



**HAL**  
open science

## Lexique et discours

Mathilde Dagnat

► **To cite this version:**

Mathilde Dagnat. Lexique et discours : Thèse d'HDR. Linguistique. Université Paris 8 - Saint-Denis; Laboratoire Structures Formelles du Langage (UMR 7023), 2023. tel-04452684

**HAL Id: tel-04452684**

**<https://hal.science/tel-04452684>**

Submitted on 15 Feb 2024

**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.

Copyright

**Université Paris 8 Vincennes – Saint-Denis**

École Doctorale Cognition, Langage, Interaction

Habilitation à Diriger des Recherches

# Lexique & discours

Mathilde Dargnat

## **Document II – Synthèse des travaux et perspectives**

<version 1.2 , 01/03/2023>

Garante : Mme **Claire Beyssade**

Professeure à l'Université Paris 8 Vincennes – Saint-Denis & SFL-CNRS (UMR 7023)

Habilitation soutenue publiquement le 17 février 2023,  
devant un jury composé de :

**Claire Beyssade**, Professeure, Université Paris 8 Vincennes – Saint-Denis & SFL, garante  
**Bernard Combettes**, Professeur émérite, Université de Lorraine & ATILE, examinateur  
**Maj-Britt Mosegaard Hansen**, Professeure, Université de Manchester, rapporteure, examinatrice  
**Fiammetta Namer**, Professeur, Université de Lorraine & ATILE, examinatrice  
**Adeline Nazarenko**, Professeure, Université Paris 13 & LIPN, examinatrice, présidente  
**Richard Waltereit**, Professeur, Université Humboldt, Berlin, examinateur

(Pré-)rapporteurs sur le dossier avant soutenance :

Mylène Blasco, Professeure, Université de Clermont-Auvergne & LRL, pré-rapporteure  
Jonathan Ginzburg, Professeur, Université Paris-Cité & LLF, pré-rapporteur  
Mathieu Valette, Professeur, INALCO & ERTIM, rapporteur



*To tell a believable story, the novelist knew,  
you just have to get the details right.*

*A Widow for one year*, John Irving, Black Swan, 2010 [1998], p. 367.



# Remerciements

Je remercie :

– Les professeurs, chercheurs et ingénieurs qui ont orienté, suivi ou évalué mes premiers pas dans la recherche, depuis la maîtrise jusqu’au doctorat : Mylène Blasco et Jacques-Philippe Saint-Gérand à l’Université Blaise Pascal, Albert Di Cristo, Marie-Christine Hazaël-Massieux, Yannick Resch et Fridrun Rinner à l’Université de Provence, Gilbert David, Lise Gauvin, Wladimir Krysinski, Walter Moser, Pierre Neveu et Pierrette Thibault à l’Université de Montréal, Françoise Gadet à l’Université Paris Nanterre, Charles Bernet et Serge Heiden à l’École Normale Supérieure de Lyon.

– Les collègues avec qui j’ai eu l’occasion de collaborer depuis ma prise de poste comme Maître de conférences à Nancy, tout particulièrement Katarina Bartkova, Bernard Combettes, Denis Jouvét et Yannick Toussaint. Outre leur sympathie à toute épreuve, ils ont su faire évoluer mes recherches et m’apprennent toujours beaucoup.

– Les collègues qui m’ont intégrée et accueillie dans leurs projets scientifiques et/ou leur projet de laboratoire, en particulier Anne Abeillé, Laurence Danlos, Danièle Godard et Jean-Marie Marandin à l’Université Paris 7, Éva Buchi, Mathieu Constant et Jean-Marie Pierrel à l’ATILE, Christophe Cerisara et Claire Gardent au LORIA, Georgeta Cislaru à l’Université Paris 3, Denis Jamet à l’Université Lyon 3 et Agnès Tutin à l’Université Grenoble 3.

– Les collègues qui ont bien voulu s’embarquer avec moi dans le projet [CODIM](#). Ce n’est que le début.

– Les collègues de l’UFR Arts, Lettres et Langues et de l’UFR Sciences du Langage de l’Université de Lorraine avec qui nous partageons nos illusions et désillusions d’enseignants-chercheurs.

– La garante de cette habilitation, Claire Beyssade, pour sa disponibilité, ses conseils avisés et bienveillants et sa grande grande patience.

– Les pré-rapporteurs pour avoir autorisé l’inscription et la soutenance.

– Les membres du jury pour avoir accepté de prendre du temps pour lire cette synthèse, peut-être pas toujours synthétique.

– Mon mari Jacques, ma fille Laure, mes parents Irène et Jean-Claude, ma sœur Aurélie et mon frère Guillaume, pour les moments où leur présence a été un soutien. Certaines périodes ont été difficiles. Pour tout le monde.  
Comme il m’arrive souvent de le dire : *je me comprends*.

– Et les personnes que j’aurais oubliées et qui me supportent pourtant, dans les deux sens.



# Table des matières

<b>Introduction</b> .....	11
Lexique et discours ?.....	11
Parcours et motivation.....	12
Présentation et organisation des documents.....	13
<b>1 L’oral représenté, l’exemple de Michel Tremblay</b> .....	15
<b>1.1 Linguistique et littérature</b> .....	17
1.1.1 Communication littéraire et ressources contextuelles.....	18
1.1.1.1 Stratification énonciative des textes littéraires.....	19
1.1.1.2 Association de ressources situationnelles et socio-culturelles.....	19
1.1.1.3 Dimension cognitive de l’interprétation.....	19
1.1.2 Structure narrative et polyphonique des textes de fiction.....	20
1.1.2.1 Personnages, oralité et parcours narratifs.....	21
1.1.2.2 Structure énonciative et polyphonie.....	22
1.1.3 Genre et style.....	25
1.1.3.1 Propriétés et contraintes du genre dramatique.....	25
1.1.3.2 La place du style.....	27
<b>1.2 Variation linguistique et contraintes de représentation</b> .....	27
1.2.1 Catégorisation linguistique et symbolique de l’oralité populaire québécoise	28
1.2.1.1 Un complexe variationnel comme référent.....	28
1.2.1.2 La question du <i>joual</i> .....	31
1.2.2 Représenter et décoder l’OPQ : la sensation d’un travail.....	33
<b>1.3 Méthodologie de corpus : donner du sens au quantitatif</b> .....	36
1.3.1 Un « sac » de données construit.....	37
1.3.1.1 Constitution des corpus <i>tremblay</i> et <i>fcapop</i> .....	37
1.3.1.2 Traitement XML et textométrie.....	39
1.3.2 L’effet personnage-locuteur et le calcul des profils linguistiques.....	40
<b>Synthèse 1</b> .....	45
<b>2 Relations de discours</b> .....	47
<b>2.1 Constructions juxtaposées</b> .....	50
2.1.1 Juxtapositions propositionnelles : propriétés générales.....	50
2.1.1.1 Syntaxe : dépendance et catégorisation.....	51
2.1.1.2 Sémantique : interprétation et (non-)compositionnalité.....	55
2.1.1.3 Contraintes sur l’enchaînement discursif.....	59
2.1.2 Juxtapositions conditionnelles : contraintes et constructionnalité.....	63
2.1.2.1 La relation de condition.....	63
2.1.2.2 Héritage et spécification.....	64
2.1.2.3 Enchaînement discursif.....	66
2.1.2.4 Valeurs modales et non-compositionnalité.....	66
2.1.3 Connexion discursive et prosodie.....	69



2.1.3.1 Contour montant.....	69
2.1.3.2 Expériences.....	71
<b>2.2 Relief argumentatif.....</b>	<b>74</b>
2.2.1 Quelques paramètres de l'argumentation.....	75
2.2.1.1 Repères.....	75
2.2.1.2 Ouvertures.....	81
2.2.2 Analyses et représentations : vers le TAL.....	83
2.2.2.1 Discours, argumentation et représentation de la structure textuelle...	83
2.2.2.2 Expériences.....	89
<b>Synthèse 2.....</b>	<b>97</b>
<b>3 Marqueurs discursifs.....</b>	<b>99</b>
<b>3.1 Que marquent les marqueurs discursifs (MD) ?.....</b>	<b>102</b>
3.1.1 Des repères dans la « jungle ».....	102
3.1.1.1 Multiplicité des approches et des étiquettes.....	102
3.1.1.2 Les propriétés pertinentes pour les MD.....	105
3.1.2 Indexicalité et dépendance contextuelle.....	106
3.1.3 Composante procédurale et fonctions de base.....	107
3.1.4 Portée.....	111
3.1.5 Intégration des MD.....	114
3.1.6 En résumé.....	118
<b>3.2 Travaux sur les connecteurs et les particules énonciatives (PEN).....</b>	<b>119</b>
3.2.1 Cartographie des items et des propriétés.....	119
3.2.1.1 Connecteurs, PEN et compagnie.....	121
3.2.1.2 (Im)possibilité d'une liste.....	126
3.2.2 Quelques aspects de la concession.....	134
3.2.2.1 De la comparaison à la concession.....	134
3.2.2.2 Concession et interaction.....	140
3.2.3 PEN et prosodie.....	142
3.2.3.1 Particule vs non-particule.....	142
3.2.3.2 De la classe aux fonctions pragmatiques.....	146
3.2.4 Combinaison de MD.....	149
3.2.4.1 <i>Coudon(c)</i> : indice de réalisme langagier.....	149
3.2.4.2 <i>Mais enfin</i> : construction et association.....	151
<b>Synthèse 3.....</b>	<b>155</b>
<b>4 Perspectives en guise de conclusion.....</b>	<b>157</b>
<b>4.1 Marqueurs de discours : combinaisons et expérimental.....</b>	<b>159</b>
4.1.1 Compositionnalité des combinaisons de marqueurs : le projet CODIM.....	159
4.1.2 Analyse en oculométrie et en potentiel évoqué de l'enchâssement des marqueurs de cause et de concession.....	161
<b>4.2 Catégorisation et représentation de la variation, suites.....</b>	<b>163</b>
4.2.1 L'oral représenté : théâtre, francophonie et littérature de jeunesse.....	163
4.2.2 Humour et variation linguistique.....	165

<b>5 Références.....</b>	<b>167</b>
5.1 Références des corpus utilisés.....	169
5.2 Références bibliographiques personnelles.....	170
5.3 Références bibliographiques générales.....	174
<b>6 Annexes.....</b>	<b>191</b>
Annexe 1 Structures de traits des constructions juxtaposées.....	193
Annexe 2 Contrastes des valeurs de pente pour l'analyse des contours prosodiques...	195
Annexe 3 Graphes initiaux, RST et argumentation.....	196
Annexe 4 Liste des propriétés des <i>pragmatic markers</i> de Brinton.....	197

## Liste des tableaux

Tableau 1. Présence de néographies phonétisantes dans les textes.....	35
Tableau 2. Liste des variables pour le profilage OPQ.....	43
Tableau 3. Relevé des occurrences des variables OPQ dans <i>L'Impromptu d'Outremont</i> ...	43
Tableau 4. Données des personnages Fernande et Lorraine pour le test de Fisher.....	43
Tableau 5. Résultats du test de Fisher pour le profilage des personnages.....	44
Tableau 6. Configurations et stimuli pour l'analyse des continuations prosodiques....	72
Tableau 7. <i>Arg-micro-texts</i> : propriétés de la redescription 1.....	92
Tableau 8. Liste indicative d'étiquettes désignant les MD.....	103
Tableau 9. Description du connecteur <i>alors, pendant que</i> et <i>quand bien même</i> .....	128
Tableau 10. Principales expressions fonctionnant comme PEN.....	129
Tableau 11. Propriétés prosodiques pour l'analyse des PEN.....	144
Tableau 12. Classification des propriétés prosodiques des PEN <i>alors, ben, bon</i> , etc.....	145
Tableau 13. <i>Alors</i> : taille des corpus, fréquence totale, etc. ....	147
Tableau 14. Relevé de <i>coudonc</i> dans les corpus <i>fricapop</i> et <i>tremblay</i> .....	152

## Liste des figures

Figure 1. Schémas énonciatifs des cinq pièces de Michel Tremblay.....	23
Figure 2. Décalage des registres à l'oral et à l'écrit.....	35
Figure 3. Espaces discursifs des pièces dans le corpus <i>tremblay</i> .....	41
Figure 4. Espaces discursifs des personnages.....	42
Figure 5. Structure de traits de la construction <i>juxta-prop</i> .....	60
Figure 6. Contrainte d'attachement de type subordination discursive.....	61
Figure 7. Contrainte d'attachement de type coordination discursive.....	62
Figure 8. Spécification des attributs CONSA, CONSB, PAUSE de <i>juxta-cond</i> .....	65
Figure 9. Spécification de l'attribut CNTXT de la construction <i>juxta-cond</i> .....	66
Figure 10. Exemple de contrainte de la construction <i>juxta-cond</i> .....	66
Figure 11. Les intonations du français (d'après Delattre).....	70
Figure 12. <i>Corpus Recorder</i> , exemple d'écran proposé aux participants .....	72

Figure 13. Exemple de graphe RST.....	84
Figure 14. Schématisation de la structure argumentative de l'exemple (51).....	87
Figure 15. Exemples de configuration de type Support et Attack.....	87
Figure 16. Schématisation de la structure argumentative de l'exemple (52).....	88
Figure 17. Compatibilité des graphes RST et ARG de l'exemple (53).....	90
Figure 18. <i>Arg-micro-texts</i> : sous-graphes de la redescription 1.....	92
Figure 19. Distribution des relations de discours entre propositions dans <i>Sci-DTB</i> .....	94
Figure 20. Sous-catégories des MD, d'après la classification de Dostie.....	120
Figure 21. Segmentation et annotation prosodiques automatiques (ex. <i>vous savez</i> ).....	143
Figure 22. Structures des stimuli pour l'expérience sur <i>quand même</i> et <i>pourtant</i> .....	162

# Symboles et sigles

## • Symboles

### **Sémantique des relations**

Condition : relation de discours de condition

Support : relation argumentative de support

### **Structures de traits**

$x \Leftrightarrow [Y]$  : modification de la valeur de la force illocutoire de  $x$  en fonction de la valeur de  $Y$

$\oplus$  : opérateur de concaténation

/ : contrainte par défaut

Italiques (ex. *mod*, *attach*, *sub*) : type

petites capitales (ex. CONSA, PAUSE) : attributs

### **Exemples**

\* : énoncé agrammatical

? : énoncé difficilement interprétable

% : énoncé dont l'interprétabilité ou la grammaticalité sont variables ou approximatives

# : énoncé interprétable dans des contextes différents du contexte ciblé ou avec une interprétation différente de l'interprétation initiale

$\Rightarrow$  : interprétation préférentielle

$\nRightarrow$  : n'implique pas l'interprétation mentionnée

## • Sigles et acronymes

CxG	Construction Grammar
HPSG	Head-driven Phrase Structure Grammar
MD	Marqueur(s) Discursif(s)
OPQ	Oralité Populaire Québécoise
PEN	Particule(s) ÉNonciative(s)
QUD	Question Under Discussion
RST	Rhetorical Structure Theory
SDRT	Segmented Discourse Representation Theory

# Introduction

## Lexique et discours ?

J'ai souhaité intituler ce document *Lexique et discours* pour deux raisons principales, qui correspondent à mon intérêt pour les processus de représentation et d'interprétation et pour celui du rôle de certaines expressions, à la fois déclencheurs et traces de ces processus.

Le discours *représente*, en ce sens qu'il propose une image du contexte où il se produit comme discours. Il peut à ce titre être perçu comme représentant des éléments de sa situation d'énonciation (instances de production et de réception, cotexte, contexte situationnel, y compris idéologique et culturel, états cognitifs, etc.).

Le discours se déploie en utilisant différentes ressources. Le choix du thème, du support (médium), du genre, du code, etc. intervient au niveau général de la communication. Dans mes recherches et dans ce document, je me suis centrée sur certaines expressions et structures de la langue qui jouent un rôle important dans la construction et l'interprétation du discours en contexte. Il s'agit en particulier d'éléments du lexique qui *marquent* l'organisation du discours et l'ancrage énonciatif (notamment le vocabulaire associé au jocal et les marqueurs de discours). Mais la fonction d'organisation est parfois assumée par des configurations (constructions) qui dépassent le cadre du mot tout en relevant comme lui du code linguistique. C'est ce qui m'a amenée à intégrer des travaux sur la juxtaposition (par ex. juxtapositions conditionnelles à l'impératif), des marqueurs discontinus (*si* + adjectif + *que* + proposition au subjonctif) et des combinaisons de marqueurs (les associés de *mais*).

