes de désignation, et de voir par là comment l'article cartographie le champ dans lequel il s'inscrit : l'univers de référence qui se constitue dans ces modes de désignation est fait d'auteurs, de textes, mais aussi d'écoles ou de courants, ou, de manière moins spécifiée encore, d'approches, d'analyses, voire enfin de points de vue dont il est dit qu'ils circulent dans le champ, ou ont circulé, sans que leur statut épistémique ni leur source locutoire ne soient davantage spécifiés. En permettant de voir comment l'auteur nomme et catégorise l'existant (i.e. les points de vue « déjà-là » par rapport auxquels il définit et analyse son objet), ces modes de désignation doivent également intéresser l'analyse de la figure de l'auteur dans son texte.

Dans cette perspective, nous traiterons la question de savoir en quoi les articles des doctorants présentent des modes de désignation spécifiques des points de vue par rapport aux articles des non-doctorants. L'hypothèse que nous avons faite avec F. Boch et F. Grossmann¹ est que le doctorant, en tant que « nouvel entrant dans le champ scientifique » (Bourdieu, 2001 : 95) a besoin de se conformer davantage aux conventions du champ académique, et produit donc plutôt des formes de désignation canoniques (i.e. les références exhaustives à un texte).

Afin de mettre l'accent sur cette variable statutaire, l'étude est menée sur le seul corpus d'articles de SCL, et plus précisément sur un sous-corpus dédié à cette analyse, et composé de 22 articles de doctorants et autant d'articles de non-doctorants. Ce sous-corpus englobe le sous-corpus des 20 articles de SCL utilisé pour la comparaison avec les 20 articles de LET. Avec la nécessaire restriction des textes retenus pour analyser en détail ces phénomènes, une autre raison préside à la centration sur les seuls articles de SCL : le fait que dans les articles de LET, les références à d'autres travaux du champ se satisfont globalement bien d'une description attendue, en termes de références dénominatives (nom d'auteur et référence exhaustive renvoyant à l'ouvrage) dont nous avons traité jusqu'ici. Si les aspects quantitatifs portent exclusivement sur les articles de SCL, nous ne nous interdirons pas toutefois d'évoquer dans l'analyse qualitative les articles de LET, lorsqu'ils sont concernés par certains modes de désignation que nous avons identifiés.

Typologie des modes de désignation des points de vue attestés dans le champ de recherches

L'univers de référence qui se constitue dans l'article implique des noms d'auteurs et des renvois à des ouvrages, qui permettent de baliser le champ à partir de travaux emblématiques. Avec les références exhaustives de ces travaux, dont on a parlé dans les parties précédentes, leur désignation est rigide, et leur identification maximale. Cependant, la traçabilité des sources tend à se dissoudre dès qu'au renvoi à une source documentaire se substituent d'autres formes de désignation, qui évoquent des « auteurs », des « conceptions », « certains travaux », ou encore des acteurs flous, « ceux qui » défendent telle ou telle position, et des points de vue attestés dont la source est passée sous silence, quand « il a été dit que » ou quel tel phénomène « a été considéré comme », sans que l'article ne permette de dire par qui, sinon quand le point de vue a-t-il été produit dans le champ social.

L'examen des modes de désignation des points de vue attestés nous a permis d'identifier 6 catégories ; celles-ci ont été mises en évidence à partir d'un premier travail de repérage de la diversité des modes de désignation dans le corpus, qui a été repris et affiné avec F. Boch et F. Grossmann que nous tenons à remercier ici pour cette collaboration. C'est avec eux que nous avons étudié les articles sous l'angle de cette typologie, en nous centrant sur la question du positionnement du scripteur en fonction de son statut. Le travail présenté ici reproduit assez largement les analyses de (Rinck, F., Boch, F. et Grossmann, F., à paraître).

1. Une première manière de désigner les références consiste à citer uniquement le nom de l'auteur, sans indiquer d'autre précision dans le reste du texte quant à la date ou au document concerné. Le nom de l'auteur peut apparaître sous une forme substantivale (« *d'après Michel Péroni* »), ou sous une forme fléchée, adjectivale (« *la théorie searlienne* »).
2. Le second mode de désignation identifié est le mode de désignation que nous dirons canonique ; c'est un document qui est cité (qu'il s'agisse d'un article ou d'un ouvrage, sur

¹ Rinck, F., Boch, F. et Grossmann, F. (à paraître).

support papier ou numérique). Au nom de l'auteur s'associent des données éditoriales : la date, le titre, la maison d'édition, le lieu d'édition. On a alors affaire à une référence exhaustive. Bien entendu, le mode de désignation est à envisager au niveau du texte pris dans son ensemble, car la référence exhaustive peut intervenir comme nous l'avons dit en notes, ou dans la bibliographie, voire au-delà des bornes de l'article dans le cas de bibliographies communes au numéro. Ce mode de désignation est donc identifié comme tel dès lors que le système de renvois « intratextuel » permet d'aboutir à une référence exhaustive : « (*Ducrot, 1995*) », « *selon Sperber et Wilson, 1989* », « (*Chevrot et al., 1983*) », etc.

3. Le troisième mode de désignation est assez proche du second, mais concerne les cas où c'est une liste de documents qui est désignée. Il est fait consécutivement référence à plusieurs documents soit d'auteurs différents, soit d'un même auteur : « *Voir Goudailler, 1997, 2002* », « (*Cf. Rossi, 1971, Di Cristo, 1978*) », etc.

Les références données au niveau du texte pris dans son ensemble sont exhaustives, mais le procédé est un peu différent du second cas dans la mesure où les différents documents sont cités conjointement, pour une caractéristique qu'ils ont en commun.

4. Avec la quatrième mode de désignation, on quitte les références dénominatives traitées dans les parties précédentes. La désignation ne concerne plus un titre de document et/ou un nom d'auteur, mais des noms de courant ou d'école (« *la praxématique* », « *selon les cognitivistes* »), de discipline ou de domaine de recherche (« *une classification rhétorique* », « *en linguistique du nom propre* »). Sont ainsi intégrées à cette catégorie les désignations d'approches dès lors qu'intervient une spécification quant à l'objet de recherche (« *les travaux attachés à la description de l'antonomase* »).

5. Le cinquième mode de désignation concerne les désignations généralisantes, sans spécification du domaine de recherche : « *la diversité des approches* », « *les analyses* », « *certaines définitions* », « *ces conceptions* », « *ces positions* », « *deux interprétations majeures* ». Le point de vue dont il est question est traité on le voit en des termes épistémiques : la désignation implique que le point de vue a été produit dans le champ scientifique. Ainsi intègre-t-on à cette catégorie des désignations généralisantes de la source, pourvu que la source en question soit mentionnée, et renvoie à un ou des locuteurs que l'on dira épistémiques (« *les auteurs qui* », « *certaines considèrent* », etc.)

6. Le sixième mode identifié se caractérise au contraire des précédents par un effacement de la source locutoire (ou les sources) à l'origine du point de vue. Les tours concernés renvoient à un « on » collectif, sans auteur ni document explicitement mentionné. Le point de vue se présente cependant comme attesté, sous l'effet de l'emploi de termes qui signalent la prise en charge de ces points de vue dans le champ social : « *cet emploi faisant l'objet de longs développements* », « *on a appelé les noms de lieux* ». Le passif est fréquemment observé dans ces formes d'effacement de la source (« *une question déjà bien souvent soulevée* », « *le nom propre est le plus souvent considéré comme* »), et, plus rarement, le *il* impersonnel (« *il a été longtemps envisagé* »).

En résumé, nous distinguons donc :

- un premier mode de désignation qui met l'accent sur l'auteur,
- un second qui correspond à la forme canonique de référence à un document,
- un troisième pour les listes de documents,
- un quatrième qui renvoie à des noms de courant, d'écoles ou de domaines,
- un cinquième aux désignations généralisantes,
- et un dernier pour l'effacement de la source des points de vue attestés.

Le dénombrement de chaque mode de référence à d'autres travaux, et du nombre total de références¹ dans les 44 articles de SCL (22 articles de doctorants et 22 articles de non-doctorants) fournit les résultats suivants. Ces résultats ici résumés sont présentés de manière plus détaillée en annexes (annexes 5.2.).

		<i>Moyenne</i>	<i>Ecart-type</i>	<i>Minimum</i>	<i>Maximum</i>	Effet du statut de l'auteur ²
Nombre de références	D	39,5	31,6	8	121	oui
	Nd	73,3	66,8	2	294	
Nom d'auteur	D	4,2	5,3	0	21	oui
	Nd	9,9	15	0	70	
Renvoi au document	D	17,9	15,6	0	53	
	Nd	28,1	36,6	0	162	
Liste de documents	D	2,6	2,2	0	7	
	Nd	4,2	5,3	0	18	
Nom de courant	D	7,8	13,7	0	59	oui
	Nd	16,7	14,5	0	56	
Etiquettes généralisantes	D	5,5	4,7	0	16	
	Nd	9,7	11,1	0	37	
Effacement de la source	D	1,5	2,4	0	10	

Figure 7.6. Modes de désignation des points de vue attestés dans le champ (articles de SCL).

Les résultats montrent d'une part une tendance dans les textes de doctorants à un nombre moindre de références par rapport aux textes de non-doctorants. D'autre part, ils signalent une différence significative également au niveau des références avec le seul nom d'auteur, moindres là encore dans les textes des doctorants, alors que l'usage des références canoniques n'est quant à lui pas discriminant. Enfin, une dernière différence se manifeste avec l'usage des noms de courant, caractéristique des pratiques expertes par opposition aux pratiques des néophytes.

Nous commencerons par commenter les formes de désignation avec nom d'auteur, en regard du mode canonique de référence exhaustive à un document. L'opposition entre les deux groupes de scripteur dans l'usage des noms d'auteur nous semble en effet valider l'hypothèse d'un rapport spécifique à la norme chez les doctorants. Nous nous intéresserons ensuite à l'usage des noms de courant, qui mettent en scène dans le texte une vue surplombante sur le champ. Pour finir, nous ouvrirons quelques pistes de questionnement sur le lien entre certains modes de référencement et la dimension argumentative, voire polémique de certains articles.

¹ S'agissant d'analyser les modes de désignation des travaux évoqués dans l'article, nous avons mis de côté les reprises anaphoriques pronominales.

² Les analyses statistiques de l'effet de la variable statut de l'auteur donnent, pour le nombre total de références (*Mann Whitney*, $U=140,5$, $p=0,0279$), pour les références avec seul nom d'auteur (*Mann Whitney*, $U=139$, $p=0,0446$), et pour les références à un nom de courant (*Anova*, $F=4,294$, $p=0,0446$) (Cf. Annexes 5.2.).

Référence au document et référence à un auteur : communauté scientifique *versus* communauté savante ?

De manière attendue, la forme canonique du renvoi aux sources, qui veut que l'on recoure à une référence explicite à un document est la plus fréquente ; ce constat est d'autant plus vrai dans les textes de doctorants qui recourent, moins que les autres scripteurs, à d'autres formes de désignation des points de vue, en particulier les références aux seuls noms d'auteurs et les références aux noms de courants.

La référence au seul nom d'auteur, sans précision quant au document concerné permet de dépasser l'idée d'un mode exclusif de renvoi aux sources qui consiste, de manière canonique, à indiquer les références exhaustives d'un texte. Une fois exclus les cas dans lesquels la mention de la source textuelle peut être trouvée en amont ou en aval dans le texte, ces références orphelines, sans mention de l'ouvrage sont relativement rares, y compris dans la catégorie des autres chercheurs. Certains articles présentent toutefois un profil spécifique à ce niveau, et plus précisément SCL035 : ce texte portant sur l'écriture de l'histoire comprend un grand nombre de références, dont 70 désignations avec le seul nom d'auteur, ce qui dépasse le nombre de références canoniques à un ouvrage qu'il comporte (58 occurrences).

Cependant, de manière générale, le recours à ce procédé, pour pouvoir être toléré, semble devoir se circonscrire à quelques configurations spécifiques. Les noms d'auteurs de l'antiquité et d'auteurs classiques sont les premiers concernés, ainsi que d'autres noms de l'histoire moderne ou contemporaine de la discipline, dont ce mode de référence signale la notoriété.

En termes phraséologiques, les tours utilisés se présentent ainsi selon des formats récurrents :

- des dérivations adjectivales (« *La séparation (foucauldienne) entre raison et folie* », « *Le respect des maximes gricéennes* », « *De cette définition découlent, dans un esprit saussurien, plusieurs conséquences* »),
- le recours à des prépositions du type *depuis*, *à partir de*, qui marquent la naissance d'un paradigme ou d'un courant (« *l'on sait depuis Bakhtine* »)
- les structures avec *chez* pour indexer le sens d'un terme à l'appareillage conceptuel spécifique à un auteur (« *la notion de notion chez Culioli* »),
- et, dans cette même visée définitoire, l'usage typique de la locution *au sens de* (« *au sens où Hawlabchs l'entendait* »).

D'autres utilisations des noms propres, plus rares, mettent l'accent sur une manière de dire plutôt que sur le contenu proprement dit : « *le mot est de Georges Duby* », « *pour paraphraser Jean-Claude Passeron* », « *même si, comme le disait Frege dans une boutade célèbre* », « *pour reprendre un mot célèbre de Garfinkel* ». On voit dans ces exemples comment la citation (« *le mot est de* ») relève de formes plus large de commentaires méta-énonciatifs, par lesquels l'auteur de l'article va pouvoir notamment joindre sa voix à celle d'un auteur bien reconnu. C'est particulièrement vrai des tours au conditionnel, que l'on observe en particulier dans des articles de LET, là encore sous des formes relativement figées (« *Pas de punctum, dirait Barthes* », « *Après une enquête sérieuse, diligentée par quelques amis, grands connaisseurs de « vieux mots », comme dirait Furetière* »). Ce qui est intéressant en LET est que ces références au seul nom d'auteur concernent aussi bien le matériau étudié que les travaux des auteurs du champ de recherches comme on le voit avec Furetière dans l'extrait cité ci-dessus.

Ce mode de renvoi non canonique à la source pourrait signaler une faiblesse du système de référencement, et résulterait dans cette optique d'une négligence du scripteur. Cependant, dans leurs travaux sur les modes de référence à autrui dans l'écrit de recherche, F.

Boch et F. Grossmann ont mis en évidence (Boch, Grossmann, 2002, Grossmann, 2002, 2003) que ce procédé qui consiste à omettre la référence explicite, apparemment non conforme à la norme, est en fait une pratique typiquement experte.

Elle manifeste d'une part la capacité du scripteur à utiliser les noms d'auteurs comme « emblèmes de position », témoignant ainsi d'une maîtrise du champ de connaissances : le scripteur sait identifier qui fait référence dans ce champ, et sait se positionner et définir son approche par rapport à ces repères. Par ailleurs, cette pratique permet de faire fonctionner un discours de connivence, appuyé sur l'allusion : les noms d'auteurs ainsi utilisés se présentent comme des repères supposés partagés, qui fondent l'univers de référence de la communauté discursive.

Ce discours de connivence est très clairement mobilisé dans l'extrait suivant, tiré d'un article de LET, où le nom d'auteur lui-même est passé sous silence, au profit d'une référence descriptive, qui n'admet cependant pas l'ambiguïté : le lecteur doit pouvoir la décoder.

Ne faut-il pas voir dans l'opposition des deux parties du poème un montage qui soudain fait éclater les apparences les plus rassurantes ? C'est un « tres doulz paradis » (v. 557) que représente l'abbaye royale de Poissy, si riche, si calme, à la vie si douce et bien réglée : « **luxe, calme et volupté** » **dirait un autre poète.** [LET029]

Au final, ces stratégies ont donc pour première caractéristique de détourner le modèle canonique de renvoi aux sources dans le texte scientifique. On reste dans la référence dénominative, mais avec un renvoi à des noms d'auteur plutôt qu'à des documents. La référence exhaustive fait fonctionner l'explicitation maximale : le lecteur doit pouvoir aller retrouver le document concerné, et ce mode de renvoi au document montre que la communauté du discours scientifique comme le savoir qu'elle produit se bâtissent sur la base de publications.

Ces stratégies nous semblent par ailleurs donner dans l'article l'image d'une communauté discursive « savante ». Tel qu'il est impliqué dans le texte, le lecteur doit convenir que la référence au seul nom d'auteur est acceptable car il doit savoir décoder une allusion fondée sur des connaissances supposées partagées dans la communauté.

Le recours aux seuls noms d'auteur traduit un rapport spécifique à la norme, expert et savant, qui expliquerait qu'il soit le fait des textes des non-doctorants par opposition aux textes des doctorants. Dans ces derniers, on observe l'expression *au sens de*, et, beaucoup plus rarement, les dérivés adjectivaux (« *'face' au sens goffmanien* ») et l'usage de noms d'auteurs sans référence explicite (« *les recherches de Labov et celles de Gumperz* »). Mais ce mode de désignation reste significativement moins utilisé. Parce qu'ils ont à maîtriser la norme, et qu'ils n'ont pas encore le statut qui leur permet véritablement de s'en jouer, les doctorants se conforment plutôt à l'usage conventionnel qui veut que l'on explicite ses sources. C'est peut-être aussi pour cette raison que dans leurs textes, il n'y a jamais moins de 8 « références » - prises au sens large de mise en scène d'un point de vue attesté dans le champ de recherches. Chez les autres chercheurs, qui en utilisent significativement plus, un nombre très restreint de références n'est inversement pas exclu, puisqu'un article en comporte seulement deux. Il convient enfin de mettre ces résultats en relation avec ceux présentés au chapitre 4 à propos des citations : la citation, plus fréquente chez les doctorants que chez les non-doctorants, alors que les formes de renvoi à l'existant sont plus fréquentes chez les seconds montre un rôle discriminant de la citation, que l'on peut comprendre comme un autre indice de ce rapport spécifique à la norme chez les doctorants.

Une vue surplombante sur le champ fournie par les noms de courant

Indicielle de leur rapport aux conventions du genre, la manière spécifique qu'ont les doctorants de se référer à d'autres points de vue semble évidemment liée aussi à leur degré de connaissance du champ. C'est vrai tout particulièrement de l'usage de noms de courants : avec ce procédé, il ne s'agit pas seulement d'indiquer la source énonciative d'un point de vue (la référence bibliographique, le ou les nom(s) d'auteur), mais d'utiliser des étiquettes généralisant des positions théoriques. Une connaissance livresque est nécessaire à qui veut positionner sa recherche dans un champ cartographié au moyen de telles catégories.

Les noms de courant, d'école ou de domaines peuvent se présenter comme communément admis et revendiqués par leurs représentants eux-mêmes (« *l'analyse conversationnelle* »), ou être plus ou moins idiomatiques. Par exemple, *l'approche « formaliste »* désigne dans l'un des textes de non-doctorant une des approches de la politesse centrée sur sa formulation en termes d'actes de langage directs ou indirects : les guillemets permettent ici à l'auteur de souligner le fait que la dénomination ne relève pas d'une terminologie pré-existante, mais d'une création *ad hoc*¹.

En homogénéisant certaines approches ou en créant des lignes de partage entre différentes approches, ces étiquettes produisent une vision scénique du champ scientifique qui, pour remplir une fonction de synthèse, tend à réifier les courants et les problématiques. La connaissance du champ ne suffit pas à expliquer seule l'usage moindre de ces noms de courant chez les doctorants. Encore faut-il que le scripteur se sente légitimé à utiliser de telles étiquettes : elles reviennent à investir une posture spécifique, celle d'un auteur qui s'octroie le privilège de la nomination des paradigmes, et se place au niveau surplombant de celui qui offre une vue panoramique du champ.

C'est aussi en ces termes que l'on peut expliquer les autres différences observées entre les deux groupes : d'abord, le taux plus faible de références chez les doctorants, toutes formes de désignations confondues car leurs textes se limiteraient aux références utiles à leur objet sans toujours les mettre en perspective dans un cadre linguistique plus général ; d'autre part, l'usage moindre chez les doctorants des dérivés adjectivaux (cf. « *l'esprit saussurien* » évoqué *supra*), de désignations de paradigmes à partir d'un nom d'auteur (« *depuis Bakhtine* »), ou de références à un mot d'auteur – qui lorsqu'il est dit « *célèbre* » renforce encore l'effet de structuration et de vision surplombante du champ.

Les différences entre les articles des doctorants et des autres chercheurs ne doivent pas masquer les différences dans chacun des deux groupes : le nombre de références et le type de désignations varient beaucoup d'un texte à l'autre. Ainsi, alors que la moyenne de noms de courants est deux fois plus faible chez les doctorants, c'est un article de doctorants qui en présente le maximum. Cette variabilité intra-groupes permet d'avancer que le statut objectif du scripteur n'est pas seul en cause pour déterminer des différences de positionnement. Outre son statut effectif, la manière dont il se perçoit en tant que chercheur-auteur doit influencer sur cette question de légitimité, et, partant, sur les pratiques d'écriture.

Si le recours aux noms de courant, de domaines ou d'école requiert une connaissance du domaine et implique que le scripteur se sente légitimé à se présenter en tant qu'expert, elles concourent en retour à la légitimation du chercheur dont l'article manifeste une telle connaissance. La posture surénonciative (Rabatel, 2004a) qui caractérise ces dénominations joue un rôle dans l'effet de vérité attaché à la cartographie du champ ainsi produite.

¹ Hormis les guillemets les indications intervenant sur le statut de la dénomination sont rares et un tel partage entre dénomination existantes et créations *ad hoc* suppose donc pour le lecteur des connaissances du champ.

La question de la fonctionnalité des noms de courant appelle alors à se demander en quoi leur usage est lié à des types d'articles, ce qui pourrait expliquer alors les différences observées en fonction du statut du scripteur, mais aussi au sein de chacun des deux groupes. Un compte-rendu de recherche expérimentale et un état des lieux sur une question n'appellent pas *a priori* le même type de référencement. Les thématiques et sous-disciplines mobilisées dans l'article peuvent en cela avoir également un effet. Ainsi, la revue *Travaux de phonétique de Strasbourg* se distingue-t-elle des autres : explicitement dédiée aux recherches phonétiques, en corpus ou expérimentales, et plus rarement, historiques et documentaires, le nombre de noms de courants y est significativement inférieur à ce qu'on trouve en moyenne dans les autres.

Y aurait-il des domaines où la multiplication des références et l'évocation des paradigmes s'imposeraient davantage dans l'article ? Il est des questions qui ont fait l'objet de nombreux travaux ; d'autres appellent un débat scientifique serré car elles sont controversées, divers points de vue s'opposant sur la manière de les aborder et de les comprendre. La multiplication des références et des noms de courant qui en résulterait pourrait ainsi être le fait de domaines ou de thématiques spécifiques mais également de certains courants de pensée ou de certaines écoles qui seraient plus polémiques que d'autres. C'est cette question de la fonction des références par rapport à la visée argumentative de l'article que nous examinerons brièvement pour clore cette étude, en nous centrant plus spécifiquement sur l'enjeu polémique de certains articles.

Des modes de désignation des points de vue au service d'un enjeu de démarcation ?

De manière très schématique, l'observation des articles suggère que les études qui accordent une large place au débat scientifique ont besoin de mobiliser certains modes de désignation des points de vue que ne requiert pas une étude qui entend s'inscrire dans un cadre prédéfini sans le rediscuter. Ce constat nous a conduit à faire l'hypothèse d'un lien entre le débat scientifique et certains modes de désignation.

Pour étayer cette hypothèse, nous proposons de nous intéresser au lien entre les types de désignation et un enjeu qu'on pourrait appeler de « démarcation » et plus précisément de démarcation oppositive : marqué dans l'introduction ou dans la conclusion, cet enjeu consisterait à y annoncer que l'objectif de l'article est d'offrir une approche qui s'oppose à d'autres¹. En cela, la visée argumentative, déclarée au lecteur, comporte une dimension polémique vis-à-vis des points de vue et approches auxquels l'approche de l'auteur s'oppose.

En annexes (annexes 5.2.) sont indiquées les données sur la présence d'un enjeu d'opposition déclaré dans l'article et sur le lien entre ce dernier et les modes de désignation de l'existant.

Notons d'abord que sur les 44 articles de notre corpus, 19 sont concernés par cet enjeu d'opposition, déclaré dans l'introduction principalement. Celui-ci, s'il n'est pas déterminant en ce qui concerne le nombre total de références, a bien, conformément à notre hypothèse, un effet significatif sur le nombre de noms de courant dans l'article².

¹ C'est ce que l'on observe par exemple dans : « *Je voudrais ici défendre la thèse selon laquelle [...]. En formulant cette thèse, je prends à contre-pied l'idée, classique, d'une distinction entre [...]* » [SCL096]. Une autre forme de démarcation, distincte de l'opposition, consiste à insister sur le caractère inédit de la recherche, comme dans : « *les travaux existants laissent tous de côté celui qui doit comprendre [les pictogrammes] : le 'lecteur'. C'est donc autour de ce lecteur [...] que s'organise notre recherche* » [SCL003]. Nous revenons sur ce type d'enjeux dans la partie suivante.

² *Mann Whitney*, $U=132,5$, $p=0,0353$ (Cf. Annexes 5.2.).

Il influe également sur le nombre de références, de type « effacement de la source », qui sont significativement plus nombreuses lorsqu'un tel enjeu est déclaré¹. La fonctionnalité spécifique de ce mode de référence est d'évoquer des points de vue attestés sans mentionner ceux qui en sont les auteurs. Cette pratique répond à une double tendance dans la rhétorique de l'écrit de recherche : en gommant la source énonciative, elle permet d'éviter l'attaque *ad personam* et se prête bien à une attaque en règle contre le sens commun.

Restreinte à un enjeu déclaré dans l'introduction, l'opposition est un des facteurs déterminant des positionnements énonciatifs spécifiques, au moins dans la manière dont sont désignés les points de vue auxquels se réfère l'auteur ; elle est indépendante cependant du statut de l'auteur, les tests statistiques entre ces deux variables montrant qu'il n'y a pas de corrélation : les articles où est annoncé un objectif d'opposition ne sont pas moins fréquents chez les doctorants, en dépit de cette spécificité de statut, et de son influence sur le positionnement énonciatif du scripteur dans son texte.

Si les pratiques des doctorants et des autres chercheurs sont similaires de ce point de vue, il faut sans doute y voir un argument à la prise en compte de certains effets de courants et d'écoles : le poids de la formation du chercheur et du domaine dont relèvent ses travaux pourraient induire ce type de pratique consistant à donner à son article un enjeu d'opposition. Nous avons vu également vu avec l'étude des styles au chapitre 6 qu'il pourrait s'agir d'un fait spécifique à certains auteurs, et relativement constant dans leurs textes : certains styles, en particulier celui des auteurs A4 et A7, doivent leur spécificité à des aspects énonciatifs possiblement indiciels d'une dimension polémique dans l'argumentation (avec les marques de personne et de non-personne, les marques d'emphase, les marques polyphoniques comme le conditionnel, la négation, certains connecteurs d'opposition ou de concession, etc.).

Au final, il nous faut revenir aussi sur la différence disciplinaire que nous avons momentanément laissée de côté. Nous nous sommes centrés ici sur les articles de SCL, comme dans l'étude des styles. Nous avons vu à propos des modes de désignation des points de vue que les articles de LET recouraient à un nombre moindre de références à d'autres travaux du champ ; ces références se satisfont globalement mieux qu'en SCL d'une approche sous l'angle de références dénominatives, à des auteurs ou à des textes, bien que les renvois à des noms de courant, et plus encore aux désignations généralisantes ne soient pas exclus.

Nous avons montré aussi à travers les différents aspects de la caractérisation énonciative du genre qu'un certain nombre d'aspects du texte pourraient être mis au compte d'un enjeu plus polémique. En particulier, la forte présence à la fois de marques personnelles et de tours impersonnels semble précisément caractéristique de textes fortement argumentatifs, voire tendanciellement polémiques. La polémique peut concerner l'objet étudié (nous en avons parlé dans le chapitre 4, avec le rapport polémique de l'auteur aux réformes de l'orthographe, ou en LET, au texte étudié) ; elle peut concerner également le champ de recherches lui-même, et ce serait davantage le cas en SCL qu'en LET. Par contraste avec les LET, les SCL se caractériseraient comme un champ où la discussion des approches, voire la remise en question des paradigmes revêt une place importante dans l'écriture de l'article.

Cette dimension polémique de l'article en SCL représente une des conclusions majeures du projet KIAP², qui vise rappelons-le à comparer l'article de recherche dans trois langues (anglais, français, norvégien), et trois disciplines (économie, linguistique, médecine), et conclut que l'identité disciplinaire de la linguistique, par opposition aux autres champs se situe précisément dans ce goût marqué pour la polémique. Les analyses de T. Kuhn (1983) sur le champ scientifique pourraient peut-être permettre de rendre raison d'une telle

¹ Mann Whitney, $U=146,5$, $p=0,0462$ (Cf. Annexes 5.2.).

² Cf. à la fin du chapitre 1 pour une présentation générale du projet.

caractéristique : les modes de renvoi aux sources ne remettent pas en question la cumulativité du savoir comme nous l'avons souligné, mais la dimension polémique des textes résulterait, et servirait, un champ de production des savoirs caractérisé par une instabilité telle que la controverse y serait centrale : la discussion tendanciellement polémique refléterait l'absence de consensus sur les paradigmes, et sur la définition du champ lui-même (par exemple, champ scientifique ou champ savant ?) et de ce qui doit être pris comme objet.

Plus largement, cette question de la polémique pose celle du positionnement du scripteur dans son texte. Avec l'analyse du renvoi aux sources, nous avons vu comment l'auteur se constitue comme un chercheur par une gestion scientifique des références. Il est aussi celui qui affirme sa singularité (cf. par exemple les formes d'auto-désignation) et se présente dans une position de surplomb dans le champ avec certains modes de désignation des points de vue qui y circulent. C'est alors à la manière dont se trouve légitimée, dans l'article, la recherche qui y est présentée qu'il nous faut nous intéresser, pour mieux cerner la question du positionnement du scripteur sous l'angle du positionnement de son approche et de son objet. Comment le point de vue du chercheur-auteur se trouve-t-il légitimé en tant que point de vue scientifique ? Comment, en même temps, se présente-t-il comme un point de vue singulier, nouveau, voire différent ?

La dernière étape de notre analyse énonciative du genre de l'article nous emmène donc du côté d'une rhétorique de la science, qui n'analyse pas en termes d'intéressement, ni *a fortiori* en termes d'intentionnalité, la question de l'inscription et du positionnement de l'auteur dans le champ de recherches, mais qui étudie comment le genre requiert de définir par rapport à ce champ l'objet que l'on se donne, et le point de vue que l'on défend. La question de la légitimation de la recherche, abordée en termes de caractéristiques textuelles qui servent une telle légitimation, nous conduira à revenir sur les références à l'existant, de même que sur la le marquage d'une individualité du chercheur constitutive de l'auctorialité, et à affiner encore les différences disciplinaires jusqu'ici analysées.

7.5. Des objectifs et une axiologie au service de la légitimation de la recherche

Poursuivant notre analyse énonciative de l'article, et la description de la subjectivité et de la polyphonie caractéristiques du genre, nous proposons de nous centrer à présent sur la manière dont le point de vue de l'auteur et son approche se trouvent légitimés dans l'article. Deux entrées textuelles nous semblent pouvoir servir l'analyse d'un tel objet.

D'une part, nous nous centrerons sur les introductions, pour voir en quels termes est introduit l'objet de l'article et son objectif. Nous prenons appui sur l'analyse des introductions d'articles de J. Swales (1990), qui propose d'y voir la création d'un espace de recherche. Sans reprendre telle quelle la structure rhétorique qu'il modélise, nous en retrouvons certaines dimensions en essayant de décrire les diverses possibilités par lesquelles le scripteur positionne son approche et en montre l'intérêt.

D'autre part, nous traiterons du lexique axiologique, en reprenant cette dénomination de la catégorisation des subjectivèmes proposée par C. Kerbrat-Orecchioni (1980) : ce lexique évaluatif impliquant un jugement de valeur fait état de paradigmes récurrents qui servent à mettre en évidence l'intérêt d'une approche ou ses limites.

L'étude prend appui sur le sous-corpus déjà utilisé, constitué de 20 articles par discipline ; au besoin, nous recourons également dans notre analyse à des extraits d'articles relevant du corpus intégral.

Nous retrouvons, avec les introductions en particulier, un certain nombre de phénomènes déjà abordés (les références à l'existant notamment). L'intérêt de ces deux entrées, l'annonce des objectifs dans l'introduction et le lexique axiologique est qu'elles

permettent de mieux cerner comment l'auteur se positionne dans son texte et comment l'objet y est défini, de manière commune ou spécifique à chacune des deux disciplines considérées ; en offrant une prise sur la construction dans le texte d'une légitimité de la recherche, elles fournissent des pistes pour une analyse du rôle des aspects énonciatifs dans la dimension argumentative des articles.

7.5.1. La définition de l'objet et de l'objectif de l'étude dans l'introduction

Nous avons en première approche considéré qu'il serait sans doute possible d'identifier des types d'articles sur la base des objectifs tels qu'ils sont annoncés dans l'introduction et repris dans la conclusion – ou du moins dans une partie introductive et/ou conclusive quand il n'est pas possible de distinguer une introduction ou une conclusion à proprement parler du reste de l'article. En particulier, il semble possible de voir si l'article dit comporter une dimension applicative dans ses enjeux (par exemple, dans le domaine de l'enseignement du français ou d'autres langues), de mettre en évidence son matériau (recueil de données sur le terrain, corpus, un texte littéraire ou le traitement d'une question transversale à plusieurs, etc.), et de montrer comment l'article positionne son objet (en termes d'opposition par rapport à d'autres approches, en en soulignant le caractère inédit, etc.). Au final, nous pouvons conclure à la difficulté qu'il y a à établir une typologie d'articles à partir d'une telle entrée ; la confrontation d'articles de LET et de SCL complique en effet l'analyse, et nous avons donc décidé de nous intéresser à la question des objectifs de l'étude mais de manière plus restrictive, en tant qu'entrée susceptible d'intéresser la question de la légitimation dans l'article de la recherche qui y est présentée.

Nous proposons en effet de nous centrer sur la « création d'un espace de recherche » dans les introductions, c'est-à-dire à la définition d'un objet d'étude et à l'annonce des objectifs de l'étude. A l'instar de l'analyse rhétorique des introductions proposées par J. Swales (1990), il s'agit de mettre au jour des schémas récurrents, relatifs à la manière dont l'auteur positionne sa recherche. L'analyse de J. Swales (1990) consiste à voir dans les introductions trois mouvements (« moves »), qui se configurent eux-mêmes en différentes étapes (« step »), obligatoires ou alternatives. Le premier mouvement consiste à établir un territoire (annoncer le caractère central de l'étude, partir de la thématique plus large à laquelle elle est indexée, ou d'appuis aux travaux antérieurs), le second à établir une niche (annoncer un enjeu d'opposition, indiquer une lacune, soulever une question ou proposer de poursuivre une tradition), et le troisième à occuper cette niche (souligner les objectifs de l'étude ou annoncer la recherche présentée, annoncer les principaux résultats et indiquer le plan de l'article).

Malgré les modifications apportées par rapport aux premières propositions de 1981, le modèle CARS (« Create a research space ») des introductions d'article (1990 :141) ne s'est avéré que peu opératoire dans notre corpus, bien que nombre de dimensions pointées dans ce modèle s'avèrent en effet récurrentes. J. Swales note que la distinction entre les deux premiers mouvements (« establishing a territory » et « establishing a niche ») a été jugée difficile, et il en va parfois de même aussi avec le troisième (« occupying the niche »). Plus largement, il n'est pas aisé d'analyser le schéma d'ensemble des introductions, tant celui-ci est variable ; la diversité au seul niveau de leur longueur en atteste : dans certains cas, elles visent nettement une définition de l'objet, alors que dans d'autres, elles consistent plutôt en un résumé, la circonscription de l'objet étant réservée à une première partie de l'article.

Plutôt que de mettre au jour un schéma rhétorique de l'introduction dans l'article, nous proposons donc de décrire les diverses possibilités offertes au scripteur pour définir son objet et/ou les objectifs de son étude, dans les parties introductives. Nous montrons ainsi comment

les introductions visent la « création d'une niche », au sens où elles servent le positionnement de l'étude par rapport à l'existant, et permettent en cela de mettre en relief son intérêt.

L'examen des introductions suggère certaines récurrences, mais aussi certains contrastes dans la manière dont sont introduits l'objet de l'article et ses objectifs. Le premier aspect dont nous traiterons concerne la présence dans l'introduction de références aux travaux du champ de recherches. Le second posera la question de savoir si la définition de l'objet et des objectifs fait intervenir un enjeu de démarcation (il s'agit de se démarquer de l'existant), ou se satisfait au contraire d'une centration sur le seul objet et son intérêt intrinsèque.

Se référer à l'existant

Voit-on se manifester dans l'introduction des références aux travaux du champ de recherche, ou, plus largement même, tout point de vue se présentant comme un point de vue attesté, à l'exclusion du point de vue des auteurs étudiés en LET, et des sujets étudiés en SCL ? Cette question est diversement résolue selon les articles, et manifeste une différence nette entre les deux disciplines : en SCL, faire référence aux travaux du champ de recherches, ou, du moins, à un point de vue attesté sur l'objet est à ce point fréquent qu'on peut y voir une norme du genre : 17 articles sur 20 sont concernés. En LET, on peut y voir une possibilité, qui se trouve exploitée dans un quart des articles (5 sur les 20).

En SCL, l'introduction demanderait donc plus qu'en LET à se positionner par rapport à l'existant, du moins quand tel n'est pas l'objet de la première partie de l'article. C'est ce que l'on observe par exemple dans cet extrait tiré d'un article de SCL, et qui en constitue le second paragraphe, après un premier paragraphe dédié à ce que J. Swales (1990 :141) appelle une « topic generalization », où l'auteur annonce qu'il va traiter de la citation dans les livres d'histoire (« *Les livres d'histoire sont truffés de citations : parce que le genre, à défaut de l'exiger, y incite ; parce que la citation...* ») :

L'intention ne sera pas de dresser une typologie des pratiques citationnelles dans l'écriture historique ni de remonter aux origines de l'érudition moderne et l'apparition des codes qui régissent l'usage des *sources*. **Si depuis un peu plus d'une vingtaine d'années, on a timidement commencé à s'intéresser à la poétique - au sens de mise en texte - du discours historique (de Certeau, 1975; Rancière, 1992; Carrard, 1998...), si on s'est penché sur la note de bas de page comme lieu de l'érudition (Grafton, 1998), il n'existe pas d'étude de la citation en histoire, en langue française du moins. Certes, des linguistes et des historiens se sont interrogés sur des points de rencontre possibles entre histoire et linguistique structurale et en ont fait la démonstration par des expériences de lexicométrie (Robin, 1973); ils se sont associés dès le début des années soixante-dix pour faire de l'analyse de discours sur des documents qui font partie de l'arsenal heuristique traditionnel de l'historien (Robin et alii, 1994); mais le discours des historiens en tant que tel n'a pas été soumis à pareille opération.** [SCL036]

Nous retrouvons avec cet extrait ce que nous avons dit du renvoi aux sources dans l'article de recherche : le paragraphe mobilise un grand nombre de références, et manifeste une quête d'exhaustivité dans cet état des lieux des travaux existants. Le recours au *on* (« *on a commencé à s'intéresser à...* », « *on s'est penché sur...* ») et à une liste de documents se clôturant par des points de suspension (« *(de Certeau, 1975; Rancière, 1992; Carrard, 1998...)* ») permet cependant de se défausser d'une tentation d'exhaustivité, de même que la restriction apportée par « en langue française du moins ». L'étude est introduite en référence à un état des lieux relativement détaillé de l'existant, et c'est le constat de l'absence d'étude sur la citation en histoire qui justifie la nécessité de l'étude.

Il convient alors d'examiner les références concernées, pour voir par rapport à quel existant les articles situent leur objet. Comme on le voit dans l'extrait ci-dessus, il peut s'agir de références à des travaux de recherche : références à un document, à une liste de documents,

à un domaine ou aux chercheurs de ce domaine (« *des linguistes et des historiens* »). Cependant, il est intéressant de constater qu'en LET, il n'est pas possible de tracer une frontière entre les références aux chercheurs du domaine, et les références aux interprètes d'une œuvre, quel que soit le contexte socio-historique de l'interprétation. C'est ce que l'on observe dans cette fin d'introduction d'un article de LET :

Enfin traduite en français, et publiée dans une collection de poche, l'œuvre romanesque de R.K. Narayan **mérite sans doute d'être réexaminée**, et rapportée à **sa réception tronquée en Occident** : quel est le réalisme de R.K. Narayan ? Dans quelle mesure celui-ci a-t-il été construit par **un regard occidental sélectif** ? Se penchant sur la dimension fantastique de l'œuvre de R.K. Narayan, **globalement éludée, ou dévaluée en Occident, mais qui apparaît dès les premiers romans** – sur lesquels reposera essentiellement l'analyse –, **le présent article a pour ambition de s'interroger sur la nature exacte d'une écriture romanesque indienne, en même temps que sur les raisons de son évaluation occidentale ambiguë**. [LET088]

L'auteur ici situe son approche en référence au champ de la critique littéraire occidentale. L'œuvre demande selon lui à être « *réexaminée* » en raison de sa réception en Occident dont il montre le caractère lacunaire (« *réception tronquée* », « *regard sélectif* », « *dimension globalement éludée* ») et problématique (l'auteur parle d'une « *évaluation occidentale ambiguë* » où la dimension du fantastique est « *dévaluée* »). La valeur distinctive ainsi conférée à l'approche présentée dans l'article va de pair avec la figure d'un auteur de LET qui se présente comme un lecteur du texte étudié, mais comme un lecteur singulier : non seulement il perçoit que la dimension fantastique « *apparaît dès les premiers romans* », mais il est suffisamment décentré pour remettre en question la réception occidentale dans son ensemble, et même en analyser les raisons.

On retrouve cette idée dans un article consacré au caractère hispanique de Don Juan : l'introduction s'inaugure sur l'idée que c'est le valet Ciutti qui suppose, dans le texte, que Don Juan est espagnol, et se poursuit en ces termes :

Sa supposition devient une évidence lors de la pièce et tout au long des différentes versions que la littérature universelle offre sur le mythe donjuanesque. Or il convient de signaler que la croyance du valet n'est fondée que sur l'attitude d'un homme qui ne pouvait être qu'un Espagnol. Depuis sa parution dans les romances espagnols chantés au Moyen Âge dans les royaumes de Galice, León et Castille, « le galant qui allait à la messe » a revêtu différentes modalités populaires jusqu'à sa cristallisation dans la pièce L'Abuseur de Séville attribuée à Tirso de Molina ; ces métamorphoses du mythe l'ont montré sous des jours nouveaux depuis les toutes premières pièces espagnoles jusqu'aux œuvres les plus récentes de tous pays. Il est étonnant que la nationalité n'ait changé dans presque aucun cas ; **force est de reconnaître que Don Juan est resté dans l'imaginaire littéraire et social comme un personnage espagnol**. Il convient d'évoquer les principales raisons qui exigent le caractère hispanique du plus grand séducteur de toute la terre. [LET103]

Cet extrait d'introduction est tout à fait similaire au précédent. Les références à l'existant ici convoquées sont « *la littérature universelle* » et « *l'imaginaire littéraire et social* » ; mieux encore que dans le premier extrait, où il était question de la réception occidentale de l'œuvre étudiée, on voit comment la prise en compte de l'interprétation de l'œuvre implique une continuité entre le champ de recherches et le champ social : les différentes versions du mythe de Don Juan en sont autant d'interprétations, de même que l'imaginaire littéraire et social qui le caractérise.

En SCL, les références à l'existant observables dans les introductions font également état d'une certaine hétérogénéité. Si les introductions d'articles de SCL se caractérisent par contraste avec celles de LET par le poids des références à d'autres travaux, on note en particulier que, de manière spécifique aux SCL, un certain nombre d'introductions

comportent des renvois à des travaux antérieurs de l'auteur lui-même. Ces formes d'auto-citation sont le fait de 5 des 20 articles du sous-corpus, et se justifient par l'économie textuelle :

Il n'est pas possible ici de décrire dans le détail le positionnement théorique sous-jacent à ce travail ; on en trouvera des éléments plus étendus ailleurs (Robillard, 2001). [SCL007]

Elles permettent d'inscrire l'objet de l'article dans un projet de recherche qui dépasse le texte concerné, et se présente comme un projet continué, qu'il s'agisse d'un projet personnel ou d'un projet d'équipe. C'est en ce sens que l'on peut interpréter l'article SCL099, bien que celui-ci se présente comme un cas tout à fait spécifique ; nous avons déjà eu l'occasion d'évoquer son auteur à la fin de la partie consacrée à l'étude du style, en raison du jeu avec les normes du genre qui caractérise ses textes. Dans l'article SCL099, l'objet d'étude consiste pour l'auteur à revenir sur sa propre approche de la problématique dénomminative, comme le montrent les extraits de l'introduction cités ci-dessous :

Cet article n'a pas pour ambition de régler le problème de la dénomination. Il s'agit — il n'est pas inutile de le souligner d'emblée pour éviter toute interprétation déplacée — **d'un simple retour sur l'approche que nous avons proposée au début des années 1980**. Loin de nous la volonté de nous approprier la problématique dénomminative et d'essayer de passer pour Monsieur Dénomination. Nous n'ignorons pas, en effet, que les problèmes posés par la dénomination vont bien au-delà de ceux que nous aborderons ici, qui se circonscrivent volontairement à notre parcours personnel. [...]

Ce retour se justifie pour au moins trois raisons [...]