Par ailleurs, l'interprétation du discours déborde l'application du code linguistique, ce qui constitue le principe de base de toute approche dite pragmatique. J'ai rencontré cette question de manière évidente à propos de l'interprétation des marques d'oralité dans des textes littéraires, mais aussi à propos de l'inférence des relations de discours non marquées par un connecteur et à propos de l'intégration de régularités contextuelles dans le processus d'évolution diachronique de connecteurs.

Pour étudier ces questions de manière systématique, j'ai dès le départ fait appel à plusieurs outils théoriques et techniques, que les domaines littéraires et linguistiques sollicitent plus ou moins régulièrement. Du point de vue théorique, j'emprunte des notions et des méthodes à la sociolinguistique, à la stylistique et à l'analyse de l'argumentation. Du point de vue technique, j'ai utilisé à la fois des cadres de représentation (structures de traits, RST et graphes argumentatifs), des outils

d'annotation et d'exploration de données (XML, automates finis), et de l'analyse quantitative (textométrie, agrégation (*clustering*), mesures d'association).

## Parcours et motivations

Je me rends compte rétrospectivement que j'ai toujours été intéressée en même temps par des aspects qualitatifs qui relèvent de l'intuition et des aspects qui relèvent d'une explicitation contrôlée des intuitions. Cela se voit dans ma formation et dans mon parcours de recherche, souvent à l'entre-deux :

- *Domaines* : linguistique, études littéraires, TAL, sciences cognitives.
- *Pays et institutions* : France, Québec, Université de Provence, Université de Montréal, École Normale Supérieure de Lyon, Université de Lorraine, Laboratoire Parole & Langage (LPL, Aix-en-Provence), Laboratoire Interactions, Corpus, Apprentissages et Représentations (ICAR, Lyon), Laboratoire Analyse et Traitement Informatique de la Langue Française (ATILF, Nancy), Laboratoire Lorrain de Recherche en Informatique Appliquée (LORIA, Nancy), Centre d'Études Linguistique (CEL, Lyon), Institut des Sciences Cognitives (ISC) Marc Jannerod (Bron). Voir le CV dans le document I.

Concrètement, j'ai souhaité travailler de cette manière depuis mes tout premiers travaux. J'ai cherché et construit l'environnement pertinent pour pouvoir mener à bien des sujets de recherche que j'ai choisis, sans m'imposer d'adhérer par principe à une école de pensée dominante dans un lieu donné. Cela a eu comme conséquences une assez grande mobilité, avec ses avantages et ses difficultés institutionnelles, mais aussi l'opportunité de développer des collaborations très diverses qui me permettent de prendre de la distance et de concevoir les résultats obtenus de manière nuancée et toujours prospective.

J'ai utilisé ce document de synthèse pour présenter mes travaux et en montrer l'évolution, même si je n'ai pas suivi une chronologie stricte. L'organisation en trois parties correspond globalement à trois moments, mais pas seulement. Elle renvoie aussi à des orientations plus ou moins centrées sur le macro ou le micro fonctionnement du discours. Un slogan commode mais simplifiant serait de dire que le document va du plus discursif au plus lexical, de la sémiotique textuelle (partie 1) à la description d'unités lexicales (partie 3) en passant par l'étude des relations de discours (partie 2). En pratique, certains points se chevauchent et se complètent, par exemple les problématiques de la stylistique et de l'argumentation, ou encore celles des relations de discours et du sens des connecteurs.

La première partie (*L'oral représenté : l'exemple de Michel Tremblay*) présente la manière dont un auteur dramatique, Michel Tremblay, travaille son expression et la structure dans ses textes pour représenter une variété linguistique symboliquement marquée dans la société québécoise. J'utilise des notions et des méthodes de théorie littéraire, de sociolinguistique et de linguistique de corpus pour montrer, d'une part, que son écriture représente bien un état de langue attesté dans des corpus authentiques et, d'autre part, que cette représentation évolue sur une période de trente ans (1968-1998).

La deuxième partie (*Relations de discours*) se concentre sur des travaux autour des relations de discours. Le premier point concerne des études traitant des juxtapositions, en particulier des juxtapositions à interprétation conditionnelle, décrites comme des constructions dans le cadre des Grammaires de construction. Il s'agit de relations de discours qui ne sont pas marquées par un connecteur (ex. *si, pourtant, mais*), mais dont la description met en évidence d'autres types de contraintes (ordre et dépendance des constituants, temps et modes verbaux, contours intonatifs) et interroge la compositionnalité de l'interprétation globale. Le deuxième point aborde des travaux plus récents ouverts sur le traitement automatique du langage (TAL), comparant la structure discursive et la structure argumentative de données génériquement homogènes et annotées à partir de cadre déjà développés (Rhetorical Structure Theory (RST) et relations argumentatives proposées par Freeman).

La troisième partie (*Marqueurs discursifs*) a pour but d'éclaircir quelques caractéristiques des marqueurs de discours (MD) que je trouve toujours complexes. Je revisite certaines notions, en particulier la notion de procédural et d'indexicalité, et j'emprunte à des théories plus générales des propriétés qui contribuent à structurer la catégorie globale des MD (projection, effet local, suspension, déplaçabilité). Cela m'amène à distinguer de manière plus précise les connecteurs et les particules énonciatives (PEN) dans leur fonctionnement prototypique, tout en ménageant la possibilité de décrire des cas hybrides. Je synthétise également des travaux portant sur des marqueurs concessifs, d'un point de vue diachronique et d'un point de vue interactionnel. Une autre dimension a fait l'objet d'un travail d'équipe, il s'agit des propriétés prosodiques des PEN et de leur corrélation avec les différentes valeurs observables dans des corpus diversifiés. Enfin, je me suis lancée plus récemment dans l'analyse quantitative des combinaisons de MD, en lien notamment avec la question de la compositionnalité de ces combinaisons. Cela rejoint des préoccupations mentionnées dans la partie 2 à propos des juxtapositions.

## Présentation et organisation des documents

Outre les diverses pièces administratives, le dossier d'Habilitation à Diriger des Recherches est composé de trois documents.

### \* Contenu des trois documents présentés

Le **document I** est composé d'un curriculum vitae détaillé et d'une liste complète des mes publications et présentations.

Le présent document est le **document II**. Il constitue la synthèse de mes recherches depuis une quinzaine d'années. Chaque partie comporte un renvoi à l'ensemble des travaux concernés par la partie.

Le **document III** est le recueil de la quasi-totalité de mes travaux présentés par ordre chronologique de parution (sauf la thèse et l'ouvrage de 2002). Parmi les 43 travaux répertoriés, 24 sont étiquetés comme représentatifs. Ils sont listés et indiqués par un astérisque dans la liste des mes publications et dans les renvois au début de chaque partie du document II.

## \* **Système de renvoi dans le texte**

### **Renvois à mes travaux**

Les renvois à mes travaux sont indiqués avec :

- Le nom de l’auteur ou des auteurs pour les publications collectives (les deux noms pour deux auteurs, et le nom du premier auteur suivi de *et al.* pour plus de deux auteurs).
- La date de publication.
- Un numéro entre crochets. Il correspond à l’ordre chronologique de mes publications. Pour faciliter leur identification, le numéro est de plus surligné en bleu dans l’impression couleur et possède un lien actif vers la liste complète (section 5.2 Références bibliographiques personnelles, p. 170).
- Dans les listes (CV, début des parties, liste des publications), un astérisque pour désigner qu’il s’agit d’une publication jugée représentative.
- Si cela est pertinent, la page ou l’ensemble de pages, l’exemple, la figure ou le tableau.

Par exemple, Dargnat et al. (2012a [16], p. 1478) renvoie à la page 1478 de la publication n°16 dans la liste de mes travaux et dans le volume III. Les pages sont celles des éditions, pas celles du volume III.

### **Renvois aux autres travaux**

Dans le corps du texte, les renvois aux travaux sont indiqués avec :

- Le nom de l’auteur ou des auteurs pour les publications collectives (les deux noms pour deux auteurs, et le nom du premier auteur suivi de *et al.* pour plus de deux auteurs).
- La date de l’édition et le cas échéant la date de la première publication entre crochets.
- Si cela est pertinent, la page ou l’ensemble de pages, l’exemple, la figure ou le tableau.

Par exemple, Bakhtine M. (Volochnikov) 1977 [1929], p. 28 renvoie à la page 28 du texte de Mikhaïl Bakhtine dans son édition en français de 1977, l’édition originale datant de 1929. Les références complètes sont indiquées dans la section 5.3 Références générales (p. 174).

### **Références des exemples**

Les exemples construits ne comportent pas de référence.

Les exemples empruntés à des auteurs et à des travaux sont référencés comme ci-dessus. Les exemples extraits de corpus sont référencés avec le nom et le code du corpus. La liste des corpus utilisés est indiquée à la section 5.1 Références des corpus utilisés (p. 169).

# 1

## L'oral représenté L'exemple de Michel Tremblay

*Qu'est-ce qu'un livre ? Une suite de petits signes. Rien de plus.  
C'est au lecteur à en tirer lui-même les formes, les couleurs et les  
sentiments auxquels ces signes correspondent.*

Anatole France, *Le Jardin d'Épicure*. Paris : Calmann Lévi, 1895, p. 44.





Cette partie est une synthèse de mes travaux sur l’oral représenté. Bien que circonscrits à un phénomène – l’oralité populaire québécoise (OPQ) – et à des données sélectionnées – des extraits de deux corpus linguistiques de français québécois et cinq pièces de théâtre de l’écrivain Michel Tremblay<sup>1</sup> –, ils ont soulevé des questions de fond sur la variation linguistique, la description d’usages attestés et leur représentation, l’analyse stylistique et la littéarité, ainsi que sur la pertinence de l’outil informatique et de la linguistique de corpus pour les analyses.

La démarche consistait en une analyse stylistique outillée à l’interface entre linguistique et littérature, que ce soit du point de vue des phénomènes étudiés ou des concepts et de la méthodologie utilisés. Ma position était, et est toujours, que pour l’interprétation de phénomènes qui mettent en jeu à la fois des observables objectifs et des représentations socio-culturelles, il est souvent plus efficace de coordonner des approches qualitatives et quantitatives.

Je rappelle brièvement quelques questions générales (section 1.1), je montre ensuite une partie du processus de définition linguistique et symbolique de la variété ciblée et les contraintes de sa représentation (section 1.2), et je termine en présentant la méthodologie de corpus et le traitement quantitatif des *profils linguistiques* (section 1.3).

#### Références des travaux concernés<sup>2</sup>

2002b [3], 2003 [4], 2005 [5], 2006 [6], 2007 [7]\*, 2008a [8]\*,  
2008c [10]\*, 2010a [13]\*, 2012d [19]\*

## 1.1 Linguistique et littérature

Quelles sont les voies les plus pertinentes pour interpréter un texte, et plus précisément un texte dramatique, sans risquer la surinterprétation ? Cette question est largement traitée par l’analyse du discours, la théorie littéraire et la sémiotique dont j’ai articulé certaines dimensions ou notions. Je les rappelle dans cette section.

Premièrement, le texte est interprété comme un discours au sein d’un schéma de communication, c’est-à-dire que son interprétation suppose des émetteurs et des récepteurs qui font appel à des ressources contextuelles (section 1.1.1.). Deuxièmement, le texte peut aussi être interprété dans son fonctionnement interne. J’ai utilisé les programmes narratifs et les fonctions actancielles pour décrire les rapports des personnages et les ressorts dramatiques (section 1.1.2). Troisièmement, le style correspond à la perception d’une singularité par rapport à différentes normes et conventions, dont les conventions génériques (section 1.1.3).

1 L’analyse comparative a été effectuée sur la base d’extraits des corpus *Sankoff-Cedergren* et *Montréal 84* sur lesquels j’ai pu travailler directement à Montréal, et sur cinq pièces de théâtre de l’écrivain québécois Michel Tremblay. J’ai traité toutes ces données moi-même pour pouvoir les exploiter dans une perspective textométrique (voir section 1.3, p. 36).

2 Quelques travaux cités ont été publiés avant l’obtention de la thèse, mais je les mentionne tout de même, car ils ont servi de base à des publications post-thèse et car ils constituent un appui pour certaines perspectives (section 4.2, p. 163).

### 1.1.1 Communication littéraire et ressources contextuelles

Le texte n'est pas un produit hors-sol. Son existence est la trace d'une production et d'une réception. L'appréhender ainsi revient à lui rendre son statut de discours, d'objet co-construit au sein d'un processus de communication et à le lier à deux subjectivités et à leurs contextes. L'intérêt de cette intersubjectivité fondamentale<sup>3</sup> est de pouvoir penser la production et la réception comme des processus distincts ancrés dans des espaces-temps non nécessairement concomitants et façonnés par un certain nombre de contraintes. Ce point est important pour la communication littéraire, qui est une communication différée et décalée, c'est-à-dire où la production et la réception sont distantes dans l'espace et dans le temps et où l'auteur ne connaît pas son lecteur et inversement, chacun n'étant finalement qu'une projection mentale de l'autre par l'intermédiaire du texte<sup>4</sup>. La complexité du processus de communication est plus ou moins bien modélisée selon que l'on insiste sur tel ou tel aspect ou telle ou telle contrainte<sup>5</sup>. Pour le texte littéraire, il est particulièrement intéressant de pointer ce que Kerbrat-Orecchioni appelle les *compétences idéologiques et culturelles* et les *contraintes de l'univers du discours*. Les compétences idéologiques et culturelles comportent notamment les connaissances implicites sur le monde et tout ce qui permet « l'interprétation et l'évaluation de l'univers référentiel ». L'univers du discours désigne un ensemble hétérogène de contraintes, présentées comme des « filtres limitant les possibilités de choix (et orientant symétriquement l'activité de décodage) » qui renvoient aux conditions concrètes de la communication et aux contraintes de genre (Kerbrat-Orecchioni 2009 [1980], p. 13-37). Ces différents éléments constituent autant de ressources contextuelles<sup>6</sup> que j'ai exploitées pour comprendre le travail d'écriture de Tremblay. Certaines des approches sémiotiques de la littérarité et du style sur lesquelles je me suis appuyée s'inscrivent aussi dans une perspective discursive adaptée aux particularités littéraires (par exemple Fisette 1996, Molinié 1998, Molino 1989 et 1994, Rastier 2001, Saint-Gérard 1993). Elles recourent plus ou moins explicitement à la sémiotique pragmatiste de Peirce pour insister sur le statut du texte comme signe, sur le processus cognitif d'interprétation et l'attribution d'une valeur littéraire à réception<sup>7</sup>. Dans mes recherches, j'ai retenu les aspects suivants : la stratification énonciative des textes littéraires (1.1.1.1), l'association des ressources contextuelles situationnelles et

3 En référence à la définition que donne Benveniste du discours, à savoir « la langue en tant qu'assurée par l'homme qui parle, et dans la condition d'*intersubjectivité*, qui seule rend possible la communication linguistique ». (1966/1, p. 266)

4 Cette idée est exploitée à travers la relation « narrateur-narrataire » de Genette (1972, p. 262), l'*implied writer* et l'*implied reader* de Booth (1983, p. 420-432), ou l'« Auteur et au Lecteur modèles » de Eco (1985 [1979], p. 77). Elle est à mettre en relation avec la dimension cognitive de l'interprétation (voir ici section 1.1.1.3, p. 19).

5 Il existe un très grand nombre de schémas de la communication. Voir par exemple Hymes (2001 [1977]), Kerbrat-Orecchioni (2009 [1980], p. 13-37), Leech (1989, p. 1-18), Lyons (1977/1, p. 32-51) et Sperber & Wilson (1995 [1986], chap. 1).

6 Pour des travaux relativement récents sur le contexte, voir Fetzer (2012 et 2017), Meibauer (2012), Renkema & Schubert (2018, chap. 7) et van Dijk (2008, p. 5-25).

7 La réflexion théorique sur le processus de signification chez Peirce m'a beaucoup intéressée et continue de m'intéresser comme cadre général. Elle est fort complexe dans le détail et comporte plusieurs versions, ce pourquoi je ne la détaille pas.

socio-culturelles (1.1.1.2) et la dimension cognitive du processus d'interprétation (1.1.1.3).

#### ***1.1.1.1 Stratification énonciative des textes littéraires***

Que l'on parle de structure polyphonique, d'enchâssement des actants de l'énonciation, de diffraction des émetteurs et des récepteurs ou de feuilleté énonciatif, l'idée est la même : il y a minimalement trois niveaux d'énonciation. Le premier niveau est celui de la communication scripteur-lectorat. Le deuxième niveau est celui des instances énonciatives textuelles proprement dites, à savoir le niveau narrateur-lecteur, plus ou moins explicite mais toujours présent. Le troisième niveau est celui des échanges entre personnages visible dans les discours rapportés et les dialogues. Dans le détail, tous les textes n'ont pas la même structure énonciative. Il y a notamment des particularités associées aux genres. Par exemple, le nombre de strates, la focalisation du narrateur, le statut des personnages, etc. varient selon que l'on se trouve dans un texte dramatique ou dans un texte romanesque. Cette manière de penser le texte et de le représenter m'a servi pour mettre en évidence la fonction des marques d'oralité populaire québécoise dans les textes de Tremblay (à quels niveaux interviennent-elles ?), pour montrer que l'évolution de leur fonction correspondait aussi à une hybridation générique.

#### ***1.1.1.2 Association des ressources situationnelles et socio-culturelles***

Pour interpréter l'écriture de Tremblay, il m'a été nécessaire de posséder ou d'acquérir certaines connaissances extérieures à ses textes, auxquelles les textes renvoient ou qu'ils codent. L'idée que les ressources situationnelles et les ressources socio-culturelles sont assez systématiquement associées lorsqu'on construit une ontologie des contextes est un point très important pour la construction du référent des marques d'oralité. Elles renvoient à une réalité linguistique objective (des prononciations, du lexique, des tournures, etc.) et c'est en cela qu'elles sont perçues comme réalistes. Pour tester leur valeur référentielle, on peut les analyser en les confrontant à des échanges linguistiques réels. Mais ces marques ne sont pas une simple transcription d'une réalité, elles en sont plutôt un « transcodage » (Gauvin 2000, p. 130) complexe qui prend en compte leur coloration sociale et culturelle. Leur usage indique au lecteur que l'échange est plutôt public ou privé, que les personnages jouent un rôle social particulier, dominant ou sont dominés, etc. Leur présence en littérature peut aussi manifester une position idéologique de l'auteur à un moment donné. Pour sonder leur richesse symbolique, il est nécessaire de les connecter au système des valeurs sociales, culturelles et idéologiques du moment de leur production, en étudiant par exemple les discours et les mouvements sociaux et esthétiques ambiants, la réception critique des pièces à leur création. J'ai abordé la définition de l'oralité populaire québécoise en prenant en compte ces deux dimensions indexicales, ce qui a nécessité une immersion dans les données et une démarche d'acculturation importante. La question du *joual*, présentée à la section 1.2.1.2, est un bon exemple d'association de toutes ces valeurs.

#### ***1.1.1.3 Dimension cognitive de l'interprétation***

Cet aspect n'est développé et explicité que dans certaines des approches citées, mais il affleure de manière récurrente dans les modèles de la communication, dans la

sémiotique de tradition peircéenne et dans l'analyse du discours et la stylistique. Il est évidemment très présent en linguistique cognitive et en psycholinguistique. C'est un aspect relativement théorique qui constitue le cadre général dans lequel j'ai pensé ma démarche d'analyse des marques d'oralité, qui est une démarche d'interprétation. J'ai cherché à expliquer un effet qui est un « effet de réel » (Barthes 1987 [1982]) et aussi un effet esthétique. Cet effet est produit par le travail d'écriture de l'auteur, qui encode son texte de telle sorte que le lecteur puisse percevoir et décoder certaines graphies, certains mots et certaines structures comme des signaux qui lui sont adressés. Dans des théories du discours récentes<sup>8</sup>, la perspective cognitive est intégrée à travers le contexte vu comme une ressource contextuelle dynamique et construite. Dynamique car remise à jour en permanence. Construite car adaptée et négociée en fonction de ce que chaque participant projette des états de croyances, des connaissances et des intentions de son interlocuteur. Cela rejoint le processus connu de l'attribution d'intention, qui relève de la théorie de l'esprit en sciences cognitives<sup>9</sup>. L'interprétation des textes varie en fonction de ce que le lecteur est en mesure de décoder, compte tenu de ses connaissances, et de ce qu'il perçoit dans le texte comme encodé par l'auteur à son intention. D'une manière générale, cela revient à dire que le lecteur attribue à l'auteur une intention d'influencer la réception de son texte. On peut aussi tenter de projeter ce fonctionnement dans l'univers fictionnel pour voir si l'auteur a doté des personnages de stratégies, de capacités de projection mentale et si les dialogues en rendent compte. L'analyse de la structure actancielle va dans ce sens pour plusieurs personnages (voir section 1.1.2.1). Dans une étude consacrée à un texte de Tremblay<sup>10</sup> hors du corpus présenté dans cette synthèse (voir Dargnat 2005 [5], p. 47-56), j'avais adapté le schéma utilisé en sémiostylistique (Molinié 1998, p. 135) pour intégrer la dimension cognitive et le point de vue de la réception. L'idée était de montrer le style essentiellement comme un processus sémiotique dynamique contraint par le texte mais « dirigé » par le lecteur.

### 1.1.2 Structure narrative et polyphonique des textes dramatiques

Si l'on se concentre sur la description de la structure interne du texte, on est confronté à au moins deux sources potentielles d'énonciation : le narrateur et les personnages. Le premier a un statut particulier et variable selon les textes. La particularité du texte dramatique est d'être un discours qui représente essentiellement des discours, les dialogues entre personnages. Canoniquement, les personnages échangent et évoluent dans un univers imaginaire séparé du monde réel de l'auteur et du lecteur. Cette étanchéité constitue d'ailleurs une des conditions de l'illusion réaliste et du contrat fictionnel entre l'auteur et le lecteur : faire *comme si* les personnages n'étaient pas des personnages mais des locuteurs réels, sans conscience de leur propre construction

8 Voir les définitions du contexte ou du *common ground* chez Sperber & Wilson (1995 [1986], p. 15), Givón (2005, p. 100 sqq.), van Dijk (2008, p. 71 sqq.), Ginzburg (2012, p. 26) ou Renkema & Schubert (2018, chap. 7).

9 La théorie de l'esprit désigne la capacité pour un sujet de se représenter les représentations des autres, en particulier celles de son interlocuteur, à des fins diverses : l'informer, le convaincre, le manipuler, susciter son empathie, etc. Un ouvrage comme celui de L. Zunshine (*Why We Read Fiction : Theory of Mind and the Novel*, Ohio State University Press, 2006) fait explicitement le lien entre la théorie littéraire et la théorie de l'esprit.

10 *C'è à ton tour*, Laura Cadieux. Montréal : Bibliothèque québécoise, 1997 [1973].

comme personnage. Le niveau du narrateur est un niveau intermédiaire. Traditionnellement, il relève plutôt du genre romanesque et sa présence n'est pas toujours déclarée. Mais, même fantomatique, il est là, sinon comme une projection explicite de l'auteur dans son texte, au moins comme un principe d'organisation générale, comme celui qui tient les fils invisibles des marionnettes, les fait se mouvoir, apparaître et disparaître. Dans les textes dramatiques, les didascalies peuvent être attribuées à cet énonciateur-narrateur. Elles sont autant d'indications permettant au lecteur de se construire une description du décor, parfois des personnages, de leur manière de parler ou de leur états mentaux. La structuration interne et la division de la voix de l'auteur en voix du narrateur et voix des personnages est très importante pour construire l'inscription et la signification textuelles des marques d'oralité et pour en montrer l'évolution. Pour étudier le plus systématiquement possible la structure interne des textes, en termes de progression de l'action, d'organisation thématique et de rapports de force entre personnages, je me suis appuyée sur la sémantique structurale de Greimas (1966), l'analyse sémiotique des textes du Groupe d'Entrevignes (1979), la sémiostylistique déjà citée et la narratologie de Genette (1972 et 2004). J'en ai retenu les fonctions actanciennes des personnages dans les programmes narratifs (1.1.2.1), la complexification de la structure polyphonique et les cas de métalepse (1.1.2.2). La mise en perspective de ces différents phénomènes m'a permis entre autres de montrer l'enrichissement du référent des marques d'oralité (on passe du réalisme social à l'affectif de l'intimité) et la thématisation de la langue au fil des textes (on passe de la langue comme habit social à la langue comme marqueur métalittéraire). Cette évolution est visible en particulier lorsqu'on cherche à situer l'emploi des marques d'oralité dans les structures narrative et énonciative des textes.

### **1.1.2.1 Personnages, oralité et parcours narratifs**

Dans l'économie générale d'une pièce, il faut repérer le moteur de l'intrigue qui fait passer de la situation initiale à la situation finale. Ce moteur correspond souvent au désir du personnage principal, au but qu'il cherche à atteindre et autour duquel les autres personnages jouent un rôle d'*adjuvant* ou d'*opposant*, dessinant des forces contraires qui se matérialisent dans des paroles et des actes. Ces paroles et ces actes concernent un ou plusieurs personnages et structurent le texte en différentes étapes. Pour représenter ce moteur de l'action, la répartition des forces et les différentes étapes de l'action, j'ai d'abord utilisé les carrés sémiotiques et les schémas actanciels (Dargnat 2002b [3], p. 183-216). Ensuite, je me suis concentrée sur les programmes – dits aussi parcours, schémas ou canevas narratifs – que j'ai appliqués à chaque texte, pour pouvoir les comparer et en percevoir la spécificité. Globalement, le programme narratif correspond à ce qui se passe entre une situation initiale et une situation finale<sup>11</sup>. Les étapes, les rôles et l'importance qui leur est donnée peuvent varier d'une approche théorique à l'autre, et, bien sûr, d'un texte à l'autre<sup>12</sup>. Dans tous les cas, la séquentialisation narrative et l'attribution d'une fonction actancielle aux personnages,

11 Dans l'art dramatique et dans les analyses scolaires des pièces, on parle traditionnellement d'exposition, de nœud, de dénouement, de péripétie et de reconnaissance.

12 Je me suis inspirée de la sémiotique des textes du Groupe d'Entrevignes (1979), pour qui la séquence narrative, ou organisation logique d'énoncés narratifs, suit le modèle Manipulation-Compétence-Performance-Sanction. Avec du recul, je trouve que ces éléments ne sont pas toujours faciles à définir et à identifier.

seuls ou en groupe, donnent un cadre pertinent pour « mesurer les récits » et les comparer et, dans mon cas, pour comprendre le rôle de la parole, des marques d'oralité et du retentissement social et émotionnel de la langue dans le sens profond et l'organisation de chaque histoire<sup>13</sup>. Par exemple, dans *Les Belles-sœurs*, le personnage principal, Germaine Lauzon, a gagné un million de timbres-primés qui lui permettront d'obtenir gratuitement tout ce que la ménagère populaire québécoise de la fin des années 1950 rêve de posséder. Afin de parvenir à coller toutes ces promesses de consommation dans les livrets associés, elle invite ses « belles-sœurs », ses homologues du quartier, pour l'aider à coller ses timbres. Mais, au fur et à mesure que l'action et les discussions avancent, deux camps se forment et un groupe, par jalousie, dérobe progressivement les timbres, si bien qu'à la fin Germaine se retrouve sans rien, son petit rêve matérialiste est anéanti. Elle tire elle-même la sanction de son échec en reconnaissant sa dépossession. La sanction vient aussi de l'extérieur, puisque Germaine termine sur scène en se lamentant sous une pluie de timbres-primés et avec l'hymne national du Canada en fond sonore. Entre temps, quelques séquences, péripéties ou programmes narratifs secondaires ont montré la « maudite vie plate » des femmes d'alors, l'hyper-correction linguistique du personnage de Lisette de Courval qui se revendique d'un voyage en « Urope », la bigoterie ambiguë, la débauche dans les clubs, etc. Les marques d'oralité concernent a priori tous les personnages, mais dans une moindre mesure la fausse noble de Courval, qui semble bien mieux « perler » que les autres. Dans cette première pièce du corpus, il y aurait beaucoup de nuances à faire et de symboles à exploiter, mais on simplifiera en disant que les marques d'oralité ont d'abord une fonction de réalisme langagier qui est de représenter la variété linguistique des foyers francophones populaires de Montréal. Elles permettent déjà de distinguer deux postures, les Germaine Lauzon gouailleuses d'un côté et les de Courval pincées de l'autre.

C'est une opposition que l'on retrouve aussi très fortement dans *L'Impromptu d'Outremont*, entre Lorraine et Fernande, et dans *Encore une fois, si vous permettez*, entre une facette du Narrateur (figure de Tremblay) et Nana (figure de sa mère). Dans ce deuxième cas, l'objectif de la pièce est de faire revivre Nana sur scène et de lui inventer une mort théâtrale en apothéose. Les marques d'oralité sont un moyen de déclencher des scènes de souvenir entre le personnage Narrateur enfant et sa mère Nana, et de la faire revivre dans la générosité, la drôlerie et le drame de sa parole.

### ***1.1.2.2 Structure énonciative et polyphonie***

Tremblay joue avec la structure énonciative. Une schématisation montre de manière simplifiée mais efficace que les pièces choisies correspondent à quatre structures différentes, deux ayant une structure identique. Les quatre figures ci-dessous correspondent aux quatre schémas rencontrés. Elles sont construites de façon à rendre compte de la perception des sources d'énonciation par le lecteur.

13 Pour l'analyse systématique des cinq pièces du corpus, voir Dargnat (2006 [6], chap. 7), dont les conclusions sont reprises dans Dargnat (2008a [8], 2008c [10], 2010a [13] et 2012d [19]).