Il n'est donc pas inutile de reprendre le problème de la dénomination à ses sources en rappelant dans une première partie les principales positions que nous avons défendues dans Kleiber (1981 et 1984). La deuxième partie s'attachera à montrer ce qui est resté et ce qui a bougé. [...] [SCL099]

On voit comment cet objet spécifique requiert de la part de l'auteur d'être expliqué, et de devancer d'éventuelles réactions du lectorat : l'auteur signale que la problématique dénomminative ne se réduit pas à ses propres travaux. Pourtant, c'est la nécessité de « *reprendre le problème de la dénomination à ses sources* » qui lui sert d'argument pour revenir sur ses travaux, de sorte que l'introduction concourt à faire de l'auteur « *Monsieur Dénomination* », alors qu'il se défie d'une intention de ce type (« *loin de nous la volonté...* »). Une autre caractéristique des introductions de SCL, également attestée en LET, se situe dans les références à la tradition scientifique prise dans son ensemble. Il est ainsi question d'« *un usage limite de la citation qui rompt les distinctions classiques entre discours cité et discours citant* » (SCL026), ou encore, de « *repenser la situation triangulaire classiquement attribuée à la vulgarisation scientifique* » (SCL090). De manière proche, mais spécifique aux SCL, nombre de références concernent la grammaire traditionnelle ou encore l'approche dictionnaire, qui tendent à être indexée à cette même tradition et vis-à-vis desquelles les articles entendent se démarquer, comme nous le verrons. Dans les références à ces « *distinctions classiques* » ou à ces « *conceptions traditionnelles* », il semble qu'on puisse parler d'une dilution de la frontière entre le champ disciplinaire et une forme de doxa. C'est également cette continuité entre certains points de vue du champ disciplinaire et des points de vue doxiques, relevant du champ social qui se manifeste dans l'extrait suivant, à propos des représentations sociales des pratiques langagières des « *adolescents urbains* » :

contrairement à une représentation assez largement partagée qui a tendance à en faire des “ handicapés linguistiques ” irrémédiablement voués à un “ enfermement pseudo-identitaire ” (Bentolila, 1998), des locuteurs détenteurs de répertoires verbaux comprenant des éléments de parlars urbains et/ou des compétences réceptrices dans une langue familiale – autre que le français –, sont capables de [...] [SCL009]

parce qu'elle est décrite comme « *assez largement partagée* », mais également en référence à un auteur cité, la représentation à laquelle l'auteur entend s'opposer est présentée comme le fait d'un point de vue doxique, endossé dans le champ social comme dans le champ de recherches.

Cette analyse des références à l'existant dans les introductions montre que c'est une possibilité attestée en LET, mais qu'elle est presque unanimement exploitée en SCL, où il peut notamment s'agir pour l'auteur de définir son objet en référence à ses propres travaux ; par ailleurs, l'existant en question ne se restreint pas à des travaux de recherches, mais se caractérise parfois par une dilution de la frontière entre le champ social et le champ disciplinaire, avec, en LET, un positionnement de l'objet de l'article par rapport à l'ensemble des interprétations d'une œuvre, et en SCL, un positionnement par rapport à un point de vue jugé traditionnel voire doxique.

Les extraits que nous avons cités suggèrent que le positionnement par rapport à l'existant est dans nombre de cas un positionnement d'ordre adversatif, et c'est ce sur quoi nous allons revenir : il s'agit plus largement de voir à partir de quels procédés la définition de l'objet et des objectifs de l'article sert la mise en relief de l'intérêt de la recherche présentée.

Un enjeu de démarcation

L'analyse des introductions d'article montre que la définition de l'objet et des objectifs de l'article peut faire intervenir un enjeu de démarcation (il s'agit de se démarquer de l'existant), ou se satisfait au contraire d'une centration sur le seul objet et son intérêt intrinsèque.

Une première possibilité pour définir l'objet de l'article consiste en effet à montrer en quoi l'étude se démarque d'autres points de vue existants sur cet objet. C'est ce que nous désignons comme un « enjeu de démarcation » dans l'introduction, et celui-ci est nettement plus représenté en SCL où il concerne 13 des 20 articles de SCL, alors que ce n'est le fait que de 4 des 20 articles de LET.

Il semble possible de distinguer deux procédés caractéristiques de cet enjeu de démarcation : d'une part le fait de signaler des lacunes dans l'existant, qui justifient qu'on s'intéresse à l'objet, et d'autre part le fait de proposer une approche qui s'oppose aux approches antérieures, second aspect que nous avons déjà évoqué à partir des modes de désignation des références dans la partie précédente.

Le fait de signaler des lacunes est un procédé signalé par J. Swales comme une des possibilités permettant d'établir une niche dans l'introduction. Ainsi l'article LET019, qui porte sur l'édition des *Essais* établie par S. Goulart, et sur la censure qui s'y exerce commence par situer cette édition « *genevoise* » et « *calviniste* » et note qu'« *on peut s'étonner* » qu'elle « *ait été si peu étudiée* » ; après avoir signalé que certains critiques ont donné une description générale des chapitres supprimés et de la censure, l'auteur annonce que « *ce sont quelques aspects de la censure de Goulart qui n'ont pas été relevés auparavant* » qu'il se propose d'étudier. De manière similaire, l'article SCL030 s'ouvre sur le constat que « *l'analyse syntaxique et fonctionnelle des séquences de discours rapporté n'a pas vraiment retenu l'attention des grammaires traditionnelles* ».

Le procédé consistant à signaler des lacunes dans l'existant peut aller jusqu'à affirmer que rien n'a été dit de l'objet¹, et sert à mettre l'accent sur le caractère inédit de l'étude présentée :

Personne ne songerait à reconnaître dans *Le Dit de Poissy* une étape majeure dans la carrière littéraire de Christine de Pizan. **Pourtant ce dit est loin d'être sans attraits**, dans les parties descriptives notamment, dont la couleur, le rythme et le ton, plus personnels que d'ordinaire, débordent le cadre des conventions du genre. **Cependant c'est surtout la construction surprenante, voire bizarre de ce poème qui le distingue et peut susciter l'intérêt** : car cette particularité fournit l'occasion de proposer quelques réflexions qui permettront peut-être de mieux comprendre **un moment important** de l'évolution littéraire et morale de Christine. [LET029]

L'objet ici non seulement n'a jamais été traité, mais n'a pas même été « *reconn[u]* » comme tel. En conformité avec ce que nous avons dit de la figure de l'auteur dans les articles de LET, l'auteur se présente comme un lecteur du matériau étudié, et comme un lecteur unique et privilégié, capable de décoder ce que « *personne* » d'autre n'a pensé à percevoir dans l'œuvre, et qui en fait précisément toute la richesse, « *l'importan[ce]* » et « *l'intérêt* ». Partant de là, c'est l'importance et l'intérêt de l'étude elle-même qui est ainsi mise en relief.

L'absence d'étude peut aller de pair avec l'idée que l'existant se réduit à une appréhension doxique de l'objet traité. C'est le cas dans l'introduction d'un article sur l'interprétation des pictogrammes, en SCL :

Depuis quelques années, on observe une augmentation appréciable du nombre des pictogrammes et de leur emploi. Nous entendons par « pictogramme » un message graphique non verbal, à but informatif ou injonctif, de nature globale (c'est-à-dire qu'un message entier est exprimé par un pictogramme) et au tracé majoritairement iconique. **De manière surprenante, si l'usage des pictogrammes se développe rapidement, il n'existe pas à notre connaissance d'étude portant sur la perception et l'interprétation de ces signes. Les travaux existants laissent tous de côté celui qui doit les comprendre: le « lecteur ».** C'est donc autour de ce lecteur et de la manière dont les pictogrammes sont interprétés que s'organise notre recherche.

Il semble que la raison pour laquelle la réception des pictogrammes n'est jamais interrogée soit qu'il est communément admis que ces signes, parce qu'ils sont non verbaux, sont des signes transparents, univoques et universellement compréhensibles. Ils seraient donc un moyen idéal de communiquer facilement et rapidement un message simple par écrit. Ils constitueraient également un écrit « alternatif » pratique, offrant le moyen d'une lecture minimale à l'usage des non-francophones et des « illettrés ».

Ce sont ces hypothèses implicites qui servent de justification à la création et à l'utilisation croissante des pictogrammes. Elles n'ont, à notre connaissance, jamais été soumises à vérification et rien n'indique que leur validité soit aussi évidente qu'il y paraît. Il en va pourtant de l'efficacité d'un mode de communication

Les « *travaux existants* » ne sont ici évoqués qu'au moyen de cette désignation généralisante. Les lacunes sur l'objet traité sont expliquées par l'auteur en référence au sens commun (« *il est communément admis que...* »). L'objectif est de se livrer à une étude inédite, qui permette de dépasser des « *hypothèses implicites* », qui ne seraient pas selon l'auteur « *aussi évidentes qu'il y paraît* ». Le positionnement est double : la niche occupée par l'auteur vient combler une lacune, puisqu'il propose de s'interroger sur ce qui n'est pas interrogé ; sur la base de l'opposition au sens commun, l'étude, inédite, s'octroie en outre des attributs de scientificité. Elle se présente comme utile à la fois au champ de recherches (tenir compte de signes dont « *on observe une augmentation appréciable du nombre et de leur emploi* »), et au champ social des usagers de ce mode de communication dont il s'agit de questionner « *l'efficacité* ».

Si le fait de signaler des lacunes peut impliquer comme nous venons de le voir une forme d'opposition à l'existant, l'opposition peut également représenter l'objet central de l'étude, tel qu'il est défini dans l'introduction. C'est le cas lorsque l'auteur propose « *une*

¹ Nous l'avons vu *supra* à propos de l'introduction de l'article SCL034, qui souligne l'absence d'études sur la citation en histoire, tout en détaillant les études sur le discours historique.

nouvelle approche », par exemple des relatives dans le cadre des recherches sur le discours rapporté (SCL021), ou dans cet article de LET :

Ceci sera l'occasion d'une **nouvelle approche** de l'énigme comme forme discursive, donc espace d'hétérogénéité pouvant permettre de **renouveler les approches des années 1970 de Todorov et Jolles** sur cette « forme simple ». [LET028]

Avec la nouveauté, c'est donc à la fois l'opposition par rapport aux approches antérieures et le caractère inédit de l'étude qui servent sa définition. L'enjeu d'opposition peut être plus explicite encore :

Je voudrais ici défendre la thèse selon laquelle la figure de rhétorique — et je pense spécialement au trope — joue un rôle argumentatif. **En formulant cette thèse, je prends à contre-pied l'idée, classique**, d'une distinction entre la rhétorique des figures et la rhétorique de l'argumentation. C'est évidemment par un rappel de cette distinction qu'il me faudra commencer, avant de fournir de la figure **une description nouvelle** — essentiellement pragmatique —, ce qui nous permettra en troisième lieu d'aborder le problème de son rôle argumentatif. [SCL096]

Dans les formes d'opposition, des travaux ou des auteurs peuvent être cités, mais généralement sous la forme de liste, l'objectif étant donc de catégoriser un ensemble d'approches : (« *De Quintilien à Fontanier au Groupe μ , de Dumarsais à Black, la plupart des auteurs associent les figures de sens à l'idée d'un sens figuré, [...] Dans tous les cas cités,...* » SCL094). A la base de l'opposition, il y a donc le plus souvent un ensemble large et relativement indéfini (« *on* »¹, « *l'idée classique* », « *le point de vue traditionnel* » « *généralement analysée* ») qui peut renvoyer à la tradition scientifique, ou, en LET, à la réception et l'interprétation d'une œuvre.

Le cas d'opposition le plus fréquent consiste à mettre en face à face l'approche proposée et un ensemble extrêmement large où sont conjointes différentes approches ; un autre procédé consiste à présenter une opposition dans l'existant, et à définir son objet comme une tierce alternative. C'est ce que l'on observe dans les deux extraits suivants, tirés respectivement d'un article de LET et d'un article de SCL :

Dans sa réflexion ironique sur la coutume et ses emprises irrationnelles, Montaigne n'épargne même pas de valeurs hautement appréciées de son temps: l'héroïsme et la bravoure militaires. Dans un passage de l'essai "De l'expérience", il examine, avec une pointe de raillerie, la valeur militaire. **Les propos sont bien connus, car ils sont à l'origine d'un débat déjà ancien** entre ceux qui croient à leur sincérité et les autres qui y entrevoient une ironie. **Nous ne reprendrons pas ici la querelle dans le sens où ces éminents montaignistes l'avaient située: l'art militaire et la guerre. Nous pensons qu'il y a bien ironie, mais elle porte sur le courage militaire, ce qui est totalement différent.** [LET005]

Afin d'éviter les éternelles disputes qui opposent ainsi l'acception artificialiste de l'opinion publique par une sociologie du soupçon qui la dénonce comme une illusion mensongère utile à la reproduction de l'ordre établi, **et son acception réaliste** par une sociologie empirique qui lui attribue le statut d'une entité réelle et donc dûment quantifiable, **cet article se propose de partir de ce qu'elle est de toute manière.** [SCL067].

Ces deux exemples, où le débat existant est taxé de « *dispute* » ou de « *querelle* », montrent que l'alternative offerte par l'auteur se présente donc comme une tierce voie, capable de dépasser les limites de chacune des deux autres, mais également d'opposer au point même qui fait débat une proposition différente.

¹ Par exemple dans « l'influence de l'*Ovide moralisé* pèse moins lourd qu'on a bien voulu le croire et le dire », LET042.

L'enjeu de démarcation dans l'article peut s'appuyer sur la mise en évidence de lacunes dans l'existant, ou de difficultés qui justifient qu'on y oppose une autre ou une nouvelle approche. La désignation de l'existant est spécifique : l'enjeu de démarcation conduit à préférer au renvoi à un document des listes d'auteurs, des noms d'approches ou de domaines, voire des désignations plus indéfinies encore, et peut aller jusqu'à signaler que rien n'a été dit. Dans tous les cas, l'effet produit est une mise en relief de la valeur distinctive de l'article : parce que l'objet et les objectifs sont dits utiles à combler des lacunes ou parce qu'ils doivent concourir à une approche nouvelle et différente, l'enjeu de démarcation fait de l'étude un apport singulier, et la légitime ainsi par rapport au champ.

Ces procédés, relativement répandus dans les deux champs et plus fréquemment encore en SCL, ne sont cependant pas toujours utilisés dans les introductions. Que l'introduction comporte ou non des références à d'autres travaux, elle semble pouvoir définir et légitimer son apport en restant centrée sur le seul objet et son intérêt intrinsèque. Mais nous verrons en quoi cette centration ne va pas sans impliquer le champ de recherches auquel l'article entend contribuer.

Une centration sur l'objet et son intérêt intrinsèque ?

Les termes dans lesquels l'objet de l'article et ses objectifs sont définis peuvent ne pas rendre compte de lacunes ou d'un enjeu d'opposition par rapport au champ de recherches et plus largement à l'existant. L'article SCL086 illustre ce cas d'introductions, qui se réfère au champ de recherches et à d'autres travaux sans pour autant entendre s'en démarquer. Nous reproduisons ici l'introduction *in extenso*.

Trouble phonétique ou trouble phonologique ? Cette question demeure centrale dans les recherches de pathologie du langage, notamment en ce qui concerne la nature des erreurs de voisement dans la parole aphasique. Pour l'aphasie de Broca ou l'apraxia of speech, les patients ont tendance à remplacer des consonnes sonores par leur corrélat sourd (Alajouanine, 1939).

L'origine de ces transformations segmentales (Valdois & Nespoulous, 1994) **est incertaine** (Nespoulous, 1979) : surviennent-elles lors de la phase phonologique (pré-motrice) ou lors de la phase phonétique (motrice) de la parole (cf. Macneilage, 1982) ? Autrement dit : s'agit-il d'une substitution phonémique ou d'une distorsion articulatoire ? En d'autres termes encore, est-ce une perturbation de sélection ou de réalisation, de sonorité ou de voisement ?

Afin de cerner des éléments d'explication, différents indices du trait de sonorité ont été examinés. Dans ce contexte, une distinction entre quatre types de déviations de voisement a été établie. Cf. plus bas pour les détails et (Janus, 2001) :

- assourdissement,
- sonorisation,
- dévoisement,
- voisement.

Le comportement du voisement est observé chez des patients et des locuteurs témoins. Dans l'ensemble, **la présente étude essaie de trouver une réponse à la question relative à l'origine des déficits chez des aphasiques de Broca. À la lumière d'analyses acoustiques, il est tenté de déterminer si c'est la sonorité ou si c'est plutôt le voisement qui est perturbé.** [SCL086]

En l'absence d'enjeu de démarcation, l'introduction n'en signale pas moins l'intérêt de l'objet d'étude dans le champ : la recherche présentée doit permettre de répondre à la question jugée « *centrale* » de la nature du trouble dans l'aphasie de Broca. L'auteur ne dit donc pas avec

cette étude se démarquer de l'existant, mais sa problématique est cependant définie en référence aux travaux menés et aux questions qu'ils posent. C'est ce positionnement par rapport à l'existant qui fait la légitimité de l'étude dans le champ¹.

De fait, y compris quand l'enjeu de démarcation n'est pas au centre de la définition de l'objet et des objectifs de l'étude, la singularité de l'apport de l'article se trouve mise en avant. Le fait que « *l'origine* [des erreurs de voisement dans la parole aphasique soit] *incertaine* » permet en contrepartie de justifier la nécessité de l'étude.

De même, dans un article de doctorant (SCL078), où l'auteur propose, à partir de références à des travaux, de s'interroger sur le troisième ton en chinois standard, l'objectif est apparemment non démarcatif (« *Le but de cette étude expérimentale est d'essayer de répondre à ces différentes questions* »), mais entre en écho avec un des constats que fait l'auteur en amont (« *Les études théoriques n'ont pas expliqué le processus de ce changement* ») : une opposition se manifeste entre les deux désignations « *étude expérimentale* » et « *étude théorique* ». L'expérimentation que propose l'auteur se justifie donc en termes de démarcation par rapport à l'existant, sans pour autant que soit commentée la différence entre les deux types d'étude ni la nouveauté de la démarche par rapport aux recherches menées jusque-là dans le champ.

En LET, l'absence d'enjeu de démarcation à proprement parler est plus fréquente encore ; l'intérêt de l'objet est alors défini par rapport à son caractère central dans l'œuvre étudiée et pour sa compréhension.

L'extrait suivant en montre un exemple :

La littérature et la philosophie du XVIII^e siècle sont hantées par **l'appel de l'inconnu**: cet esprit de conquête, cette confiance dans les pouvoirs de la raison pour affronter l'inconnu, l'irrationnel, le monde brut, et lui donner forme sont inséparables des Lumières. **L'audace tranquille de la raison** éclate dans *L'Esprit des lois*, dans l'*Essai sur les mœurs*, dans *l'Encyclopédie*, dans *Le Contrat social*. Je la vois à l'œuvre encore, mais comme dans un épilogue des Lumières, chez Casanova ou chez Potocki. **Mais elle n'apparaît nulle part aussi bien que dans *l'Histoire des voyages***, car ici, la raison défie l'inconnu, le désordre de l'univers: le moi impose sa marque au monde. C'est cette histoire que raconte Prévost. [LET107]

Les marques subjectives et dialogiques se multiplient ici (« *mais* », négation, « *nulle part aussi bien* ») et ont pour effet la mise en relief de l'importance de la question des aventures (définies par l'auteur comme « *l'appel de l'inconnu* » et « *l'audace de la raison* ») dans *l'Histoire des voyages* de Prévost.

Il s'agit donc de mettre l'accent sur l'importance de l'objet de l'étude présentée dans l'article, et d'en convaincre le lecteur comme le montre l'extrait d'introduction ci-dessous,

L'art balzacien de construire des histoires lie ainsi étroitement la question de la forme et celle de l'objet du récit, **si bien que la question « que raconte cette histoire ? » pourrait servir de fil rouge à qui voudrait approcher les procédures de narration balzaciennes.** [LET024]

où l'auteur propose une grille de lecture présentée d'une part comme intrinsèquement essentielle à l'objet (en tant que « *fil rouge* » des procédures de narration balzaciennes), et d'autre part comme utile à tout analyste de l'œuvre.

¹ Dans une perspective didactique, E. Nonnon (2002), étudiant la problématisation dans l'écrit de recherche propose deux critères qui déterminent la recevabilité d'une problématique : son « acceptabilité », qui tient à la manière dont sont mis en relation différents niveaux de questionnement, et sa validité voire sa pertinence, qui tient à la mise en jeu dans la problématique de concepts fondamentaux dans le champ disciplinaire, et de questions ou de connaissances relatives à l'état actuel des recherches (2002 :47).

Avec le caractère central du questionnement, c'est aussi la complexité de l'objet qui peut servir la légitimité de la recherche. C'est ce que l'on observe dans ce début d'introduction d'un article de LET (nous ne reproduisons pas la citation de Pavel que comporte le texte) :

Le rêve du romancier réaliste, établir une correspondance quasiment parfaite entre le monde réel et le monde produit par la fiction, a-t-il un sens ? Le projet balzacien est formulé dans l'« Avant-propos » de 1842 comme le désir de parvenir à une complétude absolue, de telle sorte que le monde de la fiction puisse venir se juxtaposer point par point avec le monde réel. Thomas Pavel, dans *L'Univers de la fiction*, énonce cette utopie du roman balzacien comme effectivement accomplie : [...].

La dimension provocatrice des propositions de Pavel ne doit pas dissimuler l'ampleur des problèmes soulevés : les relations qui unissent les deux mondes de la réalité et de la fiction **sont pour Balzac un enjeu infiniment complexe, qui détermine des stratégies narratives propres à *La Comédie humaine***. [LET023]

La complexité de l'objet (« *l'ampleur des problèmes soulevés* ») est ici signalée en référence à un auteur reconnu du champ (Pavel), mais comme vraie également du point de vue de l'auteur étudié (« *sont pour Balzac un enjeu infiniment complexe, qui détermine des stratégies narratives propres à *La Comédie humaine** ») ; le caractère central de cet objet repose donc sur le fait que le questionnement soit partagé par le champ de recherches et qu'il soit aussi au cœur de la démarche littéraire de l'auteur, et partant, à la base de ses textes.

Finalement, que l'introduction comporte ou non des références à d'autres travaux, elle sert à souligner l'intérêt de la recherche présentée dans le champ où elle s'inscrit. Même quand la recherche n'entend pas se démarquer de l'existant, des procédés linguistiques propres au genre interviennent, qui ont pour effet de mettre en relief la pertinence de l'objet et des objectifs que se donne l'auteur.

Il convient alors de se questionner sur ce que recouvre cette idée d'intérêt : nous avons vu avec les introductions que l'intérêt de l'étude se définit notamment par son utilité à combler des lacunes ou à dépasser les obstacles des approches antérieures, et par le caractère central et la complexité de la question traitée. Ces constats nous invitent à approfondir encore cette idée d'intérêt, et à voir, linguistiquement, quelles valeurs sont convoquées pour le mettre en évidence.