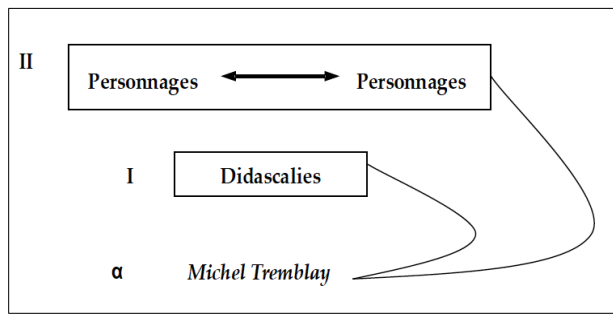
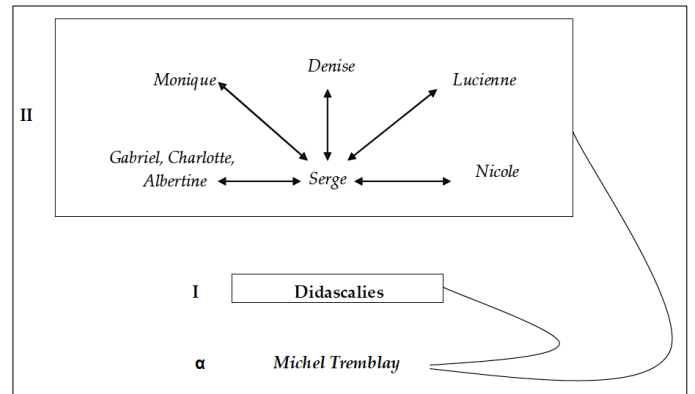
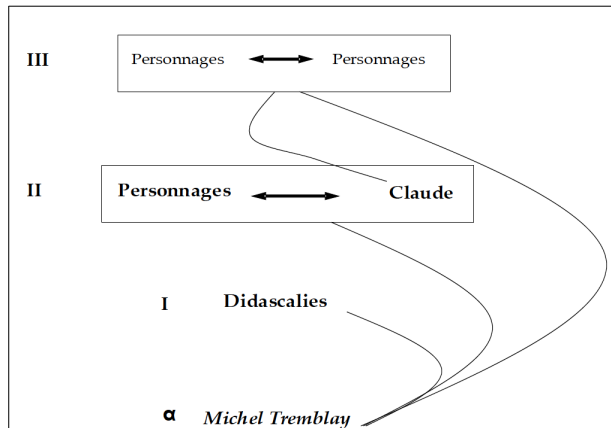
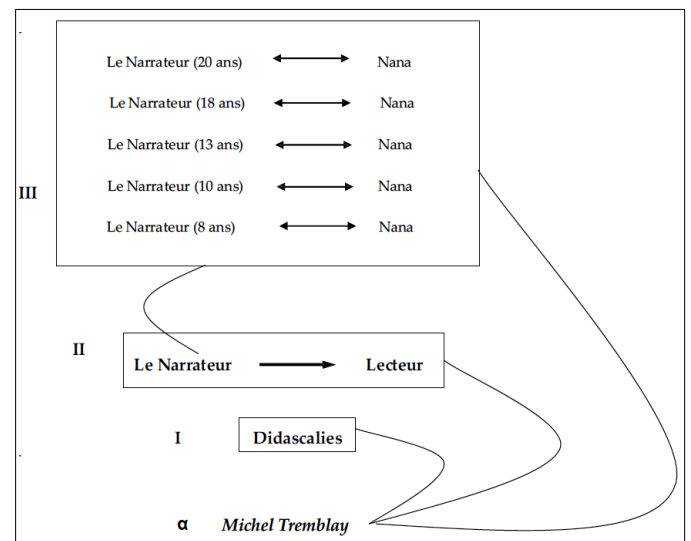
a. *Les Belles-sœurs* (1968) et *L'Impromptu d'Outremont* (1980)b. *Bonjour, là, bonjour* (1974)c. *Le Vrai Monde ?* (1987)d. *Encore une fois, si vous permettez* (1998)

Figure 1. Schémas énonciatifs des cinq pièces de Michel Tremblay

La Figure 1a concerne les deux pièces les plus conventionnelles et réalistes du point de vue de la structure, *Les Belles-sœurs* et *L'Impromptu d'Outremont*, dans le sens où elles respectent les unités de lieux, de temps et d'action. Les trois autres pièces ont des structures plus éclatées. La compréhension y repose sur la capacité qu'a le lecteur à délinéariser le texte et à faire des sauts de répliques pour reconstruire la continuité des échanges emmêlés. *Bonjour, là, bonjour* (Figure 1b) présente plusieurs dialogues simultanés entrelacés dans des lieux différents autour du personnage central Serge. *Le Vrai monde ?* (Figure 1c) expose une pièce dans la pièce, écrite par Claude, personnage qui s'essaie à l'écriture dramatique en s'inspirant de ses proches. On y voit des échanges présentés comme réels entre des personnages nommés Claude, Alex I, Madeleine I et Mariette I, entrelacés avec des échanges présentés comme fictionnels (pièce dans la pièce) entre des Alex II, Madeleine II et Mariette II. *Encore une fois, si vous permettez* (Figure 1d) montre un même personnage nommé Le Narrateur avec deux niveaux de dialogue en alternance dans des lieux et des temps différents, et à deux niveaux énonciatifs : en adresse directe au lecteur et dans des échanges rétrospectifs avec l'autre personnage Nana. L'information de l'âge est noyée dans les répliques et permet de découper l'ensemble en épisodes ou scènes.



Les marques d'oralité sont présentes dans tous les textes, mais fonctionnent différemment. A première vue, elles n'interviennent qu'au niveau des dialogues entre personnages dans les pièces correspondant aux Figures 1a, 1b et 1c. Ces marques permettent de regrouper ou de différencier les personnages selon leur profil social, c'est leur valeur référentielle de base. Dans la Figure 1d, les marques d'oralité sont distribuées autrement, elles continuent à créer l'effet d'oral et contribuent au réalisme langagier, mais elles sont réservées aux échanges entre le personnage Le Narrateur et Nana, c'est-à-dire à l'univers des souvenirs et à l'affectivité. Elles sont absentes des monologues ou répliques que Le Narrateur adresse au lecteur ou au public. Elles jouent donc un rôle de marquage des niveaux d'énonciation et de la polyphonie de ce personnage-pivot, à la fois récitant et acteur de sa propre pièce.

L'apparition d'un niveau III dans les Figures 1c et 1d correspond à une mise en abyme, une pièce dans une pièce, et à l'émergence progressive d'un personnage orchestre qui semble distribuer la parole. Il pointe déjà avec Serge, se matérialise comme écrivain amateur avec Claude, jusqu'à son aboutissement en narrateur explicite et autobiographique avec Le Narrateur, figure de Tremblay. La mise en abyme et l'adresse directe au lecteur ou au public ont été analysées en termes de valeur métalittéraire et de métalepse narrative. La valeur métalittéraire émerge à chaque fois que le lecteur perçoit une réflexivité dans la pratique littéraire. En se représentant elle-même, la littérature se « vend » comme telle et sape toute intention de réalisme. La métalepse narrative est un processus bien connu en littérature<sup>14</sup>. Elle participe à cette avancée à découvert de la production littéraire, car elle en expose les rouages et constitue une « entorse » au pacte fictionnel habituel. On rangera sous cette étiquette les cas où le narrateur surgit brusquement dans sa narration en tant que narrateur, et inversement, les cas où un personnage semble s'adresser directement à un interlocuteur appartenant à un niveau différent du sien dans les schémas énonciatifs. Quand elles permettent d'identifier ce mécanisme, comme c'est le cas avec la scission énonciative du personnage Le Narrateur, les marques d'oralité sont le support linguistique de la métalepse narrative (voir Dargnat 2010a [13]). C'est une des évolutions majeures de leur fonction, qui est corrélée à d'autres phénomènes comme l'autobiographie et l'hybridation générique.

Dans le détail, la poussée métalittéraire et métaleptique se construit plus progressivement dans d'autres exploitations de la polyphonie. Elle se cache aussi dans le caractère explicitement musical de certaines séquences, comme la « Cantate de la maudite vie plate », l'« Ode au Bingo », ou les répliques chorales dans *Les Belles-sœurs*. Dans *Bonjour, là, bonjour*, c'est la structuration entière de la pièce qui est musicale, les divisions étant intitulées « solo », « duo », etc. Cette dimension de la musicalité et du lyrisme de l'oralité doit être mise en relation avec l'hybridation générique, en particulier du côté de l'opéra et de la comédie musicale, qui sont des domaines consciemment exploités par l'auteur.

14 La formule *métalepse narrative* est empruntée à Genette (1972, 2004) puis à Pier & Schaeffer (2005). La métalepse narrative est définie comme « toute intrusion du narrateur ou narrataire extradiégétiques dans l'univers diégétique (ou des personnages diégétiques dans un univers métadiégétique, etc.), ou inversement » ou encore comme une « transgression délibérée du seuil d'enchâssement » (Genette 1972, p. 243-244). On pourrait rapprocher cela des « remontées actanciennes » dans l'approche sémiostylistique (Molinié 1998, p. 51) pour désigner le passage – délibéré et transgressif – d'un niveau énonciatif à un autre.

### 1.1.3 Genre et style

Le dernier point sur lequel je souhaite insister est celui de l'articulation entre genre de discours et style. Je me suis concentrée sur le domaine littéraire, très documenté et théorisé depuis fort longtemps, mais il s'agit d'un cas particulier d'une problématique plus générale. En effet, le genre et le style sont des notions opérationnelles à chaque fois qu'il s'agit de montrer l'appropriation individuelle de contraintes conventionnelles liées à un but communicatif. L'analyse du discours, la sociolinguistique et la linguistique de corpus supposent à un niveau général une prise en compte des formes conventionnelles de discours que sont les genres et une individuation de leur réalisation qui constitue le style<sup>15</sup>. Étant donné qu'ils sont conventionnels et partagés par une communauté (une discipline, un domaine, un groupe social), les genres sont des traits d'union entre le textuel et les contextes sociaux, socio-culturels et cognitifs évoqués précédemment (section 1.1.1.2). Je synthétise deux points en les ramenant à la question de la représentation littéraire de l'oralité : les propriétés et contraintes du genre dramatique (1.1.3.1) et la place du style (1.1.3.2).

#### 1.1.3.1 Propriétés et contraintes du genre dramatique

Les textes dramatiques sont associés à un certain nombre de conventions qui font partie des connaissances partagées et que l'analyste doit rendre explicites. Dans un texte écrit pour être joué, on s'attend à trouver minimalement des personnages qui parlent entre eux et un appareil didascalique pour guider le lecteur et le metteur en scène sur l'élocution des personnages, leur tenue, le décor, etc. Souvent, le texte comporte des divisions en actes, scènes ou autres. Les contraintes de base du genre dramatique sont bien décrites, en particulier du point de vue de sa structure énonciative<sup>16</sup>. Une des caractéristiques fondamentales est la « double énonciation », c'est-à-dire la présence d'un « discours rapporteur, dont le destinataire est l'auteur (le scripteur) et d'un discours rapporté dont le locuteur est le personnage. » (Ubersfeld 1996, p. 187-188). On retrouve là une exploitation spécifique de l'enchâssement (le feuilleté) des niveaux énonciatifs déjà mentionné pour décrire les textes littéraires en général. Si l'on s'intéresse aux représentations théâtrales des textes, il est aussi pertinent de prendre en compte une troisième « couche », celle des comédiens sur scène, qui incarnent les énonciateurs-personnages (voir Petitjean 1999). Par exemple, dans *Encore une fois, si vous permettez*, il est important de savoir que le comédien qui joue Le Narrateur sur scène, figure de l'auteur, n'est autre que le metteur en scène de prédilection de Tremblay, et le metteur en scène de la pièce elle-même. Cela apporte un niveau de complexité supplémentaire au système référentiel (mise en abyme et valeur métalittéraire).

En plus de cette double ou triple énonciation, il existe des conventions plus spécifiques qui dessinent des sous-genres, comme la comédie, la tragédie, le burlesque,

---

15 Pour des réflexions générales, voir par exemple Adam (2000, 2019), Bahktine (voir Todorov 1984, p. 123-131), Biber & Reppen (2015, partie 3), Biber & Conrad (2019), Combettes (2002), Girard (2019), Habert et al. (1997), Rastier (2001, chap. 8), Renkema & C. Schubert (2018, p. 181-207). L'idée est aujourd'hui bien admise et corrélée à celle des discours de spécialité. Les travaux académiques et les articles scientifiques sont de bons exemples pour l'application et pour l'étude des contraintes génériques (voir Bhatia 2013, Swales 2013).

16 Voir par exemple la « duplicité du dialogue théâtral » chez Maingueneau (1991, p. 141), le « dédoublement des instances énonciatives » chez Kerbrat-Orecchioni (2009 [1980], p. 171) et les réflexions générales de Biet & Triaud (2006), Pavis (1996) et Ryngaert (1993).

etc. qui imposent à qui veut en respecter les schémas classiques des contraintes de thèmes, de registres, de situations, etc. La sphère populaire et sa réalisation linguistique sont traditionnellement du côté du comique. Tremblay connaît très bien les schémas classiques et sait les utiliser pour les transgresser et les renouveler. Dans *Les Belles-sœurs*, en mêlant l'oralité populaire (le joual) et le schéma de la tragédie antique avec des séquences chorales, il casse les codes, fait scandale et les renouvelle. C'est précisément ce qui lui a été reproché et ce qui, en même temps, l'a classé comme auteur fondateur d'un théâtre québécois. La pièce est devenue un classique de la littérature québécoise et fait même partie de la *Bibliothèque idéale* de B. Pivot.

Ceci m'amène à évoquer l'idée que le genre est une catégorie dynamique, dans le sens de Bakhtine (1984 [1979], p. 263-308). Ce dynamisme peut d'abord être décrit en termes de transfert ou de transgression génériques. Pour simplifier, chez Bakhtine, les textes littéraires sont des genres secondaires et complexes parce qu'ils réutilisent des genres dits primaires, comme les conversations quotidiennes. Lorsqu'ils sont réutilisés, les genres primaires perdent la signification qu'ils avaient dans la réalité et en acquièrent une autre dans l'organisation d'ensemble de l'univers fictionnel. J'ai retenu de cette idée la déplaçabilité de certains genres et leur réutilisation pour former des genres plus complexes. Cela s'applique tout à fait bien au théâtre, qui reproduit des discours, et en particulier aux pièces de Tremblay, qui réutilisent le genre des conversations familières quotidiennes et en modifient la signification en les intégrant à son projet esthétique. C'est tout l'enjeu du référent et de sa représentation que j'ai étudiés. En réalité, le « transfert des genres » dans l'écriture de Tremblay est plus riche que cela. Le transfert d'un genre simple dans un genre complexe est associé à un transfert des registres (voir ce qui vient d'être dit des *Belles-sœurs*). Il fonctionne aussi entre séquences déjà déclarées comme littéraires, puisque, dans *Encore une fois, si vous permettez*, l'auteur gère sa propre intertextualité en réutilisant des « morceaux » déjà publiés<sup>17</sup>. Le personnage Le Narrateur est une inscription explicite du genre dramatique dans le genre narratif et inversement. On pourrait dire la même chose du récit déjà mentionné *C'tà ton tour Laura Cadieux*, hybride générique entre roman et théâtre, qui donne lui aussi « matière à transposition<sup>18</sup> ». Il s'agit là de détournements, de recyclages individuels de conventions. Corollairement, le dynamisme comme transfert ou transgression entre genres identifiés peut être l'évolution, la redéfinition du genre lui-même, selon le processus bien connu de l'avant-garde que le temps finit par ranger à côté des classiques. D'un certain point de vue, on pourrait dire que les transgressions ou transferts singuliers de conventions génériques font l'histoire littéraire. La tension entre généricité et singularité est d'ailleurs un des supports de la littérarité : percevoir l'écart, l'inattendu, qui font cette saillance sur la ligne commune, associée à la définition d'un style.

17 Un extrait du récit *Un Ange Cornu avec des ailes de tôle* (1994) constitue une séquence entière de la pièce *Encore une fois, si vous permettez*.

18 Michel Bélaïr, « Michel Tremblay publie 'Laura Cadieux' : un roman presque théâtral ». *Le Devoir*, 21 avril 1973, p. 20.

### 1.1.3.2 La place du style

Le terme *style* est très courant, dans les domaines spécialisés ou dans le langage ordinaire. Il fait l'objet de multiples attentions, et une discipline lui est même consacrée<sup>19</sup>. Il désigne parfois les ressources linguistiques associées à un genre, les formules et tournures qu'il convient d'utiliser dans telle ou telle situation. Au niveau littéraire, cela est bien décrit : dans une tragédie, la convention (classique) laisse attendre que l'auteur présentera des personnages de haut rang échangeant dans un langage soutenu, et, à l'inverse, que, dans une comédie, certains personnages peuvent être de condition inférieure et utiliser un langage familier et populaire (par exemple Pierrot dans le *Dom Juan* de Molière). Dans ce cas-là, le terme *registre* semble plus approprié<sup>20</sup>.

Le style est aussi associé à une individuation, à ce qui, dans l'expression, caractérise un individu par rapport aux autres. Autrement dit, c'est la perception d'un choix parmi des possibles ou d'un écart par rapport à une norme, quelle qu'elle soit. En sociolinguistique, cela correspond à la dimension diaphasique de la variation, et en littérature, c'est ce que l'on appelle en général un style d'auteur. En étudiant la représentation de l'oralité populaire québécoise par un auteur dans des textes d'un genre particulier, j'ai inévitablement été confrontée à ces deux valeurs du style (*registre vs idiosyncrasie*). La section suivante montre que la représentation de la variation linguistique que Tremblay propose est le résultat d'un compromis entre sa créativité individuelle et des conventions stylistiques et génériques collectives.

## 1.2 Variation linguistique et contraintes de représentation

J'ai exploité la question de la variation à plusieurs niveaux (réfèrent, représentation et fonctionnement textuel) afin de répondre aux questions suivantes :

- Quel est le réfèrent de la représentation proposée par Tremblay ? Autrement dit, qu'appelle-t-on le français québécois populaire ? Quelles en sont les marques linguistiques ?
- Dans ses textes, retranscrit-il toutes les particularités de ce réel ou opère-t-il un tri ? Selon quelles contraintes ? Quelles sont ses ressources pour introduire de l'oral dans l'écrit ?
- Du point de vue du fonctionnement interne des textes, les marques d'oralité populaire québécoise assument-elles uniquement une fonction référentielle ? Ou servent-elles la fiction ? Si l'on compare plusieurs textes, le rôle des marques d'oralité populaire québécoise sont-elles distribuées et agissent-elles de la même manière ?