Nous quittons les seules introductions pour une analyse menée sur l'ensemble du texte. L'examen des textes montre que la manière dont l'intérêt d'un point de vue ou d'une approche se présente dans l'article se trouve éclairée aussi par la manière dont sont signalées les limites d'un point de vue ou d'une approche. Qu'il s'agisse des points de vue de l'auteur ou d'autres points de vue attestés dans le champ, quels paradigmes sont mobilisés dans l'article pour mettre en évidence l'intérêt et les limites d'une approche ? Prenant comme point de départ le lexique axiologique utilisé pour évaluer les points de vue dans les textes, l'étude propose d'identifier les valeurs qui servent à caractériser l'intérêt et les limites d'un point de vue sur la base de son caractère heuristique, c'est-à-dire de la possibilité qu'il a de faire avancer le champ de recherches.

7.5.2. Les paradigmes axiologiques récurrents dans l'évaluation des points de vue et des approches

Nous avons vu en traitant de l'enjeu de démarcation dans les introductions que la définition de la recherche présentée dans l'article se base fréquemment sur une opposition entre ce que propose l'auteur et ce qui est proposé habituellement, c'est-à-dire d'abord antérieurement, de manière jugée traditionnelle, voire ancienne, mais aussi ordinairement, de manière générale

ou commune. La nouveauté par rapport à l'existant représente donc un premier critère déterminant l'intérêt d'une approche.

L'examen des moyens lexicaux utilisés pour évaluer l'intérêt ou les limites d'un point de vue, non plus seulement dans l'introduction mais dans l'ensemble de l'article, met en évidence des paradigmes axiologiques récurrents¹. Moins que les termes axiologiques eux-mêmes, ce sont les valeurs qu'ils impliquent qui se présentent comme relativement constantes et unifiées. Outre l'opposition entre la nouveauté et le caractère traditionnel ou ordinaire d'une approche, c'est sur la base de leur caractère heuristique que les points de vue sont évalués.

Nous proposons de rendre compte ici de ces paradigmes récurrents à partir de couples d'antonymes qui structurent la dimension argumentative de l'article en opposant les points de vue sur la base de leur caractère heuristique. À côté d'évaluations assez générales (« une excellente description », « une analyse judicieuse », le fait que tel ou telle approche permette de « bien décrire » l'objet), ces paradigmes axiologiques (ou ces valeurs) permettent de spécifier ce qui fait la légitimité de la recherche présentée, et ce sur quoi doit reposer selon les acteurs du champ scientifique la possibilité et la nécessité de faire avancer ce dernier.

Nous nous centrerons d'abord sur les valeurs communes aux deux disciplines, avant de voir en quoi les valeurs mobilisées sont marquées par des identités disciplinaires contrastées, y compris au sein d'un même champ.

Des valeurs propres à l'activité de recherche, et communes aux deux disciplines considérées

Parmi les axiologiques récurrents utilisés pour caractériser l'intérêt et les limites d'un point de vue sur la base de son caractère heuristique, une première relation d'antonymie se manifeste par la présence des champs sémantiques de l'obscurité et de la lumière ou la clarté. Le reproche est ainsi fait à certaines approches d'être « guett[ées] »² par « *la confusion* », par « *l'obscurité* », voire par le « *flou* ». À l'inverse, présenter un point de vue comme propre à offrir « *quelques éclaircissements* » ou « *quelques clarifications utiles* » permet d'en souligner la pertinence : l'intérêt d'une approche tient à ce qu'elle doit « *mettre au jour* » ou « *en lumière* » ce qui est parfois qualifié de « *zones d'ombres* », ou de « *points aveugles* », autrement dit permettre de « *faire le point* » « *pour y voir plus clair* ». Il est aussi question de « *reconnaître* » tel ou tel fait, ou encore, de ne pas « *masquer* » tel ou tel autre.

On s'appuie ensuite sur cette typologie, ainsi que sur la dimension dialogique de la signifiante de l'antonomase, pour **mettre en lumière des nuances sémantiques** relatives aux différentes réalisations antonomasiques. [SCL097]

En admettant plutôt que ce dernier et les autres déplacements en tête de phrase relèvent d'un même phénomène, on pourrait à la fois **faire l'économie d'une confusion inutile** et simplifier considérablement la théorie [SCL052]

En LET, l'opposition entre l'ombre et la lumière se manifeste également à propos du texte étudié et de son interprétation :

[la digression] se montre à la surface du texte, elle **s'offre au regard du lecteur. Mais celui-ci préfère détourner les yeux**, parce qu'elle lui semble inessentielle, ou parce qu'elle freine la narration de

¹ Cette étude du lexique axiologique a fait l'objet d'une publication (Rinck, 2006).

² Les segments guillemetés en italiques correspondent aux unités lexicales relevées dans le corpus d'articles des deux disciplines.

l'histoire. **Pour le lecteur, la digression est généralement ce qu'il ne lit pas**, ce qu'il lit vite et mal, ou encore ce qu'il ne retient pas. C'est un ailleurs textuel auquel on ne prête pas attention. Conscient de ce phénomène, Fielding autorise ironiquement le lecteur à sauter les « essais en forme de digression » [...] Les digressions constituent le seul intérêt de la lecture. **Elles ne sont plus les zones d'ombre du texte, ce sont elles qui projettent leur lumière sur le reste de l'œuvre**. Elles sont donc le véritable « endroit » du texte, ce qui justifie l'inversion faite par l'auteur. [LET025]

Comme nous l'avons déjà envisagé à propos du lexique transdisciplinaire, le lexique axiologique utilisé pour évaluer l'intérêt ou les limites d'un point de vue de chercheur vaut aussi, en LET, pour commenter le matériau lui-même. Cet extrait montre comment l'auteur se présente comme un lecteur privilégié, « conscient [d'un] *phénomène* » auquel le lecteur quelconque « *ne prête pas attention* », capable de mettre en lumière un fait essentiel de l'œuvre étudiée, plutôt que de l'ignorer au titre de zone d'ombre du texte.

Une autre opposition utilisée pour évaluer le potentiel heuristique d'un point de vue concerne l'angle de vue de l'analyste : les approches qui « *raisonnent dans un cadre étroit* », « *restreint* » ou « *restrictif* », à l'aide de « *notions réductrices* », sont appelées à être dépassées par des approches « *plus larges* », qui « *pren[nent] en considération* » « *différentes dimensions* » ou « *différents niveaux d'analyse* ».

C'est dans l'esprit des lois qu'apparaît cet autre usage du détail, qui **excède les considérations rhétoriques et discursives indéfiniment ramenées à l'équation trop de détails/pas de génie**, et invite à réfléchir sur la réticence pédagogique de Montesquieu [LET047]

Traiter, d'un point de vue stylistique, **ce type de phénomène comme un cas d'anacoluthie et l'interpréter comme la source d'un effet de rupture ou de surprise revient à raisonner dans le cadre étroit des relations syntaxiques et à ne pas prendre en compte le palier de traitement pertinent** [SCL048]

L'accent est mis toutefois sur les risques que comporte le rapprochement de champs, de notions ou de faits « *trop hétérogènes* » comme le « *caractère dispersé des outils d'analyse* ». Est donc montré comme heuristique le fait de « *tenir compte de différentes dimensions* », à la condition que ce soit « *de manière intégrative* », en optant pour une « *approche homogène* » ou « *unifiée* ».

Dans le même sens, s'affirme fréquemment une mise en garde à l'encontre d'une « *globalisation excessive* », accusée de « *masquer* » les problèmes et de « *dilue[r] les enjeux* » ; l'antonymie entre le large et l'étroit rejoint ainsi celle de la clarté et de l'obscurité : il est question d'éviter la confusion, en « *mainten[ant] nettement distincts les différents niveaux d'analyse* ».

Il s'agit, en d'autres termes, d'opposer le contrôle à la confusion, la rigueur aux amalgames. C'est ce que l'on voit par exemple dans cet extrait :

Classe définie de façon peu rigoureuse, huitième voire neuvième roue d'un carrosse grammatical branlant, l'interjection, « objet très flou » (Swiatkowska 1997) cristallise une série de problèmes de nature syntaxique, sémantique et pragmatique. [SCL091]

ou, de manière plus polémique, dans cette note d'un article de SCL, où l'auteur s'attaque directement à un autre (à ses connaissances) alors que nous avons vu qu'en général, et dans le corps du texte, l'opposition s'établit à l'encontre d'un ensemble d'approches :

À la lecture de ce passage et de bien d'autres textes de la même veine, **on en vient à se demander quelles connaissances de la rhétorique classique a effectivement G. Lakoff**, lui qui ne cesse

d'affirmer que "dans les théories classiques du langage, la métaphore était considérée comme une question de langue et non de pensée " (Lakoff, 1992). **On reste d'autant plus perplexe qu'il ne précise pas ce qu'il entend au juste par "** théories classiques du langage ". **Le flou de la désignation amalgame des conceptions irréductibles les unes aux autres.** [SCL095-notes]

Ce qui est reproché aux approches évaluées comme trop restrictives ou trop globalisantes, c'est de fait leur manque de profondeur. La profondeur est un critère décisif pour argumenter en faveur d'un point de vue, de même que l'importance d' « *aller plus loin* » pour saisir les problèmes dans leur complexité.

Cette façon de voir nous semble très prometteuse pour l'analyse textuelle et nous voudrions montrer, au cours de l'examen de quelques notions, que **la limitation de la variation** aux seules contraintes du système de la langue **peut être discutée** et qu'**il faut sans doute aller plus loin et s'interroger sur la** nécessaire relativisation des catégories elles-mêmes et des échelles. [SCL048]

L'intérêt de « *voir large* » va donc de pair avec celui de mener des « *études de détail* » ou un « *examen minutieux* », pour « *gagner en précision* », en « *rigueur* » et « *en profondeur* », autrement dit pour avancer dans l'analyse de l'objet. La clarté, la rigueur, la précision, la profondeur sont appelées à se substituer aux « *impressions* » et « *analyses impressionnistes* », celles dont il est dit qu'elles « *se contentent* » de « *survoler* » des phénomènes linguistiques et littéraires qui « *excèdent les interprétations simplificatrices* » ; il faut donc « *aller plus loin* », au-delà de la « *simplicité apparente* » et « *approfondir la réflexion* » pour « *creuser* » les problèmes dans toute leur « *complexité* ».

Dans la continuité de la notion de profondeur, il arrive également qu'un point de vue soit qualifié de faux ou d'erroné :

Il faut en outre récuser l'assertion que "Montaigne ne s'intéresse guère à l'enfance". **Il est faux de lui dénier tout sentiment de l'enfance. Bien au contraire,** il s'y intéresse de très près, à la sienne comme à l'enfance en général, fût-il en train de voyager en Italie, où il mentionne les mots de la langue toscane qui la désignent. Et même cette réflexion est-elle signe d'indifférence ? [LET002]

Le paradigme du vrai n'est quant à lui jamais désigné en tant que tel, et l'évaluation d'un point de vue qui consiste à le décréter faux reste très rare. Au final, il semble possible de dire que le plus souvent, les critiques portent sur des lacunes ou sur des difficultés telles qu'elles empêchent de cerner certaines questions, ou les occultent complètement. De manière positive, il s'agit prioritairement de montrer en quoi un point de vue ou une approche peuvent permettre d'avancer.

C'est cette nécessité centrale d'avancer qui est explicitement thématifiée dans l'extrait d'introduction ci-dessous, dont nous avons déjà évoqué la spécificité. L'auteur propose en effet un retour sur sa propre approche :

Ceci dit, il n'est pas facile de revenir sur ses pas, parce que, **qu'on le veuille ou non,** ce n'est pas le moyen le plus naturel d'avancer. Certains, **trop bien intentionnés,** peuvent même y voir une littérale façon de reculer ou au moins y déceler un stratagème pour masquer l'impossibilité de quitter un fonds de commerce trop douillet. [SCL099]

La nécessité d'avancer, plus souvent implicite qu'explicite dans les introductions, est cependant soulignée en conclusion, où « *les perspectives ouvertes par la recherche* » permettent d'insister sur son caractère « *prometteur* » en termes de « *gain* », de « *bénéfices* » et de « *progrès* », et sur « *la nécessité de continuer dans cette voie* ».

Le principe de la cumulativité scientifique (Bourdieu, 2001) se présente donc comme un principe central : y compris quand la proposition faite par l'auteur est celle d'un changement de paradigme, il ne s'agit pas de faire table rase de l'existant, mais de justifier par les difficultés que posent les approches attestées un tel changement de paradigme.

A côté de valeurs axiologiques plus générales (en termes de « *bien* » en particulier), la clarté, la précision, la profondeur et la rigueur fondent le système de valeurs sous-jacentes à l'argumentation dans l'écriture de recherche. Pas de thèse taxée de « farfelue », rarement d'idée « osée » ou « révolutionnaire », et aucune approche « troublante » ou « jubilatoire » ; ce système de valeurs construit une identité de chercheur et de l'activité de recherche comme un mode de penser-agir caractérisé par l'intellection.

Les valeurs ainsi mises en évidence sont communes aux deux disciplines. Les axiologiques qui contribuent à les définir sont cependant plus nombreux en SCL. Par ailleurs, des spécificités disciplinaires se manifestent dans les valeurs mobilisées pour argumenter en faveur de l'intérêt de l'objet et de l'apport de l'article. Les valeurs font l'objet d'investissements disciplinaires particuliers, qui montrent la pluralité des identités disciplinaires y compris au sein d'un même champ.

La valeur de scientificité dans les articles de SCL

Dans les articles de SCL, une des évaluations souvent convoquée par contraste avec ceux de LET a trait à la scientificité de l'approche défendue ou contestée. On le voit notamment dans les formes d'opposition, fréquentes comme nous l'avons dit dans les textes de cette discipline. Prenons-en pour exemple ces extraits d'un article de SCL, qui interviennent dans une première partie titrée « *le biais initial* ». Après une introduction qui commence en ces termes,

Nous souhaitons montrer dans cet article en quoi la synthèse de la parole éclaire la problématique du rythme de la parole et nous confronte aux **limites de nos modèles actuels**. Il sera montré que **la méthode traditionnelle de reconstruction du rythme en synthèse de la parole à partir du texte introduit un biais initial**.

l'auteur expose les ressorts du « biais initial » :

Traditionnellement, le rythme de la parole se recrée indirectement par l'effet combiné d'un ensemble de paramètres. La manipulation du rythme en synthèse de la parole s'effectue à partir des paramètres de F0 et de durées (actuellement, très peu de systèmes encore manipulent explicitement l'intensité). Pour ce faire, **il est courant d'inférer** la structure temporelle de l'énoncé (i.e., le calcul de l'ensemble des durées de chaque phone et des pauses) à partir du regroupement de mots établi sur la base des structures intonatives [...]. L'organisation temporelle de la parole **est donc le plus souvent conçue** à travers ce filtre [...]. **Ce type d'approche, motivé par des préconceptions linguistiques théoriques, entrave selon nous l'observation objective des événements de parole, et possiblement la compréhension plus profonde sur la manière dont le rythme de la parole est construit et perçu** (fig. 1). [SCL087]

L'utilisation de l'adverbe « *traditionnellement* », de désignations généralisantes (« *la méthode traditionnelle* », « *ce type d'approche* »), du *il* impersonnel et de l'adjectif « *courant* » (« *il est courant de...* »), et enfin du passif (« *est le plus souvent conçu comme* ») permettent d'éviter de s'opposer à des travaux en particulier, qui ne sont donc plus qu'une illustration parmi d'autres possibles du type d'approche incriminé. Ici, il ne s'agit pas seulement de pointer des lacunes pour montrer par contraste le caractère inédit de l'étude, mais de signaler des difficultés plus complexes qui « *entravent l'observation objective* » « *et possiblement la*

compréhension plus profonde ». L'opposition sert à mettre en évidence la scientificité de l'approche par contraste avec les approches à dépasser ; celles-ci font en effet l'objet d'une dépréciation qui remet en cause leur scientificité même (avec l'entrave à l'objectivité, il est question de « *préconceptions linguistiques* »).

On retrouve ce procédé dans nombre d'articles de SCL : le point de vue répandu ou traditionnel auquel l'auteur dit vouloir s'opposer est évalué négativement à l'aune de critères de scientificité, en termes de présupposés et d'implicite en particulier. Ces résultats se conforment donc à ce que nous avons montré jusqu'ici de l'importance tendancielle en SCL de positionner son approche en tant qu'approche scientifique, à l'aide en particulier d'un lexique qui définit un rapport scientifique à l'objet.

Cela étant, comme nous l'avons vu avec le lexique transdisciplinaire, les épistémologies au sein de la discipline sont plurielles, et peuvent aller jusqu'à évaluer négativement ce paradigme de la scientificité. C'est ce que montre cette citation qui intervient dans une note d'un article de SCL, et dont nous avons déjà commenté le positionnement épistémologique spécifique à partir des commentaires faits par l'auteur sur les termes *approche* et *méthodologie*¹. On retrouve ici une évaluation négative de la notion de démonstration et plus largement de la question de la scientificité :

" Pour cette raison aussi, **l'usage intempestif de la "scientificité" exerce une nocivité** particulière parce qu'il aboutit toujours à **poser des bornes, des douanes, au total à imposer des croyances (des intérêts) sous le vocable de "démonstration"**. Les rigidités d'armure incarnent les dogmes et l'immobilité, empêchent l'interrogation, bloquent la qualité investigatrice la plus haute qu'est l'imagination conceptuelle. **Il n'y a dorénavant de connaissance que du flou, et les poseurs de frontières ne sont que des installateurs de barbelés.** " (Porcher, 1997 : 14) [SCL007-notes].

Très virulente, la critique du paradigme de la scientificité repose sur le caractère heuristique du paradigme, conformément à ce que nous avons dit jusqu'ici de l'intérêt et des limites d'un point de vue : la démonstration et la scientificité sont incriminées en termes d'obstacles, de rigidité et de dogmatisme. Cette critique va jusqu'à revendiquer l'intérêt d'une « *connaissance du flou* », alors que cet axiologique est généralement utilisé dans un sens négatif comme nous l'avons vu. La discussion épistémologique consiste à défendre explicitement une définition de ce que doit être la connaissance, à l'encontre de critères propres à une définition de la science fondée sur les modèles dominants des sciences dures et expérimentales.

Finalement, la valeur de scientificité d'une approche peut donc être mise au rang de ses qualités ou de ses limites. Elle peut être jugée lacunaire, mais parfois aussi nocive pour la compréhension des faits linguistiques, et l'on voit donc se manifester des identités disciplinaires adverses, en référence à un modèle canonique de scientificité ; qu'il soit abordé de manière positive ou négative, celui-ci, parce qu'il se manifeste dans des formes d'opposition, tend à se présenter comme un modèle homogène et unifié.

¹ Rappelons ici un extrait du propos de l'auteur, analysé dans la partie consacrée au lexique transdisciplinaire : « *Le choix du terme 'approche' traduit une recherche de souplesse, de relativisation des dogmatismes théoriques, une ouverture au métissage épistémologique, et une modestie dans les attentes que l'on peut entretenir face à un travail de recherche, ce que ne connoterait pas nécessairement 'méthodologie'* » [SCL007].

Matérialité et portée littéraire du texte étudié en LET

Nous avons vu que dans les articles de LET, les axiologiques sont moins utilisés qu'en SCL pour évaluer l'intérêt ou les limites d'un point de vue sur l'objet étudié. Cela rejoint le constat que nous avons fait à propos des introductions : les articles de LET tendent à se centrer sur l'objet, et à préférer à la mise en perspective de l'approche avec d'autres approches du champ un plaidoyer sur l'intérêt de l'objet et son caractère central pour la compréhension de l'œuvre ou du fait littéraire analysés.

Parmi les axiologiques communs aux deux disciplines mis en évidence *supra*, ce sont en LET ceux qui concernent la nécessité de creuser les problèmes au-delà de leur apparence qui sont prioritairement utilisés. Dans les termes propres à la discipline, il s'agit de se démarquer d'une lecture du texte basée sur des impressions, comme nous le voyons dans l'extrait suivant :

Plusieurs commentateurs vantent également la valeur documentaire de cet épisode, dont le langage semble n'être que pure transparence, comme une sorte de fenêtre ouverte sur un spectacle charmant. **Une telle impression est sans doute renforcée par** la relative originalité de ce morceau, qui porte moins que d'autres la marque de conventions littéraires. **En réalité ce long passage est fort complexe et son efficacité poétique est moins due à la beauté du modèle qu'à un effet de l'art.** [LET029]

On voit ici comment se trouve opposée l'« *impression* » de plusieurs commentateurs (qui sont dits « *vanter* » la valeur documentaire du texte considéré) et la « *réalité* » du texte, c'est-à-dire, dans les termes de l'auteur, sa « *complexité* », et les ressorts de son efficacité.

Un paradigme axiologique spécifique aux LET se manifeste donc très nettement, qui tient à l'importance accordée à la réalité du texte, comme gage de la validité et de la pertinence de l'interprétation. C'est ce que l'on retrouve par exemple dans ce premier extrait, tiré d'un article où l'auteur propose d'étudier l'usage du détail dans l'œuvre de Montesquieu,

Ces éléments, *détail* et *cas particulier*, sont en quête d'un statut qui fait l'objet d'une réflexion continue dans l'œuvre de Montesquieu. **Il ne s'agit que de résister à la tentation de les renvoyer à la condamnation classique du détail « polluant », ou à sa variante positive, l'objet curieux.** [LET047]

et dans ce second extrait, tiré de l'introduction d'un article qui prend comme point de départ la digression dans l'œuvre de Balzac :

Dès qu'il y a envers, existent des possibilités et des jeux de renversement. **La tentation peut être grande en effet de renverser** sujet et hors-sujet et de voir l'essentiel d'une œuvre dans ses digressions – **certaines critiques y ont d'ailleurs cédé.** Outre le fait que **ce type de renversement semble quelque peu spécieux, il a tendance à « affadir » le phénomène digressif et lui faire perdre sa spécificité.** **En revanche, ce renversement est nettement plus intéressant – peut-être car plus surprenant** – quand il est opéré par l'auteur lui-même. C'est souvent le cas dans certains textes ironiques, que l'on a appelés les textes « digressionnistes ». [LET025]

Ces passages sont nettement critiques, puisqu'ils comportent des oppositions à des points de vue présentés comme le fait de certaines approches ou de certains commentateurs. Il est intéressant de constater que c'est la « *tentation* » de l'interprétation qui fait l'objet d'une incrimination ; aussi est-il question dans le second extrait de stigmatiser dans une incise les critiques qui ont « *cédé* » à cette tentation ainsi définie dans un sens moral, et d'opposer à cette dernière la « *spécificité* » de l'objet et la définition de son véritable intérêt.

Outre l'importance accordée à la matérialité du texte, les axiologiques utilisés en LET présentent une autre spécificité disciplinaire : celle-ci concerne de nouveau la centration sur

l'objet étudié et se situe dans les évaluations esthétiques mobilisées pour décrire l'œuvre et en vanter l'intérêt. Nous en donnerons pour exemple le début d'une introduction, où les axiologiques esthétiques abondent :

Essayer de lire "l'Art de conférer" comme si on ne l'avait jamais lu, le relire une n^{ème} fois en oubliant ou en feignant d'oublier tout ce qu'on sait ou croit savoir de Montaigne, dans le dessein de **réveiller l'énergie jubilante** de ce chapitre qui **déploie et entrelace** thèmes et images récurrents, qui **invente plus que dans tout autre des formules mémorables** et dont les "j'aime", les "je hais" rageurs imposent la présence de son "incomparable auteur", alors même que celui-ci l'écrit (dit-il) pour condamner de tels mouvements. [LET012]

De nouveau, l'auteur de l'article de LET est ce lecteur qui se présente comme un interlocuteur privilégié de l'œuvre étudié (il sait en reconnaître les qualités et accéder à ses significations profondes). Les évaluations esthétiques sont spécifiques à l'objet des articles de LET, i.e. la littérature, et sont inversement *a priori* bannies en SCL. Leur utilisation fait de l'auteur de l'article un interprète qui accède à la subtilité de l'œuvre, et on voit se dessiner une épistémologie possible en Lettres, selon laquelle l'herméneutique serait une activité fondamentalement artistique, comme nous l'a suggéré aussi l'usage de certains termes transdisciplinaires¹. Par ailleurs, l'abondance de ces évaluatifs peut également signaler la mise en œuvre dans l'article d'un « goût de l'écriture » des auteurs de LET, à l'instar de l'usage par exemple du passé simple comme nous l'avons indiqué.

L'étude des introductions et du lexique axiologique permet de retrouver et d'affiner un certain nombre de caractéristiques du genre esquissées précédemment. La définition de l'objet et des objectifs de la recherche présentée dans l'article montre qu'il s'agit à la fois de s'inscrire dans le champ et de signaler la singularité de l'apport. Les axiologiques récurrents dans l'évaluation de l'intérêt et des limites d'un point de vue se conforment à cette exigence : les valeurs convoquées se présentent comme des valeurs de scientificité (la précision, la clarté, la rigueur).

Elles se spécifient selon les disciplines, et peuvent aller jusqu'à manifester entre les SCL et les LET, et au sein des SCL, des définitions différentes de la science. Une exigence propre à ce champ est de positionner son approche par rapport à cette idée de scientificité, souvent en l'opposant à d'autres approches jugées doxiques. En LET, la centration sur l'objet permet d'en définir l'intérêt pour la compréhension de l'œuvre étudiée, et l'auteur se présente comme un interprète privilégié, qui décode le texte à l'attention de son lectorat.

Comme les introductions, les axiologiques participent de procédés caractéristiques du genre, relativement figés et routinisés, parfois aussi plus singuliers, dans certains cas polémiques, qui ont pour fonction de signaler en quoi et comment le point de vue ou l'approche peuvent faire avancer la recherche, et légitimer en cela l'apport de l'article et la figure de son auteur. Au final, cette nécessité affichée de l'avancée de la recherche légitime aussi l'activité de recherche elle-même, et le sens social de cette fonction d'« auteur qui est un chercheur ».

¹ Cf. dans LET099 l'usage spécifique de *hypothèse*, commenté dans la partie du chapitre 5 consacrée au lexique transdisciplinaire (5.3.): « *Ainsi se justifie l' hypothèse (délirante) qui a motivé cet article : « Antonio Tabucchi » est peut-être tout simplement un ultime avatar de Pessoa, hétéronyme posthume et imprévu, déniché tardivement au fin fond de la « malle pleine de gens » »*. [LIT099]

7.6. Bilan : surénonciation et argumentation

L'analyse de la subjectivité et de la polyphonie du genre de l'article montre comment celles-ci servent la construction du point de vue de l'auteur dans le texte. Que l'article soit plus ou moins explicitement argumentatif, il comporte toujours une dimension argumentative (Amossy, 2000), en faveur de ce point de vue.

L'instance auctoriale qui se présente comme responsable des contenus assertés dans le texte se construit à partir de la prise en charge énonciative et des marques de subjectivité et de polyphonie. Elle se définit comme une caution de scientificité, instanciable en tant que telle à l'envis, mais qui n'a de cesse d'affirmer sa singularité.

Cette figure de l'auteur repose sur une tension entre le marquage d'une subjectivité, la dilution de la personne énonciative dans un collectif (le lectorat du texte étudié, le sujet épistémique, cf. par exemple avec les emplois de *nous*), et enfin l'effacement énonciatif, i.e. l'énonciateur universalisant des tours impersonnels et du mode délocuté. Cette tension caractéristique du genre est résolue différemment selon les profils d'articles comme nous l'avons montré ; ils font apparaître le poids de pratiques idiomatiques, mais également de pratiques disciplinaires, et suggèrent le rôle joué par cette tension dans une dimension plus descriptive ou plus polémique de l'article.

L'instance auctoriale se construit selon une posture de surénonciation, telle que la définit A. Rabatel (2004), comme une position de surplomb dans le jeu polyphonique de prise en charge énonciative et de hiérarchisation des points de vue. Elle repose en particulier sur ce que nous pouvons appeler des coups de force dans l'assertion, en raison d'une double tendance à l'effacement énonciatif et au marquage d'une subjectivité, comme nous l'avons vu en particulier avec le *il* impersonnel, ou comme on le voit dans l'extrait suivant :

Chacun à leur manière, **tous ces exemples militent en faveur** d'une approche unifiée et homogène des citations directes et indirectes d'une part, libres ou contraintes d'autre part [SCL029]

Alors que l'approche en question a été introduite au début de l'article comme un choix théorique de l'auteur (*nous voudrions montrer, nous soutiendrons*), cette phrase inaugurant la conclusion apparaît comme un coup de force de l'assertion : par la modalisation que porte le verbe, et le statut syntactico-sémantique des exemples, ces derniers se présentent comme responsables d'une action de militantisme. L'expression, quasi lexicalisée, fait apparaître une forme de surénonciation routinisée, induite par la phraséologie même, comme c'est le cas également du *il* impersonnel.

Les tours polyphoniques sont également concernés par ces coups de force assertifs.

Ce n'est évidemment pas une redécouverte du personnage classique, du héros réaliste ; **au contraire, ce qui** se profile ici, c'est un sujet frappé d'indistinction (...) [SCL052]

L'opposition à un point de vue implicite, doublement marquée par la négation et le connecteur *au contraire*, et appuyée par la focalisation, sert l'argumentation en faveur du point de vue de L1/E1, l'adverbe *évidemment* venant encore renforcer l'autorité que ce point de vue doit à sa position de surplomb.

Les assertions, et avec elles le statut « évidentiel » du savoir (Dendale & Tasmowski, 2001) dépendent du type de confrontation des points de vue qu'introduisent les marqueurs polyphoniques et/ou épistémiques. Ainsi des injonctions telles que « *Il conviendrait assurément d'étudier de plus près...* » LET004, « *il faut sans doute aller plus loin et s'interroger sur la nécessaire relativisation...* » SCL048, se présentent comme davantage « discutées » quant à leur certitude que l'assertion citée précédemment.

La position de surplomb de l'instance auctoriale repose également sur la gestion de l'hétérogénéité des discours autres, signalés comme tels, et plus largement sur les références à l'existant. Nous avons vu que l'article rend compte de sa constitution intertextuelle et interpersonnelle dans ses différentes strates, affichant une quête de traçabilité, voire d'exhaustivité dans la représentation de la circulation des discours. En même temps, un ensemble de procédés servent à indexer la paternité du texte à son seul auteur, et à présenter ce dernier dans une position de surplomb par rapport à l'ensemble du champ de recherches en SCL, ou des interprétations du texte étudié dans le cas des articles de LET.

Par exemple, les procédés qui servent la gestion de l'attribution énonciative du dit pour éviter la confusion entre les voix peuvent aller jusqu'à nier le rôle de co-constructeur des savoirs d'un auteur cité : dans les cas où la référence à un auteur ou un texte, postposée au contenu propositionnel, est introduite par une formule comme *voir aussi*, l'auteur se présente comme seul responsable des propos ; signalant toutefois qu'ils ont été produits aussi par d'autres, il se montre fidèle à la déontologie scientifique, et retire de cette stratégie le double bénéfice de la légitimité du dit, et de son attribution.

Nous avons également montré à propos des modes de désignations de l'existant le rôle joué par les noms de courant et par les étiquettes généralisantes dans cette position de surplomb. L'auteur s'octroie le privilège de la nomination des paradigmes, et peut également se présenter dans une position d'arbitrage, s'attribuant le rôle de tiers en proposant une alternative nouvelle dans un débat. Dans les formes d'opposition, la source du point de vue est désignée de manière spécifique, de sorte que l'attaque *ad personam* est évitée, sauf parfois dans certains énoncés particulièrement polémiques, et placés en ce cas en notes.

L'évaluation axiologique de l'existant et de l'approche présentée dans l'article sur des valeurs récurrentes, caractéristiques de l'activité de recherche renforce cette position de surplomb de l'instance auctoriale ; la recherche présentée dans l'article se trouve légitimée par le fait de se présenter comme heuristique et apte à faire avancer le champ de recherches en SCL, et la compréhension de l'œuvre étudiée et de la littérature en LET.

La rhétorique de l'article est telle que la figure de son auteur repose sur la surénonciation, et se constitue *via* une instance énonciative foncièrement polyphonique. Ainsi en LET, l'auteur va jusqu'à adopter momentanément le point de vue de l'auteur étudié pour mieux faire parler le texte : il énonce au nom de l'ensemble du lectorat, mais également en tant que lecteur singulier. Plus largement, la figure de l'auteur se présente comme une instance décentrée, qui adopte provisoirement certaines positions antagonistes – la question rhétorique en est un bon exemple – selon une stratégie de reconstruction du processus de la découverte. Le jeu énonciatif est tel que l'assertion indexable à l'auteur, en tant que personne, se présente comme scientifiquement fondée et discutée, comme on en voit un nouvel exemple encore dans le dédoublement énonciatif entre le corps du texte et les notes de bas de page ci-dessous :

Le fait qu'**immanquablement**^a, pour les items lexicaux, la dénomination débouche sur une relation de sens codée permettant l'utilisation de la dénomination pour les occurrences du concept général dénommé nous a conduit in fine à poser le caractère sémantique codé comme trait définitoire des dénominations métalinguistiques : il y a dénomination, parce que l'item lexical a un sens codé ou préconstruit.

[En note :] a. Le **immanquablement** est bien sûr trop fort, puisque l'on peut fort bien, pour des items lexicaux comme *or*, *eau*, *chat*, etc., renoncer à un sens descriptif ou représentationnel et adopter, précisément à cause de leur proximité dénominative, le mode de fonctionnement causaliste (la *référence causale*) postulé pour les noms propres par Kripke (1972) et Devitt (1974 et 1976). [SCL099]

Le jeu complexe des instances énonciatives, au niveau de la note et du dédoublement qu'elle introduit signale ici un jeu avec les normes du genre, la note allant jusqu'à invalider le contenu du corps du texte ; il se révèle aussi être un procédé caractéristique du genre, spécifique de l'article par rapport à d'autres genres du discours scientifique¹ : l'instance auctoriale peut se présenter comme une instance clivée, mise en scène dans un processus de construction des savoirs qui repose sur une incessante confrontation de points de vue.

¹ C'est ce qu'a montré la comparaison de la posture de surénonciation dans deux genres du discours scientifique, les articles et les dictionnaires de sciences du langage (Grossmann et Rinck, 2004).

Conclusion

L'étude du genre de l'article de recherche dans deux disciplines de sciences humaines, les sciences du langage et les lettres, permet au final d'avancer dans la caractérisation des textes ; elle affine les descriptions existantes, et refonde certaines idées préalables que l'on peut se faire des écrits de recherche.

Le premier aspect frappant est celui de la diversité considérable des textes ; il contredit les approches qui gommant leur hétérogénéité au profit de constatations générales sur les standards en vigueur. Le second a trait à la question de la stabilité du genre. Le genre se caractérise par des traits centraux (cf. le présent, l'impersonnel, etc.) et par des traits marginaux, qui résultent de prescriptions négatives, prohibant leur usage ou appelant un usage modéré (cf. le point d'exclamation) ; ces traits marginaux sont cependant attestés, en vertu de la diversité des possibilités laissées ouverte par le genre. Si la stabilité du genre tient ainsi à ce qui est commun aux textes, elle repose également sur la manière dont leur diversité est structurée, en fonction de faisceaux de critères linguistiques. Le mode narratif présent dans certains textes dont l'objet comporte une dimension historique en est un exemple, les modèles de texte établis sur la base de leurs indices de structuration et des systèmes sémiotiques en jeu en sont un autre.

Cette stabilité du genre rend possible une analyse linguistique qui dépasse certaines difficultés posées par la description de la langue générale. Elle offre une prise pour l'analyse de la fonction des unités linguistiques dans les textes, et montre qu'il est possible de dégager des catégories relativement stables (cf. les guillemets, le lexique transdisciplinaire, les modes de références à l'existant). Elle permet ainsi de mieux cerner la diversité fonctionnelle d'une expression linguistique et la diversité des modes d'expression à partir d'une entrée notionnelle. Le genre se révèle être un point de départ productif en linguistique, pour pallier les risques d'une description trop unifiance et ceux d'un inventaire trop hétéroclite, ou non hiérarchisé, car établi sans référence à la normalisation induite par le niveau global sur le niveau local des textes.

Les approches énonciatives en linguistique, mobilisées dans l'étude d'un genre, montrent l'intérêt de prendre comme point de départ les marques traditionnellement décrites, et de les mettre en perspective les unes avec les autres pour analyser le jeu polyphonique au niveau du texte et les procédés argumentatifs qu'il sert. En retour, si les approches menées au niveau de l'énoncé sont particulièrement utiles pour décrire des marques en tant que telles, leur analyse au niveau du texte montre toute la complexité de l'énonciation représentée. Les questions soulevées dans le versant théorique comme dans le versant empirique de notre

investigation représentent une première exploration à prolonger par d'autres analyses. A titre programmatique, signalons en particulier l'hypothèse que l'énonciation représentée ne peut être décrite avec les termes qui servent à décrire l'énonciation effective ; le pronom *nous* ou la question rhétorique suggèrent l'intérêt d'envisager en termes de distinction ou de mise en communauté de l'énonciation ce qui est traité en termes de marques de soi d'une part et de l'autre d'autre part.

En ce sens, il serait intéressant de tenter de rapprocher la notion de figure de l'auteur et celle de figure du lecteur, et d'envisager éventuellement une définition énonciative de ce que U. Eco (1985) traite en termes d'auteur et de lecteur modèles, et de coopérativité interprétative. A ce titre, ce que nous avons analysé comme des procédures de mise en intrigue de la découverte nous semble représenter une entrée heuristique, quoiqu'elle mérite encore d'être mieux assise linguistiquement.

Par ailleurs, ce champ de recherches particulièrement débattu de la polyphonie appelle nous semble-t-il à contribuer à le rendre plus lisible, à l'aide notamment de synthèses, comme celle récemment proposée par P. Dendale et D. Coltier (2005) à propos de la notion de prise en charge énonciative dans la ScaPoLine. L'idée serait d'y intégrer aussi une perspective de confrontation des différentes approches ; malgré la difficulté du défi, et sans qu'il soit question ni de gommer leurs différences ni de simplifier leur propos, il s'agirait de favoriser le dialogue entre de multiples travaux et la construction, peut-être, de cadres d'analyse plus stabilisés.

Les propositions faites dans cette thèse ouvrent un certain nombre de perspectives pour l'étude des genres, des textes et des unités linguistiques, mais ce sont dans l'immédiat les approches polyphoniques qui retiennent prioritairement notre attention, et leur intérêt pour l'analyse des écrits de recherche, au sein des genres. En particulier, la confrontation des points de vue observable dans leur prolifération polyphonique au niveau de l'énoncé et de l'enchaînement des énoncés nous semble pouvoir servir d'entrée à l'analyse de la modalité épistémique et de la construction linguistique des savoirs.

En termes méthodologiques, la linguistique de corpus représente une source riche d'analyses exploratoires. Le va-et-vient entre une approche visant la quantification et une approche qualitative permet de concilier l'objectif de description globale du genre, de descriptions linguistiques fines de ses caractéristiques, et de profils diversifiés et nuancés de textes. La mise en relation des données, qui définit de manière centrale notre démarche, reste nécessairement inachevée et a à tout moment à affronter le risque de l'inconsistance de la quantification et de l'exemplification anecdotique.

En prenant comme point de départ le genre dans une analyse en corpus, nous avons montré à partir des différents traits retenus la diversité des textes, et tenté de voir notamment en quoi l'écriture de l'article se caractérise par des identités disciplinaires d'une part, et par une spécificité des pratiques des néophytes. C'est sur ces deux aspects que nous reviendrons pour finir.

La comparaison disciplinaire s'est avérée extrêmement productive. L'hypothèse d'une identité disciplinaire du genre, déjà validée pour des disciplines nettement distinctes, se vérifie également dans le cas où les deux disciplines ont des fondements épistémologiques et institutionnels tels qu'elles peuvent être considérées comme relativement proches.

Les différences constatées entre les articles de lettres et ceux de sciences du langage affectent tous les niveaux analysés, et pourraient être prolongées par l'étude de certaines tendances que nous n'avons pas pu vérifier statistiquement. Elles mettent au jour des modèles de textualité contrastés qui peuvent être interprétés en termes de rapport au lecteur et à l'objet, et sous l'angle de la figure de l'auteur et d'images diverses de la recherche.

En SCL, l'article se caractérise notamment par une exploitation du périphrase, par une distinction nette entre les sources théoriques et les données empiriques, par la nécessité de définir son approche à l'aune d'un (contre-)modèle de scientificité et par une discussion de l'existant qui signale dans certains cas une tendance nette à la polémique, avec l'importance en particulier de se positionner en s'opposant. En LET, l'article se caractérise par une textualité plus unifiée, centrée sur le corps de l'article et les notes ; *a priori* moins « reader-friendly », l'article implique cependant dans le texte une connivence avec un lecteur considéré en tant que lecteur de l'œuvre étudiée. L'auteur se présente comme un interprète privilégié, capable de décoder ce qui fait la signification profonde de l'œuvre et son importance littéraire. Son objet est l'intertexte, et l'activité de recherche se caractérise comme une herméneutique, qui n'a pas à se définir en termes de scientificité ; elle peut mettre en avant un goût de l'écriture et se présenter comme une activité fondamentalement artistique.

Cela étant, la comparaison disciplinaire montre aussi la diversité existante au sein de chacun des deux champs, et la continuité des pratiques. Nombre des entrées retenues permettent de conclure à une hétérogénéité plus importante au sein du champ des SCL, en termes textuels et épistémologiques ; il semble possible de formuler l'hypothèse d'un lien entre cette hétérogénéité des pratiques et la nécessité de se positionner par rapport à l'existant d'une part, et par rapport à un modèle canonique de scientificité défini en référence aux sciences dures et expérimentales.

En mobilisant, comme la comparaison disciplinaire, l'intérêt de la caractérisation différentielle des textes pour mettre en relief leurs traits communs et leurs spécificités, l'étude menée sur les styles d'auteur suggère des pistes d'analyse intéressantes, pour mieux rendre compte de cette diversité interne du champ des SCL : elle permet d'envisager à partir des textes une cartographie des objets centraux et des tendances adverses qui structurent une discipline. La pratique de l'exemple en SCL et en LET suggère aussi son intérêt dans une telle optique.

Au final, la comparaison disciplinaire se présente donc non comme une fin en soi, mais comme une entrée pour l'analyse des textes dans leur diversité ; cette diversité nous semble primordiale pour échapper aux descriptions unifiantes et réductrices. Si l'intérêt d'analyser l'écrit de recherche au sein de genres est de montrer qu'on n'écrit pas la science de la même manière dans les différents champs, il faut aussi convenir qu'il en va de même au sein d'un seul champ. En d'autres termes – et pour reprendre une expression dont nous avons montré l'importance dans les articles de LET – le linguiste se doit de « résister à la tentation » d'une interprétation des textes transparente vis-à-vis des pratiques supposées, en l'occurrence les pratiques culturelles d'une discipline.

En fondant l'analyse des pratiques culturelles sur la matérialité des textes, il s'agit de privilégier une approche descriptive et celle-ci peut peut-être, à terme, révéler sur les pratiques ce qu'une définition intuitive ou une approche psychologisante pourraient masquer. L'analyse de la figure de l'auteur dans les deux disciplines montre que des différences nettes interviennent, qui ont trait à des images plurielles d'une activité de recherche scientifique et savante. Cependant, dans les deux disciplines, le chercheur se manifeste textuellement dans sa complexité de sujet social, qui agit dans un but collectif, et affirme en même temps sa singularité au sein du champ. La légitimation de la recherche présentée dans l'article repose sur une phraséologie propre au genre, qui inscrit dans la langue la légitimation de l'activité elle-même, et du rôle social de ses acteurs.

L'analyse des textes en termes de pratiques culturelles est aussi aux fondements d'un second aspect que nous avons tenté d'explorer, celui d'une spécificité des textes produits par des néophytes, en l'occurrence des doctorants, par rapport aux textes de non-doctorants. Nous

avons vu que la caractérisation différentielle des textes est dans ce cas bien moins productive qu'avec la comparaison disciplinaire ; le constat n'est pas surprenant : comme nous avons affaire à des textes publiés après évaluation, ils relèvent bien du genre au titre duquel ils sont publiés. Nous vérifions cependant certaines spécificités, qui concernent en particulier un recours plus important chez les doctorants aux procédés didactiques de guidage du lecteur, à la mise en scène de soi dans les textes, mais dans des formes que nous pourrions qualifier de modérément « implicantes » (le *nous* plutôt que le *je* par exemple), et à des modes de renvoi aux sources globalement plus canoniques dans le nombre de références et les formes qu'elles prennent.

Dans une perspective de didactisation du genre, nous retiendrons prioritairement cette idée que les pratiques des doctorants tendent à davantage se conformer à un modèle relativement canonique du genre. Les différents traits qui en font la spécificité indiquent notamment une tendance à l'explicitation maximale. Ils suggèrent ainsi un rapport spécifique à la norme, qui rend complexe la question du modèle de référence à convoquer pour accompagner les scripteurs dans l'écriture d'un article. Il ne s'agit pas de didactiser l'écriture de recherche en érigeant les pratiques expertes en modèle, car la différence de compétence générique telle qu'on l'observe au niveau des textes ne peut donner lieu à l'idée d'un manque à combler ou d'un écart à réduire dans la maîtrise du genre.

En revanche, l'idée d'une conscientisation des pratiques représente une perspective intéressante : la description d'un genre peut en cela servir de préalable pour mieux comprendre sur quels ressorts se fonde dans les textes la figure de leur auteur, et cette compréhension outillée peut à son tour permettre à l'apprenti-scripteur d'asseoir progressivement sa légitimité et d'en contrôler davantage les effets. En cela, le fait d'affiner, voire de refonder les caractéristiques supposées d'un genre par des études empiriques fines de la diversité des textes, peut développer une conscientisation des pratiques qui intéresse, au-delà des seuls scripteurs doctorants, une réflexivité dans le champ scientifique.