19 Je renvoie par exemple aux multiples manuels (Cogard 2001, Gardes-Tamine 1997, etc.), aux numéros de revues (*Langages* 118 en 1995, *L'information grammaticale* 70 en 1996, *Critique* 641 en 2000, *Langue française* 135 en 2002, etc.) ou aux ouvrages qui y sont consacrés entièrement ou en partie (Adam 1997, Bordas 2008, Frédéric 1997, Molinié & Cahné 1994, Rastier 2001, p. 167-188 etc.). Je ne cite que les approches françaises, mais la critique anglo-saxonne est également très riche et moins soumise au format des concours de recrutement.

20 Pour une discussion récente sur les différences entre *genre*, *registre* et *style*, voir les propos de Biber & Conrad (2019, p. 15-23) et Schubert & Sanchez-Stockhammer (2016, p. 2-5).

Ci-dessous, j'insiste sur le processus de catégorisation du *complexe variationnel* volontairement réifié en OPQ (pour « oralité populaire québécoise ») (section 1.2.1) puis sur les contraintes et les choix qui ont pesé sur sa représentation littéraire (section 1.2.2).

## 1.2.1 Catégorisation linguistique et symbolique de l'oralité populaire québécoise (OPQ)

Si l'on dit que Tremblay représente l'OPQ dans ses textes, encore faut-il avoir une idée de ce à quoi il fait référence, et ce que doit reconstruire son lecteur à partir des indices qu'il lui donne. Par exemple, quand on lit :

ROSE OUIMET

Voyons donc, Germaine, c'est pas des moineaux, tes timbres, y s'envoleront pas ! En parlant de moineaux, ça me fait penser, j'ai été voir Bernard, mon plus vieux, dimanche passé... [...] Écoutez ça, ça vaut la peine...

*Projecteur sur Rose Ouimet.*

J'vous dis, c't'une vraie folle ! J'ris, là, pis au fond, c'est pas drôle. Entéka... À Pâques, Bernard a acheté une cage à moineaux pour ses deux p'tits. C't'un gars à 'taverne qui avait besoin d'argent pis qui y'a vendu ça pas cher... Elle, quand à l'a vu ça, est v'nue folle tu-suite, est quasiment tombée en amour avec ses oiseaux ! [...] (M. Tremblay, *Les Belles-sœurs*, 1968, p. 23)

on relève au premier coup d'œil différents éléments qui permettent d'inférer qu'il s'agit d'un texte de théâtre où quelqu'un s'adresse à quelqu'un (disposition, didascalies). Plus encore, la manière dont certains mots sont graphiés, la présence de certaines expressions mettent sur la voie de l'oralité en général (la ponctuation expressive, les apostrophes, les *y* à la place de *ils* ou *il*, les dislocations, etc.), voire d'une oralité plus particulière demandant au lecteur un effort d'acculturation (*Entéka*, *à'taverne*, *tomber en amour*, *tu-suite*, etc.). L'« effet d'oral » (Gauvin 2000, p. 127) produit par ces différents phénomènes vise à fabriquer la « parlure » (Larthomas 2001 [1972], p. 412<sup>21</sup>) du personnage, à l'habiller à l'écrit d'une manière de parler non standard perçue comme réaliste.

Une partie de mes recherches a consisté à construire le référent auquel renvoient ces phénomènes. J'ai appelé ce référent OPQ afin d'objectiver et d'associer les trois dimensions variationnelles principales que sont le médium oral, le registre populaire et l'aire linguistique québécoise. Mais c'est aussi tout un imaginaire linguistique, idéologique et culturel. La prise en compte de ces deux composantes est indispensable pour comprendre ce que représente la période du *joual*, dont Tremblay est un des emblèmes.

### 1.2.1.1 Un complexe variationnel comme référent

Parler d'une pratique linguistique comme d'une variété linguistique, c'est lui donner une unité, une identité associée à une communauté. C'est aussi la distinguer d'autres variétés ou la positionner par rapport à un repère. Autrement dit, l'étiquette « variété x » résulte d'un processus de catégorisation qui permet d'encapsuler des usages pour les associer selon certaines dimensions de la variation. En sociolinguistique, il existe au

21 Le terme vient en fait de Damourette & Pichon, qui désignent ainsi la manière de parler propre à une classe sociale. Voir *Essai de grammaire de la langue française* (1911-1927), livre 1, chap. 3, § 33 et 33, p. 46 et 50 [Gallica].

moins deux manières d’appréhender la variation. Les deux m’ont été utiles. Premièrement, la variation selon les usagers ou les usages. On parle communément de variation diachronique, diatopique, diastratique, diaphasique ou stylistique et diamésique<sup>22</sup>. Ici, c’est le médium, la position sociale et le lieu qui sont les critères affichés par le sigle OPQ. J’ai ainsi volontairement regroupé des phénomènes linguistiques jugés caractéristiques du français oral et/ou du français populaire et/ou du français québécois, sans partir d’une étiquette idéologiquement et esthétiquement marquée. Deuxièmement, la variation se calcule aussi en termes de fonctions sociales ou symboliques de variétés identifiées. On cible ainsi un ensemble d’usages employés préférentiellement ou exclusivement dans telle ou telle situation, pour telle ou telle tâche. Le modèle est celui de la diglossie. Il suppose une hiérarchie entre au moins une variété dite « basse » et une variété dite « haute », « hautement codifiée (souvent grammaticalement plus complexe), véhiculant un ensemble de littérature écrite et respecté, [...] qui est surtout étudiée dans l’éducation formelle, utilisée à l’écrit ou dans un oral formel mais n’est utilisée pour la conversation ordinaire dans aucune partie de la communauté. » (Ferguson 1959, p. 336, traduit par Calvet 1993, p. 43)<sup>23</sup>.

Transposé à mon sujet, cela pose le problème de l’utilisation de phénomènes associés à une variété basse ou basilectale, l’OPQ, pour un usage d’habitude réservé à la variété haute, ou acrolectale, (le français formel écrit, littéraire). Cette hiérarchisation ouvre la porte à l’« imaginaire linguistique » (Houdebine-Gravaux 2002) de l’OPQ, à la fois pratique déclassée et pratique littéraire, et à une intrication de valeurs symboliques, que la question *joual* a cristallisées. J’insiste sur deux moments-clés de ma démarche de construction du référent des marques d’OPQ : la sélection des traits linguistiques représentatifs de la pratique réelle ciblée et ce que signifie le *joual* au Québec et dans l’œuvre de Tremblay pour comprendre l’articulation entre propriétés et imaginaire linguistiques.

Au début de mes recherches, je n’avais pas les environnements linguistique, scientifique, ni méthodologique pour prétendre à une description fine des particularités du français populaire québécois. C’est donc par une synthèse et une recension des travaux qui y ont été consacrés que j’ai commencé. Par la suite, j’ai pu accéder à ces environnements pour m’immerger dans la culture et la langue québécoises et bénéficier des corpus et d’une méthodologie appropriée, ainsi que de fonds documentaires pertinents. Les publications dans le domaine et la possibilité de travailler directement et librement sur les corpus *Sankoff-Cedergren* et *Montréal 84* m’ont permis d’aboutir à un portrait assez complet de traits linguistiques jugés représentatifs de l’OPQ et attestés dans les corpus. Cette recherche d’exhaustivité a fait émerger des problèmes intéressants. J’en retiens deux ici : (i) les critères de sélection des phénomènes

22 Ici, je m’appuie essentiellement sur les distinctions de Gadet (2003, 2021), à savoir les dimensions diamésique (le médium, oral ou écrit), diastratique (diversité sociale des locuteurs), diachronique (évolution temporelle), diatopique (diversification spatiale, localisation du locuteur), diaphasique ou stylistique (variation interne à un locuteur, répertoire dont il dispose pour s’adapter).

23 L’utilisation de plusieurs termes/expressions comme *français parlé*, *français oral*, *français écrit*, *français graphié* oblige à dépasser la seule opposition du médium et à penser une spécificité de structure. Cette question est cruciale dans certaines approches, tant en linguistique (Blanche-Benveniste et al. 1990, Hazaël-Massieux 1993, Koch & Oesterreicher 2001), qu’en poétique (Meschonnic 1987).

définitoires de l'OPQ (oral, populaire, québécois) ; (ii) le caractère non exclusif des principales dimensions variationnelles en jeu.

(i) Les critères de représentativité soulèvent le problème fondamental de la catégorisation d'un objet complexe, et, en quelque sorte, de sa compositionnalité. Comment circonscrire linguistiquement l'OPQ ? Il s'agit d'une entreprise d'« essorage » progressif et programmé. Une première étape a consisté à faire la liste des traits associés au français parlé, au français populaire et au français québécois à partir des études disponibles, en ciblant la période contemporaine des personnages, soit globalement 1950-2000. Ceci à tous les niveaux de la description linguistique : prosodie, phonétique, morphologie, lexicale, syntaxe. Quelques traits d'ordre discursif et pragmatique ont également été commentés. Lors d'une deuxième étape, j'ai sélectionné les phénomènes qui pouvaient être jugés caractéristiques de l'OPQ, dans l'esprit de la « caractérologie stylistique » de Gicquel (1999, p. 29-40) et de la synthèse sur les traits linguistiques de l'oralité de Chénétier-Alev (2004, p. 29-40). J'ai donc utilisé les études sur le français parlé, sur le français populaire et sur le français québécois comme des garants et des filtres pour sélectionner les traits jugés représentatifs d'une pratique ainsi érigée en variété, ces traits étant de potentiels candidats pour être représentés sous la plume de Tremblay. Voici quelques exemples pour illustrer :

*Phonétique* : relâchement des voyelles hautes [i], [y] et [u] en [ɪ], [ʊ] et [ɪ] dans certains contextes (ex. *habitude* [abɪtsɔd]) ; diphthongaison des voyelles longues accentuées (ex. [garaʔʒ]) ; affrication de [t] et [d] devant [ɪ] et [y] (ex. *tu dis* [tsydzi]) ; ouverture de [ɛ] en [ɑ] devant [r] (ex. *fermer* [farmɛ]), prononciation variable du digramme « oi » (ex. *moi* [mwe], *toi* [twe], *froid* [frɛt]), etc.

*Morphosyntaxe* : paradigme des pronoms personnels différent de l'écrit standard (*il(s)* [i/iz/j], *elle(s)* [ɑ/al/az], *nous autres* [nɔzot]) ; préférence pour certains suffixes (-*oune*, -*eux*, -*ette*), etc.

*Lexique* : usage de régionalismes (ex. *chenaille*, *chicaner quelqu'un*, *une guidoune*, *magané*), d'anglicismes plus ou moins intégrés ou traduits (ex. *pitcher quelque chose*, *le fun*, *un party*, *chambre de bain*), de sacres, ces jurons empruntant au vocabulaire liturgique, plus ou moins modifiés dans leur forme (ex. *crisse*, *cibole*, *maudit*, *tabarnacle*), etc.

*Syntaxe* : négation sans *ne* (ex. *tu viens pas ?*), mais aussi concordances négatives (ex. *y a pas personne icitte ?*) ; utilisation de la particule -*tu* pour former les interrogatives et les exclamatives (ex. *C'est-tu beau*, *ça*, *hein ?*) ; structures disloquées (ex. *ma mère*, *elle a dit qu'elle le voulait bien*, *elle*, *son chapeau*) ; relatives non standard (ex. *une amie que je suis avec elle depuis longtemps*), etc.

*Au niveau discursif* : utilisation de marqueurs discursifs plus ou moins marqués populaire et québécois dans leur forme ou leur usage (ex. *écoute donc/coudonc* [kudɔ̃], *en tout cas/entéka* [ãteka], *fait que/fak* [fak], *pis* [pi], etc.)

Il s'agit d'un travail préparatoire systématique, laborieux mais nécessaire, qui a donné lieu à plusieurs listes et tableaux<sup>24</sup> qui ont servi de base pour les publications ultérieures, de 2007 à 2012 ([7], [8], [10], [13] et [19]).

(ii) Un tel recensement confronte à la non-exclusivité des dimensions variationnelles. Dans l'idéal, il faudrait pouvoir typer chaque phénomène linguistique contribuant à l'OPQ, c'est-à-dire pouvoir lui attribuer un rôle O, un rôle P ou un rôle Q<sup>25</sup>. Par exemple, certains phénomènes sont oraux sans coloration diastratique ni diatopique (les dislocations), d'autres sont québécois sans correspondre à un médium ni à un niveau social particulier (certains emprunts ou certaines traductions), etc. Mais, le plus souvent, ces trois dimensions de la variation se juxtaposent. Certaines particularités sont orales, populaires et québécoises (par exemple les sacres), d'autres orales et québécoises (les diphtongaisons, les palatalisations, la particule *-tu*, etc.), etc. Il n'est pas évident de ranger chaque phénomène linguistique dans une seule case. De plus, ces trois dimensions se rejoignent dans une certaine idée du non-standard, avec toute la difficulté et la richesse symbolique de ce que cela représente. Je vois deux conséquences potentielles à cela pour la catégorisation de l'OPQ, et pour qui cherche à la représenter : une mutualisation des effets (les traits O, P et Q s'entraînent dans la construction du non-standard) et une intensification des effets (faire oral, populaire et québécois c'est faire non-standard « au cube »). La question du standard et du non-standard est largement débattue, et elle est liée aux fonctions sociales des variétés mentionnées plus haut. Elle dépend donc fortement du contexte social et idéologique<sup>26</sup>, comme le montre la réflexion suivante :

Quant à la variété linguistique appelée STANDARD, elle désigne un dialecte ayant une fonction sociale très particulière. [...] Ce serait donc une entité abstraite qui, dans l'absolu, n'est parlée par aucun groupe linguistique. Mais pratiquement, cette variété standard qui est celle utilisée dans les communications officielles, journaux, télévision, enseignement, etc. est la variété la plus prestigieuse, celle que les groupes sociaux privilégiés tendent à s'approprier. [...] À partir de ce consensus, tout écart linguistique face à la norme est jugé comme un écart social. (Forget 1979, p. 55-56)

Au Québec, de par l'histoire de l'implantation du français en Amérique du Nord, la problématique d'une langue standard est spécifique et relativement complexe. Elle ne se réduit pas aux débats soulevés par le *joual*, mais elle y est particulièrement visible.

### 1.2.1.2 La question du joual

Je rappelle les enjeux linguistiques, symboliques et esthétiques d'une période qui a vu naître toute une vague d'auteurs oralisants (joualisants)<sup>27</sup>. Culturellement, la période

24 La liste la plus complète est disponible dans Dargnat (2006 [6] Annexes, p. 65-72). Y sont recensées systématiquement toutes les particularités du français oral, du français populaire et du français québécois relevées dans les publications utilisées. Des listes plus restreintes sélectionnant les phénomènes jugés représentatifs étaient présentées dès Dargnat (2002b [3], chap.1), et de manière plus approfondie dans Dargnat (2006 [6], p. 78-83).

25 C'est ce que j'ai cherché à faire, en partie, sur le corpus de pièces de théâtre dans Dargnat (2008a [8], p. 21-35).

26 Abécassis (2003, p. 117) et Gadet (2003, p. 21) parlent d'« idéologie du standard », en référence aux travaux de Milroy & Milroy sur l'anglais (*Authority in Language*, London/New York, Routledge, 1999 [1985]).

27 Pour un panorama historique de la situation du français au Canada et au Québec depuis les premières implantations des colons voir Plourde (2000).



dite du *joual* va de la fin des années 1950 au début des années 1980 (Gervais 2000). Le recensement de plusieurs dizaines de définitions de 1870 à 2004<sup>28</sup> montre les différentes étapes de l'enrichissement sémantique, parfois contradictoire, de ce terme. Au départ, *joual* et *jouaux* sont la prononciation populaire ou rurale des mots *cheval* et *chevaux*, attestée dans les parlers de l'Ouest de la France, régions d'origine des premières vagues d'émigration des Français en Amérique du Nord. *Joual/jouaux* n'est donc pas a priori une spécificité québécoise, mais une marque de variation régionale. Très vite, par métonymie, le terme a désigné la prononciation et la manière de parler. L'expression *parler joual* fait son apparition dès 1930 au Québec dans le journal *Le Goglu* et elle est associée à un jugement négatif : parler *joual*, c'est mal parler. La connotation péjorative est largement exploitée par la suite avec la métaphore animale, notamment dans le célèbre pamphlet rédigé par un certain Frère Untel (alias Jean-Paul Desbiens) dénonçant la qualité de l'enseignement et du français au Canada.