Revenons cependant sur le problème du statut du scripteur. A travers cette variable, nous avons cherché à analyser un degré d'acculturation au genre. Deux types d'interprétations de cette idée d'acculturation peuvent être envisagés au vu de la spécificité des textes de doctorants par rapport à ceux des non-doctorants. D'abord, l'idée d'un degré d'acculturation se base sur celle d'un degré de connaissance du champ et de ses enjeux ; c'est l'hypothèse interprétative avancée face aux modes contrastés de renvoi aux sources et à la faible présence de noms de courant dans les articles de doctorants. La connaissance du champ ne suffit cependant pas à elle seule à expliquer cet usage moindre : le scripteur doit se sentir légitimé à utiliser de telles étiquettes, qui en retour contribuent à une posture légitimante de surénonciation. Cette hypothèse revient à avancer que les pratiques des néophytes sont amenées à évoluer dans le sens des pratiques des non-doctorants.

D'autre part, l'acculturation à un genre peut être comprise en termes d'identités culturelles, à la manière des identités disciplinaires. La spécificité des textes des doctorants pourrait alors être l'indice de pratiques culturelles propres à une jeune génération de chercheurs. Elle résulterait de leur formation, et de leur pratique des textes en usage, et signalerait une possible évolution du genre ; par exemple, les doctorants de LET recourent plus que les non-doctorants aux procédés de guidage du lecteur. Cette caractéristique, que l'on retrouve dans l'article de SCL par rapport à celui de LET, suggère que le genre en LET pourrait se rapprocher des modèles observés en SCL, sous l'effet d'une généralisation des standards déjà adoptés par les jeunes chercheurs.

La question se pose cependant de savoir si de telles spécificités culturelles sont le signe d'une véritable évolution ou correspondent à des pratiques temporaires. Nous avons ainsi montré que des traits linguistiques contrastés pouvaient renvoyer à des pratiques différentes de la recherche en sciences du langage, avec du côté des doctorants des études

globalement plus empiriques. On voit alors la difficulté qu'il y a à dissocier les deux hypothèses qu'il est possible de formuler face aux différences entre les textes de doctorants et les textes de non-doctorants : rien ne permet de dire si le type d'études privilégiées par les doctorants correspond à une évolution de l'activité de recherche vers certains standards spécifiques, ou si elles sont à comprendre comme une exigence statutaire. Les normes du champ disciplinaire contraignent les choix de certains sujets et de certaines approches, mis en œuvre par le doctorant dans la thèse et dans l'article. L'évolution des normes du genre recouvre donc deux aspects : une évolution progressive et historique du champ ou une évolution en fonction du statut du scripteur. Le champ ne contraint sans doute pas de la même manière ce qui est attendu au titre d'entrée dans le champ scientifique, et dans la poursuite de la carrière.

Dans nos analyses, la spécificité des textes des doctorants a été traitée avec prudence ; les hypothèses formulées dépendent du type de phénomène concerné d'une part, et de la question d'autre part de savoir si l'effet du statut de l'auteur est mis en évidence au sein de l'une, de l'autre, ou des deux disciplines, et en parallèle ou non avec une différence disciplinaire. Aussi avons-nous mis l'accent parfois sur la difficulté posée à l'analyste face à la multiplicité interprétative de la variation observée.

Ces constats relatifs à la spécificité des pratiques des doctorants montrent en quoi l'analyse empirique des textes peut parfois poser problème, et requiert pour le moins d'éviter d'envisager de manière bi-univoque les caractéristiques des textes et les indices supposés rendre compte de leurs conditions sociales de production. Le statut de doctorant ou de non doctorant demande à déterminer ce dont il est indiciel. La description des textes valide ici l'idée que le statut du scripteur conditionne la production du texte, au sens où il exerce un effet observable sur ses caractéristiques. Cependant comme cet effet contextuel peut être compris de diverses manières, il appelle d'autres études, notamment des études diachroniques du genre, et la prise en compte d'autres entrées (par exemple, la formation des chercheurs dans le champ, le type d'étude présentée, etc.), pour une analyse multi-critériée et dynamique des genres.

Finalement, c'est de nouveau la question de savoir ce que le texte peut dire quant à ses extérieurs qui est au centre du débat ; elle signale la nécessité de ne pas assimiler trop hâtivement la variation observable dans les textes à des positions sociales prédéterminées, constat relativement bien établi aujourd'hui en sociolinguistique, et repris par M. Dabène (1991) à propos de la compétence scripturale. La variable « statut du scripteur » comme la figure de l'auteur aboutissent à une interrogation sur l'interprétation des textes à l'aune des pratiques où ils sont produits et interprétés. Elles suggèrent l'intérêt d'envisager la construction des identités et des objets sociaux depuis les textes, et dans cette perspective, la notion de genre et celle de figure de l'auteur convergent vers une même idée, celle d'aborder les textes et leurs sujets en termes de réseaux.

Références

Références bibliographiques

- ADAM, J.M., (1999), *Linguistique textuelle. Des genres de discours aux textes*, Paris, Nathan Université.
- ADAM, J.M., et HEIDMANN, U., (2004), « Des genres à la généricité. L'exemple des contes (Perrault et Grimm) », *Langages*, 153, 62-72.
- ADDA, G., MARIANI, J., PAROUBEK, P., RAJMAN, M., et LECOMTE, J., (1999), « L'action GRACE d'évaluation de l'assignation des parties du discours pour le français », *Langues*, 2 (2), 119-129.
- AFFERGAN, F., (1994), « Textualisation et métaphorisation du discours anthropologique », *Communications*, 58, Paris, Seuil, 31-43.
- ALI BOUACHA, M., (1995), « De l'ego à la classe de locuteurs : lecture linguistique des *Méditations* », *Langages*, 119, 79-94.
- AMORIM, M., (1996), *Dialogisme et altérité dans les sciences humaines*, Paris, L'Harmattan.
- AMOSSY, R., (ed), (1999), *Images de soi dans le discours. La construction de l'ethos*, Lausanne, Delachaux et Niestlé.
- AMOSSY, R., (2000), *L'argumentation dans le discours. Discours politique, littérature d'idées, fiction*, Paris, Nathan.
- AMOSSY, R., et MAINGUENEAU, D., (eds), (2004), *L'analyse du discours dans les études littéraires*, Toulouse, Presses Universitaires du Mirail.
- ANIS, J., (1988), *L'écriture, théories et descriptions*, avec la collaboration de Chiss, J.-L. et Puech C., Bruxelles, De Boeck Université.
- ANSCOMBRE, J.C. et DUCROT, O., (1976), « L'argumentation dans la langue », *Langages*, 42, 5-27.
- ARISTOTE, (1991), *Rhétorique*, Paris, Le Livre de Poche.
- AUSTIN, J.L., (1962), *How to do things with words*, Oxford, Oxford University Press.
- AUTHIER, J., (1981), « Paroles tenues à distance », in Conein, B., Courtine, J.J., Gadet, F., et Pêcheux, M., (eds), *Matérialités discursives*, Lille, Presses Universitaires de Lille, 127-142.
- AUTHIER-REVUZ J., (1982), « Hétérogénéité montrée et hétérogénéité constitutive : éléments pour une approche de l'autre en discours », *DRLAV*, 26, 91-151.
- AUTHIER-REVUZ, J., (1984), « Hétérogénéités énonciatives », *Langages*, 73, 98-111.
- AUTHIER-REVUZ, J., (1995), *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, Paris, Larousse, 2 tomes.
- AUTHIER-REVUZ, J., (1998), « Énonciation, méta-énonciation. Hétérogénéités énonciatives et problématiques du sujet », in Vion, R., *Les sujets et leurs discours. Énonciation et interactions*, Aix-en-Provence, Université de Provence, 63-79.
- BAILEY, R.W., (1969), "Statistics and styles : a historical survey", in Dolezel, L. et Bailey, R.W., (eds), *Statistics and style*, New York, Elsevier, 48-85.
- BAKHTINE, M., (1978), *Esthétique et théorie du roman*, Paris, Gallimard.
- BAKHTINE, M./ VOLOCHINOV, V.N., (1977), *Le marxisme et la philosophie du langage*, Paris, Les Editions de Minuit.
- BAKHTINE, M., (1984), *Esthétique de la création verbale*, Paris, Gallimard.
- BALANDIER, G., (1994), « L'effet d'écriture en anthropologie », *Communications*, 58, Paris, Editions du Seuil, 23-30.
- BALLY, C., (1909), *Traité de stylistique française*. Paris, Klincksieck.

- BARBEROUSSE, A., et GUYON, E.,** (2003), « Dessiner, calculer, transmettre. Écriture et création scientifique chez Pierre-Gilles de Gennes », *Genesis*, 20, Paris, Editions Jean-Michel Place, 145-166.
- BARON, R.,** (2006), « Le tournant de l'analyse du discours dans les études littéraires. Entretien avec Ruth Amossy et Dominique Maingueneau ». Disponible en ligne sur *Vox Poetica* à l'adresse <http://www.vox-poetica.com/entretiens/mainamoss.html>
- BARTHES, R.,** (1972), *Le degré zéro de l'écriture*, Paris, Seuil (édition originale 1953).
- BARTHES, R.,** (1973), *Plaisir du texte*, Paris, Editions du Seuil.
- BAUTIER, E.,** (1998), « Je ou Moi : apprentissage ou expression », *Les Cahiers Pédagogiques*, 363, 48-53.
- BAZERMAN, C.,** (1988), *Shaping written knowledge : the genre and activity of the experimental article in science*, Madison, University of Wisconsin Press.
- BEACCO, J.C.,** (eds), (1992), *Ethnolinguistique de l'écrit*, *Langages*, 105, Paris, Larousse.
- BEAUDOIN, V., et YVON, F.,** (2004), « Contribution de la métrique à la stylométrie », *Journées d'Analyse des Données Textuelles*, Louvain-la-Neuve. Disponible en ligne à l'adresse http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_009.pdf
- BEAUVISAGE, T.,** (2001), « Morpho-syntaxe et genres textuels. Exploiter des données morphosyntaxiques pour l'étude statistique des genres textuels : application au roman policier », *Traitement automatique des langues*, 42 (2), 579-608.
- BECKER, H.S.,** (2004), *Ecrire les sciences sociales. Commencer et terminer son article, sa thèse ou son livre*. Paris, Economica (édition originale 1986).
- BENVENISTE, E.,** (1966), *Problèmes de linguistique générale*, Paris, Gallimard, (t. 1).
- BENVENISTE, E.,** (1974), *Problèmes de linguistique générale*, Paris, Gallimard, (t. 2).
- BENVENISTE, E.,** (1970), « L'appareil formel de l'énonciation », *Langages*, 17, 12-18.
- BENZECRI, J.P.,** (1973), *L'analyse de données. La taxinomie* (t.1), *L'analyse des correspondances* (t.2), Paris, Dunod.
- BENZECRI, J.P.,** (1981), *Pratique de l'analyse des données : linguistique et lexicologie*, Paris, Dunod.
- BENZÉCRI J.P.,** (1992), *Correspondence analysis Handbook*, New York, Dekker.
- BERKENKOTTER, C., et HUCKIN, T.,** (1995), *Genre knowledge in disciplinary communication : cognition/culture/power*, Hillsdale, N.J. Lawrence Erlbaum.
- BERNET, C.,** (1983), *Le vocabulaire des tragédies de Racine*, Paris, Slatkine/ Champion.
- BERNSTEIN, B.,** (1975), *Langage et classes sociales. Codes sociolinguistiques et contrôle social*, Paris, Editions de Minuit.
- BIBER, D.,** (1988), *Variation across speech and writing*, Cambridge, Cambridge University Press.
- BIBER, D.,** (1989), "A typology of English texts", *Linguistics*, 27, 3-43.
- BIBER, D.,** (1993), "Using register-diversified corpora for general language studies", *Computational Linguistics*, 19 (2), 243-258.
- BIBER, D.,** (1995), *Dimensions of register variation : a cross-linguistic comparison*, Cambridge, Cambridge University Press.
- BILGER, M.,** (ed.), (2000), *Corpus, méthodologies et applications linguistiques*, Paris, Honoré Champion.
- BIZZELL, P.,** (1992), *Academic discourse and critical consciousness*, Pittsburg, University of Pittsburg Press.
- BLOQUET, G.,** (2004), *Vers une modélisation des collocations verbales du lexique méthodologique dans les écrits scientifiques : application au TAL*, Mémoire de DEA, sous la direction de Grossmann, F., et Tutin A., Grenoble, Université Stendhal.

-
- BLOQUET, G., et TUTIN, A.,** (2004), « Lexique transdisciplinaire et structure des articles scientifiques », *Colloque sciences et écritures*, mai 2004, Besançon. Disponible en ligne à l'adresse http://w3.u-grenoble3.fr/tutin/Publis/Bloquet_tutin.pdf
- BOCH, F.,** (1999), *Pratiques d'écriture et de réécriture à l'université. La prise de notes, entre texte source et texte cible*, Thèse de doctorat de Sciences du Langage, Presses Universitaires du Septentrion.
- BOCH, F., et GROSSMANN, F.,** (eds), (2001) : *Apprendre à citer le discours d'autrui*, *Lidil*, 24, Grenoble, Université Stendhal.
- BOCH, F., et GROSSMANN, F.,** (2001a), « De l'usage des citations dans le discours théorique. Des constats aux propositions didactiques », *Lidil*, 24, 91-111.
- BOCH, F., et GROSSMANN, F.,** (2002), Se référer au discours d'autrui : comparaison entre néophytes et experts, *Enjeux*, 54, 41-51.
- BORILLO, A.,** (1981), « Quelques aspects de la question rhétorique en français », *DRLAV*, 25, 1-33.
- BOSSON, J.L., FAUCONNIER, J., BONHOMME, I., FRANÇOIS, P., WEIL, G., CARPENTIER, P.H., et ROBERT, C.,** (1995), « Evaluation des méthodes statistiques dans les thèses de médecine », Grenoble, GRESIM. Disponible en ligne à l'adresse <http://www-sante.ujf-grenoble.fr/SANTE/statistique/p-value.html>
- BOUQUET, S.,** (1999), « La linguistique générale de Ferdinand de Saussure : textes et retour aux textes », Communication inédite présentée au Congrès ICHOLS. Disponible en ligne sur *Texto !* à l'adresse http://www.revue-texto.net/Saussure/Sur_Saussure/Bouquet_Linguist-gen.html
- BOURDIEU, P.,** (1979), *La distinction. Critique sociale du jugement*, Paris, Les éditions de Minuit.
- BOURDIEU, P.,** (ed), (1993), *La misère du monde*, Paris, Editions du Seuil.
- BOURDIEU, P.,** (1994), *Raisons pratiques. Sur la théorie de l'action*, Paris, Editions du Seuil.
- BOURDIEU, P.,** (2001), *Science de la science et réflexivité*, Paris, Raisons d'agir.
- BOURGAIN, D.,** (1988), *Discours sur l'écriture. Analyse des représentations sociales de l'écriture en milieu professionnel*, Thèse de doctorat d'état, sous la direction de Peytard, J., Besançon.
- BOYER, H.,** (1988), « Scription et écriture dans la communication journalistique », in Charaudeau, P., *La presse : produit, production et réception*, Paris, Didier Érudition, 71-92.
- BRANCA-ROSOFF, S.,** (1996), « Retour aux genres », in Auroux, S., Delesalle, S. et Meschonnic, H., *Histoire et grammaire du sens. Hommage à J.C. Chevalier*, Paris, Armand Colin, 189-203.
- BRANCA-ROSOFF, S.,** (ed), (1999), *Types, modes et genres de discours, Langage et Société*, 87, Paris, Maison des Sciences de l'Homme.
- BRANCA-ROSOFF, S.,** (1999a), « Entre langue et discours », *Langage et Société*, 87, 5-24.
- BRANCA-ROSOFF, S.,** (1999b), « Des innovations et des fonctionnements de langue rapportés à des genres », *Langage et Société*, 87, 115-129.
- BREIVEGA, K., DAHL, T. et FLØTTUM K.,** (2002), "Traces of self and others in research articles. A comparative pilot study of English, French and Norwegian articles in medicine, economics and linguistics", *International Journal of applied linguistics*, 12 (2), 218-239.
- BRES, J., DELAMOTTE-LEGRAND, R., MADRAY-LESIGNE, F., et SIBLOT, P.,** (eds), (1999), *L'autre en discours*, Montpellier - Rouen, Université Paul Valéry/ Praxiling.
- BRES, J.,** (1999), « Entendre des voix : de quelques marqueurs dialogiques en français », in Bres et al., (eds), *L'autre en discours*, Montpellier - Rouen, Université Paul Valéry/ Praxiling, 191-212.

- BRES, J.**, et **VERINE, B.**, (2002), « Le bruissement des voix dans le discours : dialogisme et discours rapporté », in Rosier, L., (ed), *Le discours rapporté, Faits de langue*, 19, Paris, Ophrys, 159-169.
- BRONCKART, J.P.**, (1985), *Le fonctionnement des discours. Un modèle psychologique et une méthode d'analyse*, Neuchâtel, Delachaux et Niestlé.
- BRONCKART, J.P.**, (1996), *Activité langagière, texte et discours. Pour un interactionnisme socio-discursif*, Neuchâtel - Paris, Delachaux et Niestlé.
- BRUNET, E.**, (1978), *Le vocabulaire de Jean Giraudoux*, Genève-Paris, Slatkine/ Champion.
- BRUNER, J.S.**, (1983), *Le développement de l'enfant : savoir faire, savoir dire*, Paris, Presses Universitaires de France.
- BRUNET, E.**, (1991), "What Do Statistics Tell Us?", *Research in Humanities Computing*, 1, 70-92.
- BRUNET, E.**, et **MULLER, C.**, (1988), « La statistique résout-elle les problèmes d'attribution? », *Strumenti critici*, Sett., 367-387.
- BUVAT, I.**, (2000), *Introduction à l'approche Bootstrap*, Paris, INSERM (U.494). Disponible en ligne sur <http://www.guillemet.org/irene/equipe4/coursem/Bootstrap.pdf>
- CAMPBELL, P.N.**, (1975), "The personae of scientific discourse", *Quarterly Journal of Speech*, 61, 391-405.
- CANVAT, K.**, (1996), « Types de textes et genres textuels. Problématique et enjeux », *Enjeux*, 37/38, 5-29.
- CASSANAS, A.**, **DEMANGE, A.**, **LAURENT, B.**, et **LECLER, A.**, (eds), (2003), *Dialogisme et nomination, Actes du troisième colloque Jeunes Chercheurs Praxiling*, Montpellier, Presses de l'Université Paul Valéry.
- CERVONI, J.**, (1987), *L'énonciation*, Paris, Presses Universitaires de France.
- CHARAUDEAU, P.**, et **MAINGUENEAU, D.**, (eds), (2002), *Dictionnaire d'analyse du discours*, Paris, Editions du Seuil.
- CHISS, J.-L.**, (1987), « Malaise dans la classification », *Langue française*, 74, 10-28.
- CLYNE, M.**, (1998), "Cultural differences in the organization of academic texts : English and German", in Cheschire, J., et Trudgill, P., *The sociolinguistics reader*, 2., London, Arnold, 315-347.
- COLLETTA, J.M.**, (1994), « Identité et légitimité des interactions au tribunal », in Moirand, S., Bouacha, A., Beacco, J.C., et Collinot, A., (eds), *Parcours linguistiques de discours spécialisés*, Berne, Peter Lang, 87-99.
- COMBETTES, B.**, (2002), « Analyse linguistique des textes et stylistique », *Langue Française*, 135, 95-112.
- COMPAGNON, A.**, (2003), « La notion de genre », Cours donné à l'Université de Paris IV. Disponible en ligne à l'adresse <http://www.fabula.org/compagnon/genre.php>
- CONDAMINES, A.**, (eds), (1999), « Corpus et traitement automatique des langues : pour une réflexion méthodologique », *Actes de l'atelier thématique TALN*, Cargèse.
- COSERIU, E.**, (1981), *Textlinguistik - Eine Einführung*, Tübingen, Narr.
- COTTERET, J.M.**, **MOREAU, R.**, (1969), *Le vocabulaire du Général de Gaulle*, Paris, Armand Colin.
- COXHEAD, A.**, (2000), "A new academic word list", *TESOL Quarterly*, 34 (2), 213-238.
- COXHEAD, A.**, (2002), "The Academic Word List : A Corpus-based Word List for Academic Purposes" in *Teaching and Language Corpora, TALC 2000*, Conference Proceedings, Atlanta, Rodopi.
- CRESSOT, M.**, et **JAMES, L.**, (1991), *Le style et ses techniques. Précis d'analyse stylistique*. Paris, Presses Universitaires de France (édition originale 1947).
- CULIOLI, A.**, (1990), *Pour une linguistique de l'énonciation. Opérations et représentations*, t.1, Paris, Ophrys.

-
- CULIOLI, A., (1999), *Pour une linguistique de l'énonciation. Formalisation et opérations de repérage*, t.2, Paris, Ophrys.
- CULIOLI, A., (1999), *Pour une linguistique de l'énonciation. Domaine notionnel*, t.3, Paris, Ophrys.
- DABENE, M., (1987), *L'adulte et l'écriture. Contribution à une didactique de l'écrit en langue maternelle*, Bruxelles, De Boeck.
- DABENE, M., (1991), « Un modèle didactique de la compétence scripturale », *Repères*, 4, Paris, INRP, 9-21.
- DABENE, M., (1996), « Aspects socio-didactiques de l'acculturation au scriptural », in Barré-De Miniac, C., (ed.), *Vers une didactique de l'écriture. Pour une approche pluridisciplinaire*, Paris-Bruxelles, INRP/ De Boeck.
- DAHL, T., (2003), « Metadiscourse in research articles », in Fløttum, F. et Rastier, F., (eds), *Academic discourse. Multidisciplinary approaches*, Oslo, Novus Press, 120-138.
- DAHLET, V., (2003), *Ponctuation et énonciation*, Presses Universitaires Créoles, Ibis Rouge éditions.
- DAMOURETTE, J., et PICHON, E., (1911), *Des Mots à la pensée. Essai de grammaire de la langue française*, Paris, Editions d'Artrey.
- DAUNAY, B., (2002), « Le lecteur distant. Positions du scripteur dans l'écriture du commentaire », *Pratiques*, 113-114, Metz, CRESEF, 135-153.
- DELAMOTTE-LEGRAND, R., (1996), « Polyphonie dans l'écriture », *Cahiers du français contemporain*, 3, 193-210.
- DELAMOTTE-LEGRAND, R., (2002), « Faire savoir son savoir. Polyphonie et mise en débat d'idées dans des copies d'étudiants », *Enjeux*, n° 54, Namur, CEDOCEF, 28-40.
- DELCAMBRE, I., (1990), « De l'argumentation à la dissertation. Analyse d'une démarche d'apprentissage », *Pratiques*, 68, 69-88.
- DELCAMBRE, I., et JOVENET, A.-M., (eds), (2002), *Lire-écrire dans le supérieur, Spirale*, 29, Lille 3.
- DELCAMBRE, I., et LABORDE-MILAA, I., (2002), « Diversité des modes d'investissement du scripteur dans l'introduction du mémoire professionnel », *Enjeux*, 53, Namur, CEDOCEF, 11-22.
- DELCAMBRE, I., et REUTER, Y., (2002), « Images du scripteur et rapports à l'écriture », *Pratiques*, 113-114, 7-28.
- DENDALE, P., et TASMOWSKI, L., (eds), (1994), *Les sources du savoir et leurs marques linguistiques*, Langue Française, 102, Paris, Larousse.
- DENDALE, P., et TASMOWSKI, L., (eds), (2001), *Le conditionnel en français, Recherches linguistiques*, 25, Metz, Université de Metz, CELTED.
- DENDALE, P., et COLTIER, D., (2005), « La notion de prise en charge ou de responsabilité dans la théorie scandinave de la polyphonie linguistique », in Bres, J., Haillet, P., Mellet, S., Nølke, H., et Rosier, L., *Dialogisme et polyphonie. Approches linguistiques*, Bruxelles, De Boeck/ Duculot, 125-140.
- DENIZOT, N., (2002), « Une démarche de lecture-écriture au lycée autour des images du scripteur », *Pratiques*, 113-114, 191-206.
- DETRIE, C., SIBLOT, P., et VERINE, B., (eds), (2001), *Termes et concepts pour l'analyse du discours, une approche praxématique*, Paris, Honoré Champion.
- DOLINE, C., (1999), « Le problème des genres de discours 45 ans après Bakhtine », *Langage et Société*, 87, 25-40.
- DONAHUE, C., (2002a), « Quelles stratégies pour aider l'étudiant écrivain à gérer la polyphonie énonciative ? », *Enjeux*, 54, 67-83.
- DONAHUE, C., (2002b), « Effets de l'écrit sur la construction du sujet textuel à l'université », *Spirale*, 29, 75-99.

- DUCROT, O., (1981), « La valeur argumentative de la phrase interrogative », in Berrendonner, A., et Grize, J.B., (eds), *Logique, argumentation, conversation*, Berne -Francfort, Peter Lang.
- DUCROT, O., (1982), « Note sur l'argumentation et l'acte d'argumenter », *Cahiers de linguistique française*, 4, 143-163.
- DUCROT, O., (1984), *Le dire et le dit*, Paris, Les éditions de Minuit.
- DUCROT, O., (2001), « Quelques raisons de distinguer « locuteurs » et « énonciateurs » », *Polyphonie – linguistique et littéraire*, 3. Disponible en ligne à l'adresse http://www.hum.au.dk/romansk/polyfoni/Polyphonie_III/Oswald_Ducrot.htm
- DUSZAK, A., (ed), (1997), *Culture and styles of Academic Discourse*, Berlin, Mouton de Gruyter.
- ECO, U., (1984), *Sémiotique et philosophie du langage*, Paris, Presses Universitaires de France.
- ECO, U., (1985), *Lector in fabula*, Paris, Grasset.
- EFRON, B., et TIBISHIRANI, R.J., (1993), *An introduction to the Bootstrap*, New York, Chapman and Hall.
- FABRE-COLS, C., (2001), « Trouver sa voix dans la rumeur des discours », *Lidil*, 24, 49-66.
- FLØTTUM, K., (2001), « Les liens énonciatifs : tentative d'une nouvelle typologie », *Polyphonie - linguistique et littéraire*, 3, 67-86. Disponible en ligne à l'adresse http://www.hum.au.dk/romansk/polyfoni/Polyphonie_III/Kjersti_Flottum.htm
- FLØTTUM K., (2002a), « Polyphonie au niveau textuel », *Romansk Forum*, 16, 339-350, Disponible en ligne à l'adresse <http://www.digbib.uio.no/roman/Art/Rf-16-02-2/fra/Flottum.pdf>
- FLØTTUM, K., (2002b), « Fragments guillemetés dans une perspective polyphonique », *Tribune* 13, Skriftserie for Romansk institutt. Disponible en ligne à l'adresse http://www.hum.au.dk/romansk/polyfoni/Diverse%20artikler/Flottum_guillemets.htm
- FLØTTUM K., (2005), « The self and the others : polyphonic visibility in research articles », *International Journal of applied linguistics*, 15 (1), 29-44.
- FLØTTUM, K., DAHL, T., et KINN, T., (2006), *Academic Voices. Across languages and disciplines*, Amsterdam - Philadelphia, John Benjamins.
- FLØTTUM, K., GJESDAL, A., KINN, T., VOLD, E.T., et DAHL, T., “Snapshots from the KIAP Project”, *International Symposium, Academic Voices in contrast*, University of Bergen, mai 2006.
- FONTANIER, P., (1968), *Les figures du discours*, Paris, Flammarion (édition originale 1821).
- FOUCAULT, M., (1969a), *L'archéologie du savoir*, Paris, Gallimard.
- FOUCAULT, M., (1969b) : « Qu'est-ce qu'un auteur ? », *Bulletin de la société française de philosophie*, 63 (3), 73-104.
- FRANÇOIS, F., (1980), *Linguistique*, Paris, Presses Universitaires de France.
- FRANÇOIS, G., (à venir), « Guillemets et italiques: Marques de circulation de discours? » *3ème Colloque International Ci-Dit*, Université Laval, Québec, octobre 2006.
- GADET, F., (1989), *Le français ordinaire*, Paris, Armand Colin.
- GADET, F., et PECHEUX, M., (1981), *La langue introuvable*, Paris, Maspero.
- GARDES-TAMINES, J., (1996), *La rhétorique*, Paris, Armand Colin.
- GENETTE, G., (1972), *Figures III*. Paris, Editions du Seuil.
- GENETTE, G., (1982), *Palimpsestes*, Paris, Editions du Seuil.
- GENETTE, G., (1983), *Nouveau discours sur le récit*. Paris, Editions du Seuil.
- GENETTE, G., (1987), *Seuils*, Paris, Editions du Seuil.
- GOODY, J., (1979), *La raison graphique ; la domestication de la pensée sauvage*, Paris, Les éditions de Minuit.
- GOODY, J., (1993), *Entre l'oralité et l'écriture*, Paris, Presses Universitaires de France.

- GOUGENGHEIM, G., MICHEA, R., RIVENC, P., et SAUVAGEOT, A., (1956), *L'élaboration du français élémentaire : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Paris, Didier.**
- GOUGENGHEIM, G., MICHEA, R., RIVENC, P., et SAUVAGEOT, A., (1964), *L'élaboration du français fondamental : étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Paris, Didier.**
- GRIZE J., (1990), *Logique et langage*, Gap - Paris, Ophrys.**
- GROSSMANN, F., (2002), « Les modes de référence à autrui chez les experts : l'exemple de la revue *Langages* », *Faits de langue*, 19, 255-262.**
- GROSSMANN, F., (2003), « Du discours rapporté au discours autorisé, le maniement des noms d'auteur dans l'article en Sciences Humaines », *Estudios de Lengua y Literatura francesas*. 14, 9-26.**
- GROSSMANN, F., et RABATEL, A., (eds), (à paraître 2007), Marquage et construction de la figure de l'auteur, construction du sujet apprenant dans les documents et situations didactiques, *Lidil*, 35, Grenoble, Université Stendhal.**
- GROSSMANN, F., et RINCK, F., (2004), « La surénonciation comme norme du genre. L'exemple de l'article de recherche et du dictionnaire en linguistique », *Langages*, 156, Paris, Larousse, 34-50.**
- GUESPIN, L., (1971), « Problématique des travaux sur le discours politique », *Langages*, 23, 3-24.**
- GUIBERT, R., (1989), *Jeux énonciatifs-enjeux évaluatifs. Rapport au texte d'étudiants-adultes préparant des diplômes de fin d'études*, Thèse de doctorat, sous la direction de Matalon, B., Paris, Ecole des Hautes Etudes en Sciences Sociales.**
- GUIBERT, R., (2001), « Citer et se situer. L'apprentissage de l'écriture avec les discours d'autrui », *Lidil*, 24, 29-47.**
- GUIBERT, R., (2002), « L'entraînement à la synthèse comme apprentissage du dialogisme », *Spirale*, 29, 145-163.**
- GUILBERT, L., et PEYARD, J., (eds), (1973), « Les vocabulaires techniques et scientifiques », *Langue française*, 17, Paris, Larousse.**
- GUILLEMIN-FLESCHER, J., (1984), « Enonciation, perception et traduction », *Langages*, 73, 74-97.**
- GUMPERZ, J.J., et HYMES, D.H., (eds), (1964), *The Ethnography of Communication. American Anthropologist*, 66, Special Publication. Washington, American Anthropological Association.**
- GUMPERZ, J.J., et HYMES, D.H., (eds), (1972), *Directions in Sociolinguistics: The Ethnography of Communication*, New York - Holt, Rinehart/ Winston.**
- HABERMAS, J., (1973), *Théorie de l'agir communicationnel*, t. 1 et 2, Paris, Fayard.**
- HABERT, B., FABRE, C. et ISSAC, F., (1998), *De l'écrit au numérique : constituer, normaliser, exploiter les corpus électroniques*, Paris, InterÉditions/ Masson.**
- HABERT, B., NAZARENKO, A., et SALEM, A., (1997), *Les linguistiques de corpus*, Paris, Masson/Armand Colin.**
- HALLIDAY, M.A.K., (1985), *An Introduction to Functional Grammar*, London, Edward Arnold.**
- HALLIDAY, M.A.K., (2004), *The language of science*, London, Continuum.**
- HALLYN, F., (2004), *Les structures rhétoriques de la science*, Paris, Editions du Seuil.**
- HAMANN, J., (1996), « La stylométrie au programme », *28^e journées de statistique*, Université de Laval. Disponible en ligne à l'adresse <http://www.scom.ulaval.ca/Au.fil.des.evenements/1996/06.20/mots.html>**
- HAROCHE, C., HENRY, P., et PECHEUX, M., (1971), « La sémantique et la coupure saussurienne : langue, langage, discours », *Langages*, 24, 93-106.**

- HENRY, P.**, (1977), *Le mauvais outil. Langue, sujet et discours*, Paris, Editions Klincksiek.
- HESLOT, J.**, (1980), « La formation des chercheurs à l'expression scientifique écrite », *Langage et Société*, supplément au n° 12, Paris, Maison des Sciences de l'Homme, 35-40.
- HOLMES, D.I.**, (1994). "Autorship Attribution", *Computers and the Humanities*, 28, 87-106.
- HOWELL, D.C.**, (1998), *Méthodes statistiques en sciences humaines*, Paris, De Boeck Université.
- HUNT, N.**, (1996), Boxplots in Excel, *The Spreadsheet User 3-2*. Version remaniée disponible en ligne à l'adresse <http://www.mis.coventry.ac.uk/~nhunt/boxplot.htm>
- HYLAND, K.**, (1998), "Persuasion and context : the pragmatics of academic metadiscourse", *Journal of pragmatics*, 30, 437-455.
- HYLAND, K.**, (2000), *Disciplinary discourses. Social interactions in academic writing*, London, Longman.
- HYLAND, K.**, (2001), "Humble servants of the discipline? Self mention in research articles", *English for Specific Purposes*, 20, 207-226.
- HYLAND, K.**, (2002), "Authority and invisibility: authorial identity in academic writing", *Journal of Pragmatics*, 34 (8), 1091-1112.
- HYLAND, K.**, (2005), "Stance and engagement: a model of interaction in academic discourse", *Discourse Studies*, 7 (2), 173-191.
- JACOBI, D.**, (1999), *La Communication scientifique : discours, figures, modèles*, Saint-Martin-d'Hères, Presses Universitaires de Grenoble.
- JACQUES, F.**, (1983), « La mise en communauté de l'énonciation », *Langages*, 70, 47-71.
- JAKOBSON, R.**, (1963), *Essais de linguistique générale*, Paris, Les éditions de Minuit.
- JANE BLEVINS, S.**, et **LE BIGOT, O.**, (2003), « Sur la genèse de résultats théoriques en physique », *Genesis*, 20, Paris, Editions Jean-Michel Place, 129-144.
- JAUSS, H.R.**, (1978), *Pour une esthétique de la réception*, Paris, Gallimard.
- KARA, M.**, (2004), « Pratiques de la citation dans les mémoires de maîtrise », *Pratiques*, 121-122, 111-142.
- KERBRAT-ORECCHIONI, C.**, (1980), *L'énonciation. De la subjectivité dans le langage*, Paris, Armand Colin.
- KILANI, M.**, (1994), « Du terrain au texte. Sur l'écriture de l'anthropologie », *Communications*, 58, Paris, Editions du Seuil, 45-60.
- KLEIBER, G.**, (1994), « Contexte, interprétation et mémoire : approche standard versus approche cognitive », *Langue Française*, 103, 9-22.
- KOCOUREK, R.**, (1982), *La langue française de la technique et de la science*, Wiesbaden, Brandstetter Verlag.
- KRAIF O.**, et **TUTIN A.**, (2006), "Looking for Semi-Frozen Expressions using an Aligned Corpus : an Application for Academic Writing for EFL Learners", *Colloque TALC 2006*, juillet 2006, Paris.
- KRISTEVA, J.**, (1969), *Séméiotikè. Recherche pour une sémanalyse*, Paris, Editions du Seuil.
- KRONNING, H.**, (1996), *Modalité, cognition et polysémie : sémantique du verbe modal devoir*, Uppsala, Acta Universitatis Upsaliensis, Coll. Studia Romanica Upsaliensia, 54.
- KUHN, T. S.**, (1983), *La structure des révolutions scientifiques*, Paris, Flammarion.
- KYLANDER, B.M.**, (1995), *Le vocabulaire de Molière*, Goteborg, Acta Universitatis Gothoburgensis.
- LABBE, D.**, (1983), *François Mitterrand : essai sur le discours*, Grenoble, La pensée sauvage.
- LABORDE-MILAA, I.**, (2002), « Polyphonie énonciative. Représentations d'étudiants en position d'évaluateurs de leurs pairs », *Spirale*, 29, 181-200.
- LABORDE-MILAA, I.**, **BOCH, F.** et **REUTER, Y.**, (eds), (2004), *Les écrits universitaires*, *Pratiques*, 121-122, Metz, CRESEF.

-
- LABOV, W., (1976), *Sociolinguistique*, Paris, Les éditions de Minuit.
- LABOV, W., (1978), *Le parler ordinaire*, t. 1 et 2, Paris, Les éditions de Minuit.
- LATOUR, B., et WOOLGAR, S., (1979), *Laboratory Life : the Social Construction of Scientific Facts*, Los Angeles - Londres, Sage.
- LAUFER, R., (1979), « Guillemets et marques de discours direct », in *La ponctuation, Recherches historiques et actuelles, Actes de la table ronde internationale CNRS*, Paris, Publications CNRS-HESO, 235-251.
- LEBART, L., (2002), *Procédures de validation en Text Mining*, Paris, ENSA (Club SAS). Disponible en ligne à l'adresse http://club-sas-stat.jeannot.org/journees/docs20020620/Lebart_SAS2002.pdf
- LEBART, L., (2004), « Validité des visualisations de données textuelles » *Actes des 7emes Journées internationales d'Analyse statistique des Données Textuelles*, Louvain, Presses universitaires de Louvain, 708-715.
- LEBART, L., et SALEM, A., (1994), *Statistique textuelle*, Paris, Dunod.
- LEON, P., (1971), « La joncture externe en français : nature et fonctions linguistiques et expressives, Essai de Phonostylistique », *Studia Phonetica*, 4, 57-66.
- LOFFLER-LAURIAN, A.M., (1980), « L'expression du locuteur dans les discours scientifiques », *Revue de linguistique romane*, 44, 135-157.
- MAINGUENEAU, D., (1992), « Le tour ethnolinguistique de l'analyse du discours », *Langages*, 105, 114-125.
- MAINGUENEAU, D., (1994), *L'énonciation en linguistique française*, Paris, Hachette.
- MAINGUENEAU, D., (1995), « Présentation », *Les analyses du discours en France, Langages*, 117, 5-12.
- MAINGUENEAU, D., (1997), *L'Analyse du discours*, Paris, Hachette.
- MAINGUENEAU, D., (1998), *Analyser les textes de communication*, Paris, Dunod.
- MAINGUENEAU, D., (2002), « Problèmes d'ethos », *Pratiques*, 113-114, 55-67.
- MAINGUENEAU, D., et COSSUTTA, F., (1995), « L'Analyse des discours constituants », *Langages*, 117, 112-125.
- MALRIEU, D., et RASTIER, F., (2001), « Genres et variations morpho-syntaxiques », *Traitement Automatique des Langues*, 42 (2), 548-577. Disponible en ligne sur *Texto !* à l'adresse http://www.revue-texto.net/Inedits/Malrieu_Rastier/Malrieu-Rastier_Genres.html.
- MARTIN, R., (1983), *Pour une logique du sens*, Paris, Presses Universitaires de France.
- MAURANEN, A., (1993), « Contrastive ESP rhetoric : metatext in Finnish-English economics texts », *English for Specific Purposes*, 18 (1), 3-26.
- MAYAFFRE, D., (2000), *Le poids des mots. Le discours de gauche et de droite dans l'entre-deux-guerres. Maurice Thorez, Léon Blum, Pierre-Etienne Flandin et André Tardieu (1928-1939)*, Paris, Honoré Champion.
- MAYAFFRE, D., (2002), « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus*, 1, Corpus et recherches linguistiques, Université de Nice. Disponible en ligne à l'adresse <http://revel.unice.fr/corpus/document.html?id=11>
- MEAD, G.H., (1934), *Mind, Self, and Society*, Chicago, University of Chicago.
- MEUNIER, A., (1974), « Modalités et communication », *Langue française*, 21, 8-25.
- MEZAILLE, T., (2003), « Thématiques littéraires : enseignement des textes numériques ». Disponible en ligne sur *Texto !* à l'adresse http://www.revue-texto.net/Inedits/Mezaille_Thematiques.html
- MOIRAND, S., (1988), *Une histoire de discours... Une analyse des discours de la revue Le français dans le monde 1961-1981*, Paris, Hachette.

- MOIRAND, S., (1990), « Décrire des discours produits dans des situations professionnelles », in Beacco, J.C., et Lehmann, D., (eds), *Publics spécifiques et communication spécialisée, Le Français dans le monde*, Numéro spécial, Paris, Hachette, 52-62.
- MOLINIE, G., et CAHNE, P., (1994), *Qu'est-ce que le style ?* Paris, Presses Universitaires de France.
- MONIERE, D., et LABBE, D., (2002), « Essai de stylistique quantitative. Duplessis, Bourassa et Lévesque ». Disponible en ligne à l'adresse <http://www.upmf-grenoble.fr/cerat/Recherche/PagesPerso/LabbeMoniere.pdf>
- MORTUREUX, M.F., (1997) *La lexicologie entre langue et discours*, Paris, Editions Sedes.
- MULLER, C., (1968), *Étude de statistique lexicale, le vocabulaire du théâtre de P. Corneille*. Paris, Larousse.
- NØLKE H., (1994), « La dilution linguistique des responsabilités : essai de description polyphonique des marqueurs évidentiels *il semble que* et *il paraît que* », *Langue française* 102, 84-94.
- NØLKE, H., (2001), « La ScaPoLine 2001 Version révisée de la théorie Scandinave de la Polyphonie Linguistique », *Polyphonie - linguistique et littéraire*, 3, 43-65. Disponible en ligne sur http://www.hum.au.dk/romansk/polyfoni/Polyphonie_III/Henning_Nolke.htm
- NØLKE, H., FLØTTUM, K., et NORÉN, C., (2004), *ScaPoLine. La théorie scandinave de la polyphonie linguistique*, Paris, Éditions Kimé.
- NØLKE, H., et OLSEN, M., (2000), « Polyphonie : théorie et terminologie », *Polyphonie – linguistique et littéraire*, 2. Disponible en ligne à l'adresse http://www.hum.au.dk/romansk/polyfoni/Polyphonie_II/poly2_NolkeOlsen_article.htm
- NONNON, E., (1998), « La notion de point de vue dans le discours », *Pratiques*, 100, Metz, CRESEF, 99-123.
- NONNON, E., (2002), « Formulations de problématiques et mouvements de problématisation dans les textes réflexifs. Un point aveugle pour l'enseignant ? », *Spirale*, 29, 29-69.
- NOREN, C., (2000), « Remarques sur la notion de point de vue », *Polyphonie – linguistique et littéraire*, 2. Disponible en ligne à l'adresse http://www.hum.au.dk/romansk/polyfoni/Polyphonie_II/poly2_CocoNoren.htm
- NUCHEZE, V. de, (1998), « Approche pragmatico-énonciative du discours de recherche (à l'usage des apprentis-chercheurs), *Lidil*, 17, 38-47.
- OLSEN, M., (2002), « Remarques sur le dialogisme et la polyphonie », *Polyphonie – linguistique et littéraire*, 6, Numéro spécial. Disponible en ligne sur <http://www.hum.au.dk/romansk/polyfoni/fraind.htm>
- OLSON, D.R., (1998), *L'Univers de l'écrit. Comment la culture écrite donne forme à la pensée*, Paris, Editions Retz.
- OUELLET, P., (1989), « 'Par exemple...' : statut cognitif et portée argumentative de l'exemplification dans les sciences du langage », in Reicher-Béguelin, M.J., (ed), *Perspectives méthodologiques et épistémologiques en sciences du langage, Actes du Colloque de Fribourg 11-12 mars 1988*, Berne, Editions Peter Lang, 95-114.
- PALM, R., (2002), « Utilisation du *Bootstrap* pour les problèmes statistiques liés à l'estimation des paramètres », *BASE*, 6 (3), Gembloux, Presses agronomiques de Gembloux. Disponible en ligne à l'adresse <http://www.bib.fsagx.ac.be/library/base/text/v6n3/143.pdf>
- PALTRIDGE, B., (1997), *Genres, frames and writing in research settings*, Amsterdam, Philadelphia, John Benjamins.
- PECHEUX M., (1969), *Analyse automatique du discours*, Paris, Dunod.
- PECHEUX, M., (1975), *Les vérités de la Palice*, Paris, Maspero.