Il y a proportion entre la chose et le nom qui la désigne. Le mot est odieux et la chose est odieuse. Le mot *joual* est une espèce de description ramassée de ce que c'est que le parler *joual* : parler *joual*, c'est précisément dire *joual* au lieu de *cheval*. C'est parler comme on peut supposer que les chevaux parleraient s'ils n'avaient pas déjà opté pour le silence et le sourire de Fernandel. [...] Le vice est profond, il est au niveau de la syntaxe. Il est aussi au niveau de la prononciation [...] Le *joual* est une langue désossée : les consonnes sont toutes escamotées, un peu comme les langues que parlent (je suppose, d'après certains disques) les danseuses des Îles-sous-le vent : oula-oula-alao-alao. On dit : « chu pas apable », au lieu de : « je ne suis pas capable ». [...] Cette absence de langue qu'est le *joual* est un cas de notre existence, à nous, les Canadiens français. On n'étudiera jamais assez le langage. Le langage est le lieu de toutes les significations. Notre inaptitude à nous affirmer, notre refus de l'avenir, notre obsession du passé, tout cela se reflète dans le *joual*, qui est vraiment notre langue. [...] Bien sûr qu'entre *jouaux*, ils se comprennent. La question est de savoir si on peut faire sa vie entre *jouaux*.<sup>29</sup>

Ce texte pose le *joual* comme le symptôme d'une blessure identitaire profonde, d'un sentiment de domination et d'aliénation, par les Canadiens anglophones ou la France, Mère-patrie. Les années 1960 et 1970, années dites de la Révolution tranquille au Québec, sont un tournant majeur. On assiste alors à une libération des carcans sociaux de la tradition classique et de l'église catholique et de revendication du peuple sur les élites. Dans ce bain idéologique révolutionnaire, l'identification du peuple québécois à cette manière de parler jadis dévalorisée et signe d'aliénation sert de pivot à un basculement des valeurs. Utiliser le *joual*, ce n'est pas seulement mal parler ou faire local, c'est adopter une position sociale, idéologique et politique active ; position dont la culture et la littérature s'emparent pour renverser les valeurs traditionnelles. Au niveau artistique général, cela se traduit par la culture pop – le Ti-Pop étant sa traduction québécoise<sup>30</sup> – qui récupère, transforme et profane les codes traditionnels et religieux au moyen de thèmes et procédés empruntés à la vie de tous les jours, au peuple et à la société de consommation. Plus précisément, au Québec, dans le domaine littéraire, on voit un grand nombre de productions mettre en scène et en texte des personnages de la

28 Le relevé exhaustif est consultable dans Dargnat (2006 [6] Annexes, p. 54 sq.).

29 Desbiens J.-P. (pseudonyme : Frère Untel), *Les insolences du Frère Untel*. Montréal : Les Éditions de l'Homme. Réédition de 2000, p. 23-24-25.

30 Dans des travaux plus anciens, j'ai exploré un contexte idéologique et esthétique plus large, celui de l'hypermélangisme, du Pop'Art et du Ti-Pop.

classe populaire, dans des situations et avec des thèmes de la plus grande banalité, faisant usage d'une langue non normative qui comporte des particularités phonétiques, syntaxiques et lexicales associées à l'imaginaire du *joual*. Un peu plus tard, le terme *joual* élargit son spectre référentiel pour désigner une pratique d'écrivain cherchant l'effet d'oralité en littérature, comme l'illustrent les propos de Godbout, essayiste et écrivain :

Qu'est-ce que le joual ? Peu importe qui baptisa ainsi le franglais de Montréal, avant André Laurendeau, avant le Frère Untel ou leur mère l'Université Laval, il reste qu'aujourd'hui ce mot décrit, dans la pensée populaire, le langage populaire. Le joual ce n'est plus le nom commun qui dit la dislocation du français des champs au contact de l'anglais des villes. Le joual est devenu une appellation contrôlée de l'un des niveaux de langage, à la disposition de l'écrivain québécois comme tous les autres niveaux langagiers.<sup>31</sup>

Plus encore, c'est tout l'imaginaire du non-standard qui est convoqué chez Poirier, pour qui le sens du terme s'étend jusqu'à « toute variété linguistique considérée comme déviante par rapport à une norme donnée » (1998, p. 327). Beaucoup d'études ont été consacrées à la langue et aux discours sur la langue au Québec ainsi qu'au *joual*. Il en ressort qu'il est un symptôme ambivalent de l'intrication d'usages linguistiques et de valeurs symboliques.

Je retiens de ces approches et de mes travaux que le *joual* est tout autant une pratique linguistique qu'un épiphénomène social, culturel et politique. Je m'y suis intéressée de près, car l'œuvre de Tremblay y est associée depuis le début, et en est même un emblème. En effet, *Les Belles-sœurs* a été qualifiée de première pièce en *joual*. Comme le référent du *joual* s'est diversifié avec le temps et les événements, le rapport de l'écrivain à la langue a muté. Il s'agit donc d'un référent complexe et changeant, ce qui en fait une porte d'entrée privilégiée dans le fonctionnement sémiotique global de la langue en usage.

J'ai d'abord appréhendé l'OPQ comme un ensemble de traits linguistiques jugés caractéristiques du français populaire québécois. Pour en étudier la représentation, j'ai enrichi cette première définition de l'imaginaire associé au non standard et à la problématique du *joual*. Le référent de Tremblay est donc un double complexe variationnel, au niveau des composantes descriptives (médium, classe sociale, lieu) et au niveau des valeurs symboliques (dénonciation, identification, création).

### 1.2.2 Représenter et décoder l'OPQ : la sensation d'un travail

Comment Tremblay réussit-il à faire effet d'OPQ dans ses textes, à construire une langue de fait fictionnelle mais perçue comme réaliste ?

Cette question soulève celle du réalisme dans les arts. Le réalisme langagier en littérature amène à son tour plusieurs problèmes théoriques qui font se rejoindre philosophie du langage, linguistique, stylistique et théorie littéraire. Une des idées centrales est celle de la lisibilité comme tension entre fidélité et efficacité<sup>32</sup>. L'écrivain réaliste n'agit pas comme un linguiste qui transcrirait des enregistrements en utilisant scrupuleusement l'alphabet phonétique, ce qui le rendrait illisible. Il filtre ce qui, dans

31 J. Godbout, « Entre l'Académie et l'écurie ». *Liberté*, vol. 16, n° 3, (mai-juin), 1974, p. 27-28.

32 En référence à la formule de Paul Valéry : « L'art commence par le sacrifice de la fidélité à l'efficacité » (« Poïétique », *Cahiers*, 1974, p. 993).

la pratique linguistique réelle qu'il cible, sera sémiotiquement le plus rentable pour en donner l'illusion à son lecteur.

Cette activité conduit à la perception d'un désir de l'écrivain de produire un effet, ici un effet d'OPQ. En même temps que le lecteur décode le référent vers lequel tend l'écriture qu'il lit, il reconnaît la présence de l'écrivain dans les choix opérés et les écarts volontaires. C'est la « sensation directe d'un travail »<sup>33</sup>. Le fait de style est précisément constitué par la détection de ce travail. Concrètement, les marques d'OPQ dans les textes que j'ai étudiés sont le résultat d'un ensemble de choix singuliers à l'intérieur de contraintes générales, notamment celles liées (i) à la lisibilité, (ii) à la catégorisation linguistique et symbolique évoquées ci-dessus et (iii) au genre dramatique.

(i) Concernant la lisibilité, la contrainte la plus évidente est celle de la recherche d'un équilibre entre l'écart par rapport à la norme orthographique, qui reflète par exemple une réalité phonétique, et l'effort à fournir pour le lecteur pour récupérer cette réalité. Si la forme graphique entraîne un traitement trop lourd, trop inhabituel, l'effet de réel est raté, car le lecteur bute sur la matérialité du signe. C'est la raison pour laquelle, à mon avis, les diphtongaisons et les palatalisations ne sont pas représentées, bien que très symptomatiques du français québécois, populaire ou non. Par exemple, un mot comme *habitude* ([abitsød]) ou une séquence comme *tu dis* ([tsvdzi]) ne sont pas graphiés « abetsed » ou « tzedzé ». Même remarque pour le découpage de la chaîne parlée, qui, sauf quelques rares exceptions passées dans l'usage, demeure très respectueux de l'unité lexicale. L'emploi non standard de l'apostrophe apparaît comme un bon compromis (*t'sais*, *not'faute*, etc.). Le Tableau 1 sur les « néographies phonétisantes »<sup>34</sup>, repris de Dargnat (2008a [8], p. 24-25), illustre le travail de filtrage à partir du référent phonétique<sup>35</sup>. Les exemples-types ont été relevés dans les cinq pièces étudiées.

#### VOYELLES

Phénomène non standard	représenté	non-représenté	exemples-types
<b>Voyelles [i,y,u]</b>			
• relâchement	X		menute
• désonorisation		X	
• syncope		X	
<b>Voyelle centrale [ə]</b>			
• syncope	X		v'nir
• harmonisation en [e]	X		bédaine, dis-lé
• harmonisation en [a]	X		sacrament
<b>Voyelle mi-ouverte [ɛ]</b>			
• ouverture en [a] devant [r]	X		farmer
• neutralisation [ɛ,e]	X		pére
<b>Voyelle [A]</b>			
• postériorisation de [a] en [ɑ]	X		cârosse

33 « La présence de Proust, de Joyce, de Faulkner dans leur texte, c'est moins l'apparente perfection des effets, l'inimitable justesse du timbre, que la sensation à moi transmise, d'un effort, d'une tension, d'une décision jamais reprise, d'une énergie qui se dépense tout entière à transformer l'écrit en parole. C'est la sensation directe d'un travail. » (G. A. Vachon. 1974. « Le colonisé parle ». *Études françaises* 10/1, p. 66-67)

34 Par *néographie phonétisante*, il faut entendre : « sans jugement de valeur, ni positif, ni négatif, des graphies qui s'écartent délibérément de la norme orthographique. Ce caractère délibéré se manifeste par la saillance de procédés tels que l'abréviation, la simplification phonétisante, la transcription de prononciations s'écartant du français soutenu, etc. » (Anis 1999, p. 86)

35 Pour d'autres phénomènes, voir Dargnat (2002b [3]), (2006 [6], chap. 3) et (2007 [7]).

Phénomène non standard	représenté	non-représenté	exemples-types
• arrondissement en [ɔ/o]	X		popa
• fermeture en [e]	X		pétates
<b>Diphthongaison</b>		X	
<b>Prononciations du graphème standard « OI »</b>			
• [we]	X		moé
• [wɛ]		X	
• [wa] ou [wɔ]		X	
• [e]	X		j'cré
• [ɛ]	X		frette
• [ɔ]	X		mognons
<b>CONSONNES</b>			
<b>Affrications de [t] et [d]</b>			
		X	
<b>Prononciation d'un [t] final</b>			
• orthographique	X		litte, bouttes
• non orthographique	X		icitte, légerte
<b>Réductions des groupes consonantiques</b>			
• à l'initiale ou interne au mot	X		pus, quequ'chose
• en finale de mot	X		vot', sarcasses
<b>Désonorisation</b>			
• de [ʒ] en [ʃ]	X		ch'suis, chus
• de [g] en [k]	X		fatiqué
<b>Palatalisation/yodisation</b>			
• yodisation de [d] + [i,j]	X		yable, bonyeu
• yodisation de ([ɔ]+) [g]	X		yeule, cataloye
<b>Hésitations circonscrites</b>			
• entre [l] et [n]	X		canneçon
• entre [l] et [r]		X	

Tableau 1. Présence de néographies phonétisantes dans les cinq pièces

(ii) Concernant les contraintes de catégorisation linguistique et symbolique, un des paramètres qui pèse sur l'interprétation des phénomènes perçus est celui du décalage des registres à l'oral et à l'écrit, représenté dans la Figure 2.

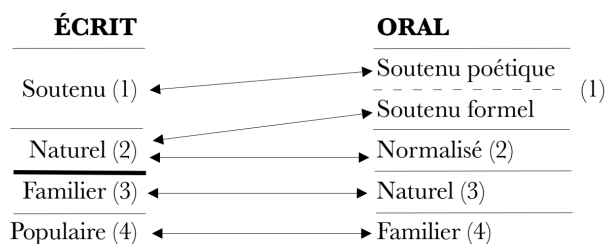


Figure 2. Décalage des registres à l'oral et à l'écrit (d'après Anis 1981, p. 20).

L'idée de base est celle d'un décalage de perception lors du passage de l'oral à l'écrit. Par exemple, ce qui est perçu comme naturel à l'oral devient familier à l'écrit, ce qui est familier à l'oral devient populaire à l'écrit, et vice-versa. L'approche intuitive de Anis reste un peu floue sur l'identité des « formes » (quelles formes ?) et sur ce qu'il faut entendre par *oral* et *écrit* (canal, variété ?). Cependant, la Figure 2 est très utile. A un niveau général, elle souligne le rôle complexe et ambigu de la norme et de l'écrit, ainsi

que celui des formes linguistiques dans la cognition sociale<sup>36</sup> (en particulier l'effet produit par la perturbation volontaire de la norme orthographique). A un niveau plus local, elle apporte un cadre pour expliquer certains choix de Tremblay et comprendre ce qu'on pourrait appeler l'édulcoration de son imitation sur certains aspects pourtant très spécifiques de l'OPQ, sa « recollection sélective » (Bourdieu 1983)<sup>37</sup>. Finalement, pas besoin de transposer fidèlement toutes les caractéristiques linguistiques de l'OPQ, quelques particularités orales (naturelles) et familières suffiront pour faire « sombrer » les personnages dans le populaire.

(iii) Concernant enfin les contraintes liées au genre dramatique, je renvoie à ce qui a été dit plus haut sur le genre et le style (section 1.1.3, p. 25) et sur la démarche particulière de Tremblay dans l'agencement polyphonique et la distribution des « parlures » entre les personnages, et même entre différentes facettes d'un personnage (section 1.1.2, p. 27).

L'étude des saillances (écarts perçus) sur lesquelles repose en partie l'évaluation stylistique de l'oralité chez Tremblay a nécessairement été multiple. Elle a supposé des comparaisons à plusieurs niveaux, qui s'enchevêtrent parfois : écarts entre la réalité et l'imaginaire d'un français « standard » et d'un complexe variationnel (OPQ) ; écarts entre la réalité linguistique associée à l'OPQ et la manière dont Tremblay la représente ; écarts entre plusieurs manières dont Tremblay représente l'OPQ. Ce qui était nouveau dans mon travail n'était pas vraiment son domaine – la question de l'oral représenté n'est pas nouvelle en littérature –, mais ce sont sa systématisme, sa dimension sémiotique et son approche stylistique diachronique. L'apport réside dans l'étude de l'évolution de l'interprétation de cette imitation sur une période de 30 ans et de son articulation avec le contexte.

La section suivante présente les enjeux principaux de la démarche méthodologique que j'ai construite pour mener une analyse outillée fine des données linguistiques qui prenne en compte aussi leur dimension textuelle.

### 1.3 Méthodologie de corpus : donner du sens au quantitatif

Sur quelles données baser l'analyse de la représentation de l'oral ? Comment traiter les données pour les exploiter de manière systématique et automatique et servir les hypothèses avancées, qui partent de phénomènes saillants et cherchent à montrer leur fonction référentielle (leur valeur réaliste) et leur fonction de structuration énonciative (à qui ces phénomènes sont-ils attribués dans les textes) ?

36 Par *cognition sociale*, on cible les ressources et les processus cognitifs qui permettent à un individu de faire sens de l'environnement social. Le langage met en jeu des relations de dominance et de prestige. Le locuteur peut chercher à récupérer à travers ses usages linguistiques une partie du prestige du groupe valorisé. Inversement, un usage linguistique jugé déviant par rapport à un standard peut marginaliser un individu.