-
- PEIRCE**, C. S., (1978), *Écrits sur le signe* (rassemblés, traduits et commentés par Deledalle, G.), Paris, Editions du Seuil.
- PERROT**, M., et **SOUDIERE**, M. de la, (eds), (1994), *L'écriture des Sciences de l'homme, Communications*, 58, Paris, Editions du Seuil.
- PERY-WOODLEY**, M.P., (1995), « Quels corpus pour quels traitements automatiques ? », *Traitement Automatique des Langues*, 36 (1-2), 213-232.
- PETITJEAN**, A., (1989), « Les typologiques textuelles », *Pratiques*, 62, 86-125.
- PEYTARD**, J., (1970), « Oral et scriptural : deux ordres de situation et de description linguistiques », *Langue Française*, 6, 35-47.
- PHAL**, A., (ed), (1971), *Vocabulaire général d'orientation scientifique : la part du lexique commun dans l'expression scientifique*, Paris, CREDIF.
- PHILIPPE**, G., (2002), « L'appareil formel de l'effacement énonciatif et la pragmatique des textes sans locuteur » in Amossy, R., (ed), *Pragmatique et analyse des textes*, Tel-Aviv, Université de Tel-Aviv, 17-34.
- PIEGAY-GROS**, N., (1996), *Introduction à l'intertextualité*, Paris, Dunod.
- PINON**, L., et **BARBEROUSSE**, A., (eds), (2003), *Ecriture scientifique, Genesis*, 20, Paris, Editions Jean-Michel Place.
- POLLET**, M.C., **BOCH**, F., (eds), (2002), *L'écrit dans l'enseignement supérieur t. 1 et 2, Enjeux*, 53 et 54, Namur, CEDOCEF.
- POLLET**, M.C., et **PIETTE**, V., (2002), Citation, reformulation du discours d'autrui. Une clé pour enseigner l'écriture de recherche ?, *Spirale*, 29, 165-179.
- PONTILLE**, D., (2001), « L'auteur scientifique en question : pratiques en psychologie et en sciences biomédicales », *Social Science Information*, 40, 433-453.
- PONTILLE**, D., (2004), *La signature scientifique. Une sociologie pragmatique de l'attribution*, Paris, CNRS Editions.
- PONTILLE**, D., (2006), « Qu'est-ce qu'un auteur scientifique ? », *Sciences de la Société*, 67, 76-93.
- POUDAT**, C., (2003), "Characterization of French linguistic research papers with morphosyntactic variables" in Fløttum, F., et Rastier, F., (eds), *Academic discourse. Multidisciplinary Approaches*, Oslo, Novus Press, 77-96.
- POUDAT**, C., (2006a), *Etude contrastive de l'article scientifique de revue linguistique*, Thèse de doctorat de Sciences du langage, sous la direction de Bergounioux, G., Université d'Orléans.
- POUDAT**, C., (2006b), « Typologie des concepts de linguistique : évaluation et élaboration en corpus de critères discriminants », in Duteil-Mongel, C., et Foulquié B., (eds), *Corpus en Lettres et Sciences sociales. Des documents numériques à l'interprétation, Actes de colloque*, Albi, 10-14 juillet 2006. Disponible en ligne sur *Texto!* à l'adresse http://www.revue-texto.net/Archives/Albi_2006/Actes.html
- POUDAT**, C. et **LOISEAU**, S., (2005), "Authorial presence in academic genres", *Studies in Corpus Linguistics*, 19, Amsterdam - Philadelphie, John Benjamins, 51-68.
- POUDAT**, C., et **RINCK**, F., (2006), « Contrastes internes et variations stylistiques du genre de l'article en linguistique », *Journées d'Analyse des données textuelles, JADT06*, avril 2006, Université de Besançon.
- POUDAT**, C., et **RINCK**, F., (à paraître), « Genres scientifiques et styles d'auteur : des variations stylistiques de l'article de revue linguistique », *Actes des Journées linguistique de corpus*, septembre 2005, Lorient, Université de Bretagne Sud.
- POUILLON**, J., (1946), *Temps et roman*, Paris, Gallimard.
- QUIVY**, R., et **VAN CANPENDHOUDT**, L., (1998), *Manuel de recherches en sciences sociales*, Paris, Dunod.

- RABATEL, A.**, (ed), (2004), *Effacement énonciatif et discours rapportés*, *Langages*, 156, Paris, Larousse.
- RABATEL, A.**, (1997), *Une histoire du point de vue*, *Recherches textuelles*, 2, Metz, Université de Metz.
- RABATEL, A.**, (1998), *La construction textuelle du point de vue*, Lausanne, Delachaux et Niestlé.
- RABATEL, A.**, (2002), « Le sous-énonciateur dans les montages citationnels », *Enjeux*, 54, 52-66.
- RABATEL, A.**, (2003), « L'effacement énonciatif et ses effets pragmatiques de sous- et de surénonciation », *Estudios de Lengua y Literatura francesas*, 14, Cadix, Université de Cadix, 33-61.
- RABATEL, A.**, (2004a), « L'effacement énonciatif dans les discours rapportés et ses effets pragmatiques », *Langages*, 156, 3-17.
- RABATEL, A.**, (2004b), « Stratégies d'effacement énonciatif et posture de surénonciation dans le dictionnaire philosophique de Comte-Sponville », *Langages*, 156, 18-33.
- RABATEL, A.**, (2004c), « Effacement argumentatif et effets argumentatifs indirects dans l'incipit du *Mort qu'il faut* de Semprum », *Semen*, 17, 111-132.
- RASTIER, F.**, (1987), *Sémantique interprétative*, Paris, Presses Universitaires de France.
- RASTIER, F.**, (1989), *Sens et textualité*, Paris, Hachette.
- RASTIER, F.**, (1996), « Pour une sémantique des textes. Question d'épistémologie », in Rastier, F., (ed), *Sens et Textes*, Paris, Didier, 9-38.
- RASTIER, F.**, (2001), *Arts et sciences du texte*, Paris, Presses Universitaires de France.
- RASTIER, F.**, (2002), « Enjeux épistémologiques de la linguistique de corpus ». Disponible en ligne sur *Texto!* à l'adresse http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html
- RASTIER, F.**, et Pincemin, B., (1999), « Des genres à l'intertexte », *Cahiers de praxématique*, 33, 83-111.
- REUTER, Y.**, (1998), « De quelques obstacles à l'écriture de recherche », *Lidil*, 17, 11-24.
- REY-DEBOVE, J.**, (1978), *Le métalangage*, Paris, Le Robert.
- RICOEUR, P.**, (1990), *Soi-même comme un autre*, Paris, Editions du Seuil.
- RIEGEL, M.**, **PELLAT, J.C.**, et **RIOUL, R.**, (1994), *Grammaire méthodique du français*, Paris, Presses Universitaires de France.
- RINCK, F.**, (2006), « Ecrire au nom de la science et de sa discipline : la figure de l'auteur dans l'article en sciences humaines », *Sciences de la Société*, 67, 94-111.
- RINCK, F.**, **BOCH, F.** et **GROSSMANN, F.**, (à paraître), *Quelques lieux de variation du positionnement énonciatif dans l'article de recherche*, Ouvrage d'hommages.
- RINCK, F.**, **TUTIN, A.**, et **HERMANN, C.**, (2004), « Analyse et annotation des passages entre guillemets dans des écrits scientifiques », *La construction du savoir scientifique dans la langue*, Workshop TCAN (CNRS), octobre 2004, Grenoble.
- ROCHE, J.**, (1971), *Le style des candidats à la Présidence de la République 1965, 1969*, Toulouse, Privat.
- ROSIER, L.**, (2002), « *Je n'ai jamais appris à citer*. Gestion de la « polyphonie » dans certaines écritures scientifiques », *Enjeux*, 54, 11-27.
- ROUANET, H.**, **LE ROUX, B.**, et **BERT, M.C.**, (1987), *Statistique en sciences humaines : procédures naturelles*, Paris, Dunod.
- SANSOT, P.**, (1994), « Le goût de l'écriture : une dérive épistémologique ? », *Communications*, 58, 61-67.
- SAUSSURE, F.** de, (1995), *Cours de linguistique générale*, édité par Bally, C. et Sechehaye, A. avec la collaboration de Riedlinger, A., Lausanne-Paris, Payot (édition originale 1916).

-
- SAUSSURE, F. de, (2001), *Ecrits de linguistique générale*, édités par Bouquet, S. et Engler, R., Paris, Gallimard.
- SCHLICHTER, H., (1999), *Trois imitations vocales comparées à leur modèle : aspects phonétiques et acoustiques*, Thèse de Doctorat en Sciences du Langage, Université Strasbourg 2.
- SCHNEDECKER, C., (2002), *Lire, comprendre, rédiger des textes théoriques*, Bruxelles, De Boeck Université.
- SCHNEUWLY, B., (1994), « Genres et types de discours : considérations psychologiques et ontogénétiques », in Reuter, Y., (ed), *Les interactions lecture-écriture*, Actes du colloque Théodile-Crel, Peter Lang, 155-173.
- SCHNEUWLY, B., (2001), Les genres de textes – objets d’enseignement de la production orale et écrite. In Garcia-Debanc, C., Confais, J.P., et Grandaty, M., (eds), *Quelles grammaires enseigner à l’école et au collège ? Discours, genres, texte, phrase*, Paris, Delgrave, 83-94.
- SHAPIN, S., et SCHAEFFER, S., (1985), *Leviathan and the Air Pump : Hobbes, Locke, and the experimental life*, Princeton, Princeton University Press.
- SIBLOT, P., (2001), « De la dénomination à la nomination. Les dynamiques de la signifiante nominale et le propre du nom », *Cahiers de Praxématique*, 36, 189-214.
- SIMONIN GRUMBACH J., (1975), « Pour une typologie des discours », in Kristeva J., Milner, J.-C., et Ruwet, N., (eds), *Langue, discours, société : pour Émile Benveniste*, Paris, Editions du Seuil, 85-121.
- SLATKA, D., (1975), « L’ordre du texte », *Etudes de linguistique appliquée*, 19, 30-42.
- STEVENS, S., (1959), « Measurement, psychophysics, and utility”, in Churchman, W. C. et Ratoosh, P., (eds), *Measurement Definitions and Theories*, Chapter 2, New York, Wiley, 18-37.
- SVARTVIK, J., (ed), (1990), *The London Corpus of Spoken English : Description and Research. Lund Studies in English*, 82, Lund, Lund University Press.
- SWALES, J., (1990), *Genre analysis. English in academic and research settings*, Cambridge, Cambridge University Press.
- THOMPSON, S.E., (2003), “Text structuring metadiscourse, intonation and the signalling of organisation in academic lectures”, *Journal of english for academic purposes*, 2, 5-20.
- TODOROV, T., (1978), *Les genres du discours*, Paris, Editions du Seuil.
- TODOROV, T., (1981), *Mikhail Bakhtine. Le principe dialogique*, Paris, Editions du Seuil.
- TUKEY, J.W., (1977), *Exploratory Data Analysis*, Reading, AddisonWesley.
- TUKIA, M., (1983), « Observations sur le vocabulaire, sur les marques d’énonciateur, et sur la construction dans le discours scientifique », *Etudes de linguistique appliquée*, 51, 34-44.
- TUTIN, A., (2005), « Collocations du lexique transdisciplinaire des écrits scientifiques : annotation et extraction des propriétés syntaxiques », *La phraséologie dans tous ses états, Colloque interdisciplinaire Phraséologie 2005*, octobre 2005, Louvain-la-Neuve.
- VALLI, A. et VERONIS, J., (1999), « Etiquetage grammatical des corpus de parole : problèmes et perspectives, *Revue Française de Linguistique Appliquée*, 4 (2), 113-133.
- VASSILEVA, I., (1998), “Who am I/who are we in academic writing? A contrastive analysis of authorial presence in English, German, French, Russian and Bulgarian”, *International Journal of applied linguistics*, 8 (2), 163-190.
- VERONIS, J., (2000), « Alignement de corpus multilingues », in Pierrel, J.M., (ed), *Ingénierie des langues*, Paris, Editions Hermès, 151-171.
- VION, R., (ed), (1998), *Les sujets et leurs discours. Énonciation et interaction*, Aix-en-Provence, Presses de l’Université de Provence.

- VION, R., (2001), « 'Effacement énonciatif' et stratégies discursives », in De Mattia, M. et Joly, A., (eds), *De la syntaxe à la narratologie énonciative*, Gap, Orphrys, 331-354.
- VION, R., (2004), « Modalités, modalisations et discours représentés », *Langages*, 156, 96-110.
- VISOZ, M., (1989), *La part du lecteur*, mémoire de DEA, sous la direction de Dabène, M., Grenoble, Université Stendhal.
- VOLD, E.T., (2005), "Expressions of uncertainty in scientific research articles", *Akademisk Prosa*, 3, 113-127.
- VOLD, E.T., (2006), "Epistemic modality markers in research articles : a cross-linguistic and cross-disciplinary study", *International Journal of Applied Linguistics*, 16 (1), 61-87.
- VYGOTSKY, L.S., (1934), *Pensée et langage*, Paris, Editions sociales (1985 pour la traduction française).
- WEINRICH, H., (1973), *Le temps*, Paris, Seuil.
- WITTGENSTEIN, L., (1961), *Tractatus logico-philosophicus*, suivi de *Investigations philosophiques*, Paris, Gallimard.

Groupes, projets de recherche, et ressources en ligne

- Cediscor, *Centre de recherches sur les discours ordinaires et spécialisés*. Université Paris 3. En ligne à l'adresse <http://www.univ-paris12.fr/www/labos/ceditec/cediscor.html>
- Ci-Dit, *Groupe de recherches International sur la circulation des discours*. En ligne à l'adresse <http://www.ulb.ac.be/philo/serlifra/ci-dit>
- KIAP, *Cultural identity in academic prose*, projet piloté par K. Fløttum (ed), University of Bergen, Romansk Institutt, 2002-2006. <http://www.uib.no/kiap>
- PolyphoniE, site dédié aux notions de polyphonie et de dialogisme, P. Dendale, Université d'Anvers - Ceted, Metz, avec la collaboration de M. Kara et S. Leroy, 2003-2006. En ligne à l'adresse <http://webhost.ua.ac.be/polyphon>
- Projet PPF (Plan Pluri-Formations), Développement de ressources pour la didactique du français à l'aide d'outils de TAL : étude des marqueurs linguistiques de la subjectivité, piloté par G. Antoniadis et F. Grossmann, Université Stendhal Grenoble 3, Laboratoire Lidilem, 2003-2006. En ligne à l'adresse <http://ppfesc.free.fr> (site en construction).
- Projet Emergence, *l'Acquisition semi-automatique de traductions d'expressions semi-figées pour l'élaboration d'outils d'aide à la rédaction scientifique et technique*, piloté par O. Kraif et A. Tutin, Université Stendhal Grenoble 3, Laboratoire Lidilem, 2004.
- Texto ! *Textes et cultures*, revue électronique, Equipe sémantique des textes, sous la direction de F. Rastier, 1996-2006. En ligne à l'adresse <http://www.revue-texto.net>

Corpus et outils

- TEXT ENCODING INITIATIVE. En ligne à l'adresse <http://www.tei-c.org/>
- DTM. Logiciel de Data and Text Mining, développé par L. Lebart. En ligne sur <http://www.enst.fr/egsh/lebart/> (rubrique Logiciel).
- STAVIEW. Version 5.0, SAS Institue INC., 1992-1998. « Manuel de référence » (1996) et « Addendum Statview 5.0 pour la version française » (SAS Institue INC. et Alsyl, 1998).

-
- CORDIAL PROFESSIONNEL.** Version 9.0. Correcteur et analyseur linguistique, Synapse Développement, 1994-1998. « Manuel utilisateur de Cordial Analyseur » (2002).
- FRANTEXT.** Outil de consultation de ressources informatisées sur la langue française, Atilf - CNRS, Université de Nancy 2. En ligne sur <http://www.atilf.fr/frantext.htm>
- WCONCORD.** Concordancier, Version 2.0., Z. Martinek, avec la collaboration de L. Siegrist, Technische Hochschule Darmstadt, 1996.
- TEXT PAD.** Editeur de texte, Version 4.7.3., Helios Software Solutions, 1992-2004. En ligne à l'adresse <http://www.textpad.com/>
- LOB.** The Lancaster-Oslo/Bergen Corpus, S. Johansson et G. N. Leech, University of Oslo - University of Lancaster, compilé en 1970-1978.
- LLC.** The London-Lund Corpus of Spoken English, J. Svartvik, Lund University, compilé en 1975-81 et 1985-88 (*cf.* Svartvik, 1990).

Index

Les entrées retenues correspondent aux principaux traits linguistiques analysés dans le genre, et à l'usage qui en est fait en fonction des disciplines et du statut de l'auteur (cf. chapitres 4, 5 et 7) et par chaque auteur en fonction de son style (cf. chapitre 6). Par exemple, « phrase » renvoie aux occurrences relatives à la caractérisation de la phrase dans l'article (problèmes méthodologiques, nombre de phrases et longueur en nombre de mots, ponctuations).

Annexes : 108, 118, 137-139, 146-150, 157-158, 184-185, 246, 250.

Argumentation (et dimension argumentative) : 48, 65-66, 82, 92, 110, 226-228, 230, 242-246, 297, 299-301, 303, 307, 319, 344-346, 350, 353, 363, 367, 375, 382, 386.

Autonyme : 88, 181, 204, 218, 220, 232-233, 235, 239, 279, 305, 322-323.

Axiologiques (lexique axiologique) : 57, 85, 225, 236, 256, 272, 304, 319-321, 327, 335, 352, 362-371.

Balises méta-textuelles : 42, 204-208, 262, 282-285, 308, 311, 316, 320, 330.

Bibliographies : 108, 115-116, 118, 122-123, 126-127, 137-138, 146-147, 150-158, 171, 177-178, 184-185, 171, 177-178, 184-185, 190, 246, 331, 338, 345.

Citation : 36, 47-48, 50, 65, 92, 108-109, 121, 126, 130, 149, 172-184, 190, 221-223, 232-240, 250, 279, 304-305, 309, 315-316, 322, 331, 336-341, 347-348, 356.

Commentaire méta-énonciatif (dédoublé méta-énonciatif) : 87, 155, 237-240, 251, 309, 329, 347.

Conclusions : 159-162, 203, 256, 300, 350.

Connecteurs : 164, 200, 204, 218, 270, 278, 279, 281, 283-286, 320, 351.

Deux-points : 215-218, 274, 280-281, 297.

Discours autres et discours représentés : 50-51, 172, 331-332, 339, 371.

Épistémologie : 171, 181, 184-185, 255-258, 275, 366-369.

Exemples : 108-109, 121, 126, 129-130, 134, 136-138, 164, 172-185, 190-191, 194, 219-221, 226, 233-235, 260, 269, 274, 279, 305, 315, 332.

Formalismes : 118, 128, 137-138, 165-171, 184, 279, 296-298, 314.

Futur : 197, 202, 204-208, 283, 286, 296, 298, 308.

Gestion des références : 154-157, 337-338 (voir aussi *Renvoi aux sources*).

Guillemets : 88, 109-110, 128-129, 126, 172, 187, 213, 217, 220, 231-232-240, 251, 305, 341-342, 349, 373.

Il Impersonnel : voir *Impersonnel*.

Imparfait de l'indicatif : 197-201, 208, 277-278, 296.

Impératif : 96, 197, 202, 204-208, 257, 296, 311-312.

Impersonnel (tours impersonnel) : 128, 228-229, 270, 282-284, 296-297, 304, 316-330, 332-333, 345, 351, 366, 370, 373.

Introductions : 41, 108, 159-162, 171, 182, 184-185, 205, 226, 324, 339, 352-357, 360, 362, 365, 368-369.

Je : voir *Marques de personne et Pronoms personnels*.

Lexique transdisciplinaire : 128, 130, 133-135, 188, 232, 240-262, 270-273, 283, 285-291, 364, 367-369, 373.

Marques de personne : 85, 91, 135, 187-188, 196, 208-213, 262, 283, 285, 304-317, 322-330, 351.

Mise en intrigue de la découverte : 227, 229-231, 310, 312, 316, 374.

Modaux : 117, 270, 284, 296-297, 320.

Mode narratif : 117, 199-202, 208, 278, 296-299, 389.

Modes de désignation (des points de vue et de l'existant) : 129, 325, 221-332, 342-352, 357, 371.

Mots-outils, mots signifiants : 121, 185, 193-195, 270, 274-276.

Négation : 42, 59-61, 65-67, 77, 85-86, 92, 226-231, 238, 257, 283, 284-286, 319, 321, 351, 361, 370.

Notes (de bas de page ou de fin) : 105, 108, 123, 126-127, 130, 134-142, 145-146, 156-158, 162, 168, 171, 178, 184, 190-193, 206, 219, 246, 331, 334-338, 340, 365, 375.

Nous : voir *Marques de personne et Pronoms personnels*.

Objectifs de l'article : 134, 304, 350, 352-362, 370.

On : 212-213, 225, 255, 282, 296-297, 316, 321-330, 354, 359.

Opposition : 66, 87, 118, 285, 350-351, 359-360.

Paratexte : 147-148.

Parenthèses : 213, 215-223, 255-256, 296, 305, 309, 336, 340.

Passé composé : 197, 200-203, 208, 309.

Passé simple : 197-203, 208, 277-278, 296, 369.

Passif : 78, 112, 322, 325-326, 328-330, 345, 366.

Péritexte : 124, 146-158, 163, 171, 183-184, 336-338, 375.

Phrase : 126, 130, 187-196, 213-218, 220, 223-224, 262, 270, 280, 282-283, 291.

Plans de texte : voir *Structure des textes*.

Points d'exclamation : 215-217, 220-226, 230-231, 274, 373.

Points d'interrogation : 107, 166, 168, 215-217, 220, 222, 226-227, 230, 274, 281-283, 296.

Points de suspension : 187, 215-218, 220-224, 280.

Points-virgules : 213, 215-218, 280.

Polémique : 42, 65-66, 222-225, 229, 262, 283-285, 262, 283-285, 291, 329-330, 346, 351-352, 364, 369-371, 375.

Possessifs : 208-212, 267, 271, 293, 299, 302, 321.

Présent de l'indicatif : 142, 196-204, 108, 277, 284, 296-298, 328, 389.

Pronoms personnels : 57-58, 63, 66, 112, 208-212, 255, 282, 302.

Question rhétorique : 226-231, 281, 312, 328, 371, 374 (voir aussi *Points d'interrogation*).

Reader-friendly (principes reader-friendly) : 41, 162, 171, 184, 206-297, 219, 375.

Références à l'existant : 109, 331-352, 355-357, 371-373.

Renvoi aux sources : 152-156, 175, 190, 218-220, 331-332, 336-352, 354, 376 (voir aussi *Gestion des références*).

Résumé : 41, 108, 119, 138, 147-148.

Sections et sous-sections : 57, 158-167, 170.

Signature : 100, 103, 147-149, 157.

Structure des textes : 41, 137, 158-165, 219-221, 270-271, 278-279, 291, 296-298, 373.

Subjonctif : 197, 201, 286.

Subordonnées : 192-195, 218, 281.

Surénonciation : 81-82, 90, 231, 370-371, 376.

Temps verbaux : 94, 112, 117, 123, 135, 187-188, 196-208, 262, 277, 290-291, 296, 310-311.

Tirets : 164-166, 170, 215-216, 221-223, 229, 231, 279.

Titres et sous-titres : 107-118, 146-148, 156-159, 161-163, 184, 230-231, 236-239, 350

Virgules : 213, 215-217, 280-281, 290.

Volume de l'article : 123, 127, 137-148, 150-154, 157, 168.