37 La sélection de certains traits peut avoir aussi comme effet de les hypertrophier. Cet effet à la fois de loupe et de distanciation est bien connu dans le domaine artistique, pour mettre au premier plan le processus de création de l'effet de réel plus que le réel imité lui-même (voir le *Verfremdungseffekt*).

Pour répondre à ces questions, j'ai dû construire un corpus pertinent, puis développer une analyse outillée des données qu'il comporte. Cela supposait une numérisation, un encodage et un environnement logiciel (section 1.3.1) ainsi qu'un traitement quantitatif et son interprétation qualitative (section 1.3.2). Une telle démarche permet de confirmer empiriquement l'intuition d'une évolution de la représentation étudiée.

En mettant l'accent sur un aspect méthodologique de mes recherches, je veux montrer l'intérêt qu'il y a à étendre le domaine de la textométrie et de la linguistique de corpus à des données que l'on peut dire récalcitrantes à l'analyse automatique. Une telle méthodologie qui interroge « en aveugle » la qualité d'une écriture et d'un texte rencontre l'analyse stylistique et littéraire. Bien loin de détruire l'effet littéraire, elle permet de l'expliquer, de le sonder et de le mettre en évidence. Par souci de cohérence, je me restreins dans cette section à ce qui a été fait sur l'oral représenté, mais, dans des travaux plus récents, je continue à faire appel à la méthodologie de corpus et aux approches quantitatives, en particulier pour étudier les combinaisons de marqueurs discursifs (voir section 3.2.4, p. 149).

### 1.3.1 Un « sac » de données compartimenté

Les données sélectionnées n'ont pas été choisies au hasard. Il ne s'agit pas d'un gros « sac de mots » (Rastier 2005, p. 1, Valette & Slodzian 2008, p. 122) où se trouveraient pêle-mêle des éléments disparates et recueillis aléatoirement. Le corpus, comme ensemble de données organisées, a été pensé comme pertinent par rapport au phénomène étudié et à la problématique, qui nécessitait un relevé fin exhaustif de propriétés ou d'items dans une logique de comparaison. Les contraintes à respecter étaient les suivantes :

- un ensemble de textes homogènes du point de vue du genre (textes dramatiques), répartis sur une période ayant un sens, et en nombre suffisant mais pas excessif pour que le traitement soit possible,
- des données linguistiques authentiques correspondant à la variété de français oral populaire pour la période donnée (des locuteurs correspondant à la provenance fictionnelle des personnages dans les pièces),
- un format numérique et un traitement les rendant comparables.

#### 1.3.1.1 Constitution des corpus tremblay et frcapop

La visée comparative nécessitant deux types de données (transcriptions d'entretiens authentiques et textes dramatiques), il y a donc eu toute une phase de sélection, en fonction de critères explicites jugés pertinents, mais aussi de contraintes matérielles, comme l'accès ou la qualité des données.

La dynamique de constitution du corpus littéraire est liée au fait que deux pièces se font volontairement écho à trente ans d'intervalle : *Les Belles-sœurs* (publiée une première fois en 1965 et montée en 1968) et *Encore une fois, si vous permettez* (montée et publiée en 1998). Elles ont donc constitué les textes-bornes. Ensuite, il a fallu chercher les meilleures « pièces à conviction » (Boulanger 2001) pour montrer une évolution dans la représentation de l'OPQ. Les critères ont été les suivants :

- une périodisation des trente ans la plus équilibrée possible,
- des pièces montées au théâtre, publiées et reçues par la critique,

- des pièces ne provenant pas uniquement du Cycle des *Belles-sœurs*,
- des textes affichant toutes des marques d’OPQ, mais un usage sensiblement différent, dans la forme ou la distribution,
- des textes structurés différemment du point de vue énonciatif (diffraction d’un personnage, mise en parallèle de dialogues, schéma musical, mise en abyme) (voir ici Figure 1, p. 23).

Le tri a abouti à la sélection suivante : *Les Belles-sœurs* (édition de 1965<sup>38</sup>), *Bonjour, là, bonjour* (1974), *L’Impromptu d’Outremont* (1980), *Le Vrai monde* (1987) et *Encore une fois, si vous permettez* (1998). Seuls les textes imprimés étaient disponibles, il n’a pas été possible de les récupérer sous format numérique.

Concernant les données linguistiques, il existe plusieurs corpus de français québécois, utilisés comme corpus de référence dans les études consultées, mais ils n’étaient pas forcément disponibles (voir Blondeau et al. 2021). Grâce à Pierrette Thibault, j’ai pu bénéficier d’un accès privilégié au laboratoire d’ethnolinguistique à l’Université de Montréal, en particulier aux enregistrements, aux transcriptions et aux métadonnées des corpus *Sankoff-Cedergren* et *Montréal 84*<sup>39</sup>. Les deux ont été faits sur le même modèle. Il s’agit de deux enquêtes sociolinguistiques qui consistent en des entretiens semi-dirigés de longueurs très variables (de moins d’une heure à plusieurs heures). Commandés et subventionnés par le gouvernement québécois, leur visée était de décrire le français parlé par les francophones de l’agglomération montréalaise. La totalité des enquêtés représente les différentes classes d’âges, les deux sexes, et toute l’échelle des milieux socio-professionnels. La particularité de ces deux corpus est de se répondre à treize ans d’intervalle, l’objectif du travail de 1984 étant de sonder l’évolution linguistique de locuteurs déjà interrogés en 1971<sup>40</sup>. J’ai constitué un corpus de référence pour l’OPQ à partir de ces deux importants corpus, en cherchant la meilleure représentativité possible. La sélection des entretiens s’est faite principalement sur critères sociolinguistiques en utilisant les profils des locuteurs-cibles (L2 dans les transcriptions). Ces profils sont déductibles des fiches signalétiques de chaque locuteur et des tableaux synthétiques consultables avec chaque corpus. Les contraintes de représentativité retenues étaient les suivantes :

- parité des sexes,
- « cote de marché linguistique »<sup>41</sup> basse correspondant à l’image du populaire (classe ouvrière urbaine peu scolarisée),
- locuteur originaire de Montréal et résidant à Montréal dans les quartiers dits populaires de l’époque (en particulier le plateau Mont-Royal),
- locuteur francophone de langue maternelle,

38 Il existe deux éditions du texte : 1965 chez Holt, Rinehart et Winston, avant sa création scénique, et 1971 chez Leméac. Il existe quelques différences dans les marques d’OPQ, et à la césure entre les deux actes. Le principe a été de conserver les premières éditions des textes.

39 Le premier a été réalisé par D. Sankoff et H. Cedergren et date de 1971 (voir Sankoff et al. 1976), le second a été collecté par D. Vincent et P. Thibault en 1984 (voir Thibault & Vincent 1990).

40 Pour un descriptif plus complet, voir le paragraphe qu’y consacre C. Blanche-Benveniste et C. Jeanjean (1987, p. 50-52), et le descriptif très complet de P. Thibault et D. Vincent (1990). Les questionnaires des entretiens et des jugements sur la langue sont reproduits dans Dargnat (2006 [6] Annexes, p. 51-53).

41 Voir Laberge (1977, p. 41) et Sankoff & Laberge (1978). Cette donnée n’était disponible que pour le corpus *Sankoff-Cedergren*.

– représentativité des classes d'âges et des deux corpus.

J'ai finalement récupéré 22 entretiens transcrits dans un format numérique de type traitement de texte, avec des normes de transcription assez sommaires mais relativement systématiques et une conservation de la norme orthographique. J'ai pu récupérer également les fichiers sons correspondant<sup>42</sup>.

Ces deux ensembles sélectionnés ont été considérés comme deux corpus, le corpus *tremblay* pour les pièces et le corpus *frcapop* pour les entretiens linguistiques. Afin de pouvoir les exploiter, une phase de prétraitement et d'uniformisation des données a été nécessaire pour les récupérer dans un format textuel basique (.txt). Cette étape, très chronophage, a le mérite d'obliger à une connaissance quasi-parfaite des données. Le problème de l'uniformisation de données hétérogènes est un problème récurrent que je rencontre encore dans d'autres contextes quand je fais appel à plusieurs corpus déjà constitués<sup>43</sup>.

### 1.3.1.2 Traitement XML et textométrie

Une fois numérisées, les données ont pu être structurées en fonction des hypothèses et des langages disponibles pour les structurer. L'objectif principal était de pouvoir isoler le matériau linguistique des locuteurs-cibles (les personnages dans les pièces de théâtre, L2 dans les entretiens) en laissant de côté, sans les perdre pour autant, les autres interventions (didascalies, interventions de l'enquêteur).

Cette partie du travail a pu être effectuée en collaboration avec l'équipe de l'École Normale Supérieure de Lyon, à l'occasion d'un séjour scientifique pour le traitement de mes données, leur intégration et leur exploitation au moyen de l'interface *Weblex* (Heiden 2002). La structuration des deux corpus a été faite avec le langage de balises XML (*Extensible Markup Language*) en suivant et en adaptant le principe de la TEI (*Text Encoding Initiative*)<sup>44</sup>. Ainsi, les documents structurés sont éditables, échangeables et interrogeables dans des environnements courants et encore largement partagés aujourd'hui, par exemple celui de *TXM*<sup>45</sup> (Heiden et al. 2010).

L'intérêt principal de cette démarche pour l'analyse des deux corpus dans le cadre d'une étude des marques d'OPQ a été de pouvoir structurer les textes de manière hiérarchique, de les interroger finement et systématiquement<sup>46</sup> en sélectionnant tel ou tel locuteur ou personnage et de comparer les résultats. Les données étant non normatives, il n'a pas été possible d'utiliser d'étiqueteur en parties du discours comme *Treetagger*, le résultat aurait été catastrophique. Le problème serait identique aujourd'hui

42 Les fichiers sons n'ont pas été utilisés pour l'étude phonétique et prosodique, mais ils ont servi de référence ponctuelle. Chaque transcription d'entretien a été écoutée.

43 Voir les travaux sur la prosodie des marqueurs discursifs (par ex. Bartkova et al. 2017a [27]), la comparaison de structures discursives et argumentatives (par ex. Huber et al. 2019c [32]) et sur les combinaisons de marqueurs (par ex. Dargnat 2022a [42]).

44 Les recommandations et les exemples sont disponibles en ligne et sont présentées par types de documents, qui rejoignent globalement les genres de discours et/ou le format des données. Il existe également une TEI simplifiée, ou TEI Lite, produite par Burnard & Sperberg-McQueen, et traduite en français. <https://tei-c.org>

45 L'interface de *Weblex* n'est plus accessible aujourd'hui, mais mes données sont transposables sans difficulté pour être exploitées avec son successeur *TXM*, qui accepte plusieurs formats, dont les formats XML. Une partie du corpus *tremblay* a d'ailleurs été transférée pour la démonstration dans le cadre d'un cours.

46 La syntaxe d'interrogation de *Weblex*, comme celle de *TXM* actuellement, est basée sur CQP, pour *Corpus Query Processing*. Cette syntaxe a été développée initialement à l'Université de Stuttgart. Pour un descriptif actuel : <http://cwb.sourceforge.net/documentation.php>



et c'est un problème constant pour le traitement de corpus « oralisants ». C'est en cela que j'ai parlé plus haut de données récalcitrantes. Il a donc fallu trouver des moyens détournés pour extraire des patrons syntaxiques, par exemple les dislocations, les interrogatives, etc. et revenir parfois, dans un deuxième temps, à un tri manuel. Les contraintes typographique et lexicale sur les requêtes ont donc été très fortes, celles-ci ne pouvant se faire que sur la base d'expressions régulières mentionnant la présence ou l'absence de signes, de mots ou de suite de mots. Toutes les analyses évoquées dans la section 1.2.2 (p. 33) ont été rendues possibles grâce à cette méthodologie.

Dans la mesure où j'ai constitué, annoté et exploité le corpus avec des hypothèses préalables et des requêtes ciblées, on peut dire que la démarche première a été une démarche *top-down*. Dans cette perspective, l'approche informatique est un outil, un soutien pour interroger les données et les textes de manière systématique, quantitative et comparative et produire des relevés exhaustifs, ce qu'une étude intuitive et manuelle ne pourrait atteindre que très difficilement. Mais, en réalité, l'informatique n'est pas qu'un outil, son utilisation aide aussi en amont à penser la structure des données et les limites des phénomènes interrogeables ; les résultats obtenus vont parfois au-delà de la simple preuve des hypothèses testées et d'une lecture mécanique. Ils ouvrent d'autres perspectives, permettent des distinctions qui n'étaient pas perçues. Dans ce cas-là, l'approche devient davantage *bottom-up*. Le caractère herméneutique se double d'une « plus-value heuristique » (voir Habert et al. 1997, p. 184 sqq., Mayaffre 2006, p. 8, Rastier 2005, chap. 3).

### 1.3.2 L'effet personnage-locuteur et le calcul des profils linguistiques

Dans le domaine littéraire, les différences linguistiques entre les personnages sont souvent un élément exploité pour construire l'identité des personnages, et même parfois pour structurer l'action. Le théâtre en est une illustration prototypique. L'« effet-personnage », d'abord défini pour le roman (Jouve 1992), s'applique tout naturellement au texte dramatique, pour lequel on peut même parler d'un *effet-personnage-locuteur* (Dargnat 2010a [13], p. 10). L'idée que le personnage est une image mentale probabiliste faite d'hypothèses que le lecteur se construit à partir de stimuli textuels correspond à ma problématique d'identification des personnages par leur profil linguistique. Cette image prend forme en s'appuyant à la fois sur la connaissance qu'a le lecteur du monde qui l'entoure et sur sa capacité à faire des liens intertextuels. J'ai considéré que les marques d'oralité étaient des stimuli textuels contribuant à cet effet personnage-locuteur, qui serait l'image d'un sujet parlant autonome que se construit le lecteur à partir de ce qu'il lit, en fonction d'une réalité linguistique qu'il connaît plus ou moins (le français montréalais), mais aussi en fonction des habitudes de transcription du parlé spontané qu'il s'est forgées et du décalage des registres (voir ici Figure 2, p. 35) auquel il a été confronté en fréquentant d'autres textes. Dans cette perspective, l'effet personnage-locuteur est fait d'un mélange de projections et de souvenirs de paroles et de parlures déclenchés par les stimuli textuels que sont ici les marques d'OPQ. L'illusion consiste à faire croire à l'autonomie discursive de segments textuels extraits d'une production unique, celle de l'auteur, à faire d'une simple « fonction sémiologique d'énonciateur » un véritable locuteur (Ducrot 1984, p. 106).

De ce point de vue, comment exploiter au mieux les données balisées pour servir les hypothèses de départ ? Globalement, le premier intérêt était de pouvoir comparer systématiquement la langue des locuteurs réels et celle des personnages des pièces de théâtre, en faisant des requêtes assez fines ou, sur la base du lexique complet, en faisant la liste des éléments saillants, c'est-à-dire dont l'orthographe était déviante. Cette comparaison systématique a mis en évidence les choix de l'écrivain, et certaines habitudes d'intégration de l'oral dans l'écrit. En utilisant davantage les éléments de structure des textes dramatiques balisés, il a été possible de différencier ou de rapprocher les personnages entre eux, à l'intérieur d'une pièce ou entre les pièces, et de faire correspondre à des ressemblances et des différences des éléments de la structure dramatique, actantielle et énonciative des textes. La voie choisie pour montrer ces différences, leur fonction textuelle et fictionnelle est celle des profils linguistiques et de leur significativité statistique.

J'ai proposé l'idée d'*espace discursif* pour désigner le matériau verbal constitué par les paroles des personnages et l'idée des *profils linguistiques* pour classer les personnages selon que leur matériau linguistique présentait plus ou moins de marques d'OPQ. Chaque texte a été défini comme un espace discursif, c'est-à-dire comme un monde de paroles échangées par les personnages. Et chaque personnage a pu être quantifié en nombre de mots et nombre de prises de parole. Le corpus *tremblay* a d'abord été considéré comme un macro *espace discursif* divisible en sous-espaces discursifs (par pièce, par acte ou division de texte, par groupe de locuteurs, par locuteur, et même par sous-locuteur). Comme le montre la Figure 3 ci-dessous, les cinq pièces sont d'une taille assez comparable en termes de nombre de mots. Autrement dit, les tailles des sous-corpus du corpus *tremblay* sont assez homogènes, ce qui constitue plutôt un atout pour l'approche statistique.

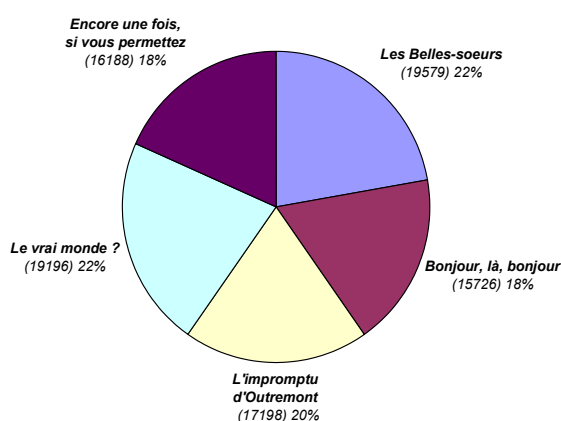


Figure 3. Espaces discursifs des pièces dans le corpus *tremblay* (nombre de mots et %)

En revanche, le nombre de personnages et la répartition du matériau linguistique entre eux à l'intérieur de chaque pièce varient beaucoup. Les deux figures ci-dessous

montrent la répartition pour les deux pièces-bornes, *Les Belles-sœurs* et *Encore une fois, si vous permettez*.

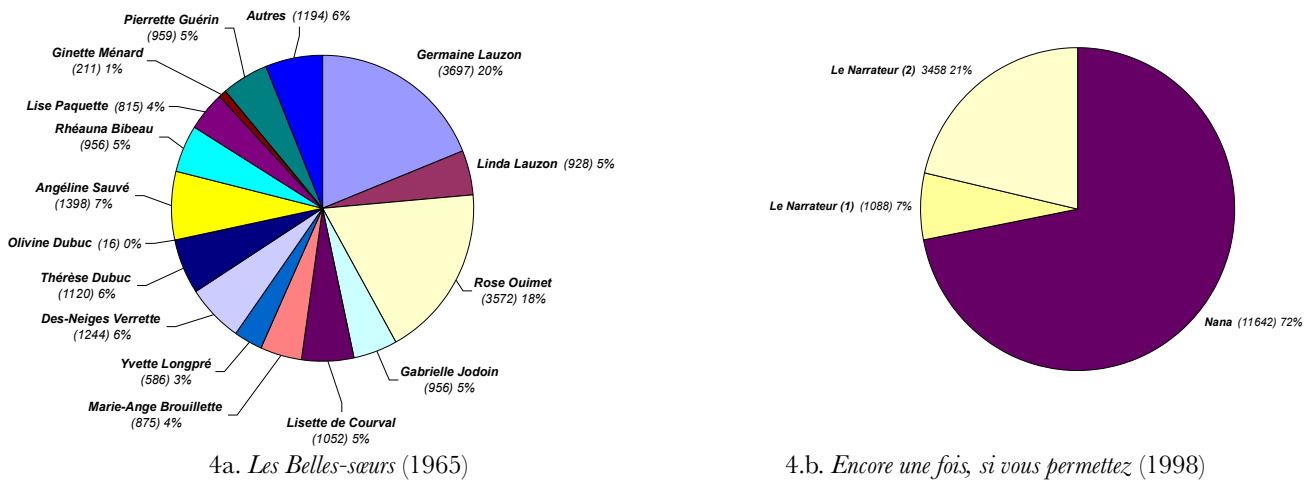


Figure 4 . Espaces discursifs des personnages dans les pièces (nombre de mots et %)

Les trois autres pièces présentent également des particularités. Globalement, on repère une différence dans le nombre de personnages-locuteurs, une disproportion ou une équivalence des espaces discursifs alloués à chaque locuteur, et la nécessité d'une catégorie « autre » pour désigner les prises de parole collectives. Ces particularités de taille et de répartition correspondent parfois à des rôles actanciels ou à une position énonciative dans la structure du texte. Je ne détaille pas ce point et présente seulement le profilage des personnages, qui visait à trouver des corrélations entre la qualité +/- OPQ de la parole des personnages et leur rôle et leur position.

Le profil linguistique est une représentation de chaque personnage comme un vecteur constitué de propriétés associées à des phénomènes linguistiques jugés rentables, pertinents ou représentatifs pour faire effet d'OPQ. Je n'ai pas pris en compte l'ensemble des phénomènes exposés à la section 1.2.1.1 (p. 28), ni même ceux listés à la section 1.2.2 (p. 33), car les profils se devaient d'être restreints et efficaces. Les contraintes pour les composer ont été les suivantes :

- la possibilité de comptage des phénomènes retenus pour chaque personnage dans l'ensemble du corpus *tremblay*, avec les fonctionnalités de *Weblex* sur la base de l'annotation XML. Les phénomènes ciblés doivent être « attrapables »,
- une couverture des différentes composantes linguistiques (prononciation, syntaxe, lexique).

Concrètement, les variables retenues étaient :

1. Phonographie	2. Morpho-syntaxe	3. Lexique
<p><b>a.</b> <i>pis</i></p> <p><b>b.</b> <i>ben</i></p> <p><b>c.</b> <i>toé, moé</i></p> <p><b>d.</b> ouverture de [ɛ] en [ɑ] devant [ʀ] et ses variantes (ex. <i>farmer, narveuse</i>)</p>	<p><b>f.</b> <i>à et dans + N</i> (ex. <i>à'soir, dans'cuisine</i>)</p> <p><b>g.</b> marqueur discursif <i>ça fait que</i></p> <p><b>h.</b> particule interrogative et exclamative <i>tu</i> (ex. <i>c'est-tu beau ! A'vient-tu ?</i>)</p>	<p><b>j.</b> <i>sacres</i> (ex. <i>crisse, câlisse</i>)</p> <p><b>k.</b> québécoisismes (ex. <i>garrocher, magané</i>)</p> <p><b>l.</b> anglicismes (ex. <i>pinottes, grocerie</i>)</p>

<b>e.</b> apostrophes devant consonnes (ex. <i>v'nir, à'veulait</i> )	<b>i.</b> négation en <i>pas</i> sans <i>ne</i> (ex. <i>j'en veux pas</i> )	
---	---	--

Tableau 2. Liste des variables pour le profilage OPQ

Les profils peuvent paraître succincts, mais ils résultent de plusieurs cycles d'essorage, pour reprendre une image déjà employée, qui avait pour but d'accroître leur pertinence référentielle et littéraire. Ils agissent comme des traceurs d'OPQ. Un relevé brut de leur présence ou de leur absence dans le discours de tel ou tel personnage est certes intéressant mais demeure insuffisant compte tenu de la disparité des espaces discursifs. Une expression en pourcentage donne des résultats plus représentatifs mais ne dit rien de la significativité des différences observées. On peut illustrer cela en partant des données brutes de *L'Impromptu d'Outremont*. Le Tableau 3 présente le nombre d'occurrences pour chaque variable recensée dans le Tableau 2 (colonnes *a* à *l*), ainsi que le nombre total de ces occurrences (colonne *total*) et leur proportion (colonne *%*) par rapport au nombre total de mots pour chaque personnage (colonne *nb.mots*).

<b>IO</b>	<b>nb.mots</b>	<b>a.</b>	<b>b.</b>	<b>c.</b>	<b>d.</b>	<b>e.</b>	<b>f.</b>	<b>g.</b>	<b>h.</b>	<b>i.</b>	<b>j.</b>	<b>k.</b>	<b>l.</b>	<b>total</b>	<b>%</b>
Fernande	6 238	1	0	0	0	16	0	0	0	24	0	13	14	68	1,09
Lorraine	3 241	23	8	0	0	46	3	0	4	40	0	12	9	145	4,47
Lucille	4 826	2	0	0	0	45	0	0	0	44	0	8	16	115	2,38
Yvette	2 891	1	0	0	0	14	0	0	0	29	0	5	2	51	1,76
Total IO	17 196	26	8	0	0	121	3	0	4	137	0	38	41	379	2,2

Tableau 3. Relevé des occurrences des variables OPQ par personnage dans *L'Impromptu d'Outremont*

Une fois le profil OPQ exprimé pour chaque personnage dans les cinq pièces<sup>47</sup>, on peut tester si les différences de distribution observées sont significatives ou non, c'est-à-dire dues ou non au hasard. En guise d'illustration, je présente brièvement les résultats du test de Fisher pour les données du Tableau 3. Il s'agit d'un test très simple qui correspond à des situations d'analyse courantes en stylistique littéraire et sociolinguistique, dans la mesure où il permet aisément de tester la significativité de différences entre des locuteurs, des auteurs ou des personnages. Le test de Fisher est le test le plus fiable pour ce type de données. Le principe est de comparer les scores de deux catégories (ici deux personnages) par rapport à un critère binaire, ici le fait qu'un mot est une marque d'OPQ ou pas. Par exemple, pour le personnage de Fernande, sur l'ensemble de son espace discursif en nombre de mots (6238) on a 68 marques d'OPQ toutes variables confondues, et 6238-68 autres marques. Pour Lorraine, on a 145 marques d'OPQ et 3241-145 autres marques, etc. Le test de Fisher prend en entrée le tableau suivant :

	Fernande	Lorraine
+OPQ	68	145
- OPQ	6 170	3 096

Tableau 4. Données des personnages Fernande et Lorraine pour le test de Fisher

Le test calcule ensuite la probabilité d'observer une distribution telle que celle du Tableau 4. Dans l'exemple, en sortie, on obtient une probabilité quasiment nulle, ce qui

<sup>47</sup> Compte tenu de la petite taille du sous-corpus *tremblay* et des faibles occurrences de certains phénomènes, il a fallu regrouper les scores pour toutes les variables du Tableau 2.

revient à dire que la possibilité que la répartition des marques d’OPQ observée entre Fernande et Lorraine soit due au hasard est très faible. Ce test est exécuté pour chaque couple de personnages. Pour être lisibles, les résultats sont présentés sous la forme de tableaux à double entrée dans lesquels les personnages sont comparés deux à deux<sup>48</sup>. L’intersection d’un personnage avec un autre a été notée 1 quand la différence était significative au seuil de 5 %, c’est-à-dire quand la probabilité renvoyée par le test était inférieure ou égale à 5 %. Quand rien n’est noté (case vide), cela signifie que le test renvoie une probabilité supérieure à 5 %. C’est ce qui est présenté ci-dessous pour deux pièces, *L’Impromptu d’Outremont* (Tableau 5a) et *Encore une fois, si vous permettez* (Tableau 5b).

	Fernande	Lorraine	Lucille	Yvette		Nana	Le Narrateur	N1	N2
Fernande		1	1	1	Nana		1	1	
Lorraine	1		1	1	Le Narrateur	1		1	1
Lucille	1	1			N1	1	1		1
Yvette	1	1			N2		1	1	

5a. *L’Impromptu d’Outremont*5b. *Encore une fois, si vous permettez*

Tableau 5. Résultats du test de Fisher pour le profilage des personnages dans *L’Impromptu d’Outremont* et dans *Encore une fois, si vous permettez*.

Ainsi, on arrive à déterminer des profils + ou – OPQ, qu’il faut comprendre comme une catégorisation sur un continuum, dans la mesure où tous les personnages présentent au moins une variable. Dans *L’Impromptu d’Outremont*, trois groupes de personnages émergent : Fernande qui correspond au pôle /– OPQ/ et Lorraine au pôle /+ OPQ/. Yvette et Lucille sont associées à un profil intermédiaire. Le cas de *Encore une fois, si vous permettez* est plus complexe. Il n’y a en réalité que deux personnages déclarés dans le texte, Nana et Le Narrateur. N1 et N2 renvoient à la décomposition du personnage Le Narrateur en fonction de son interlocuteur : le public pour N1, Nana pour N2. Cette décomposition, codée dans le fichier XML, a pu faire l’objet d’une analyse statistique, qui a révélé que N1 et N2 avaient des profils OPQ très différents. Ainsi, on peut déterminer trois groupes : N2 et Nana au profil /+OPQ/, N1 au profil /–OPQ/, et, pour retrouver le personnage Le Narrateur, un profil hybride, collage de N1 et N2.

L’analyse aurait pu s’arrêter là puisque les variables ont fait ressortir plusieurs profils linguistiques, ce qui montre l’efficacité et le rôle structurant de certaines marques d’OPQ. C’est un premier niveau d’articulation du quantitatif et du qualitatif. J’ai choisi d’aller plus loin en connectant systématiquement les profils linguistiques et la structure actancielle, en comparant l’analyse de chacune des pièces pour mettre en évidence l’évolution de l’usage de l’OPQ chez Tremblay, et en connectant cette évolution à l’évolution des modèles esthétiques du théâtre québécois. Je ne reprends pas ici cet aspect socio-critique, qui est cependant une composante essentielle de l’orientation sémiostylistique que j’ai adoptée et qui justifie la pertinence des données littéraires et linguistiques sélectionnées.

48 La comparaison Lorraine vs Fernande équivaut à la comparaison Fernande vs Lorraine.

## Synthèse 1

Dans cette première partie, qui correspond chronologiquement au premier « moment » de mes travaux, j'ai présenté la façon dont un texte littéraire reflète et construit une image de la variation linguistique. Cela m'a obligée à mener de front et à articuler deux stratégies qui relevaient de deux champs disciplinaires distincts, la description linguistique empirique et les études littéraires.

Pour les conduire, j'ai constitué deux corpus de référence, le corpus *ficapop* et le corpus *tremblay*. Le premier a servi à catégoriser et à décrire la variété « oralité populaire québécoise » (OPQ) dans une perspective sociolinguistique. Dans le second, j'ai repéré de manière systématique les phénomènes linguistiques censés évoquer la variété en question.

D'un point de vue théorique, j'ai utilisé les notions linguistiques qui me semblaient utiles pour circonscrire la variété OPQ et les notions que m'offraient la théorie littéraire et la stylistique pour aborder la question de la représentation des pratiques linguistiques et son ancrage contextuel (culture et société québécoises, évolution des formes littéraires).

Afin d'étayer les pistes de lecture que l'approche théorique pouvait me suggérer par rapport à la représentation de l'OPQ dans le corpus littéraire, j'ai construit une grille d'analyse me permettant de « profiler » les personnages en fonction de la présence de phénomènes qui me semblaient pertinents dans les textes (analyse qualitative). J'ai mis cette grille à l'épreuve d'un test statistique robuste pour évaluer son efficacité (analyse quantitative).

Même si ces travaux sont les plus anciens, ils occupent toujours une place importante dans ma pratique d'enseignant-chercheur, puisque j'y ai recours régulièrement et que j'ai parmi mes perspectives pour les prochaines années d'étendre ce type d'analyse à d'autres données (section 4.2, p. 163). J'ai aussi su faire évoluer l'aspect méthodologique (linguistique de corpus et articulation analyse qualitative/analyse quantitative) dans des travaux plus récents sur les marqueurs discursifs, sur leur prosodie (section 3.2.3, p. 142) et sur leur combinaison (section 3.2.4, p. 149).



# 2

## Relations de discours

*Lorsqu'un énoncé défile devant l'auditeur qui cherche à le comprendre, ses mots ne déversent pas l'un après l'autre, comme les godets dans la chaîne d'une noria, leur contenu individuel, qui viendrait s'ajouter au contenu véhiculé par les précédents.*

Oswald Ducrot et al., *Les mots du discours*. Paris, Les Éditions de Minuit, 1980, p. 9.





Cette partie correspond à une deuxième étape dans mon parcours scientifique. A partir des années 2008, bien que conservant un intérêt particulier pour les supports littéraires et la problématique de la représentation stylisée de la variation linguistique, et bien que restant dans le domaine de l'analyse du texte et du discours, je me suis progressivement focalisée sur la description de phénomènes considérés habituellement comme plus linguistiques. Le français québécois, les contraintes génériques et la question de l'évolution des formes ressurgissent cependant, certes différemment mais régulièrement, à travers des exemples, des types de données et une dimension diachronique. Il s'agit donc d'un tournant, mais pas d'une rupture, dû à des choix scientifiques personnels, mais aussi à un changement de cadre institutionnel et à de nouvelles collaborations.

Cette étape couvre une période assez longue, puisque elle est toujours d'actualité. Les travaux qui y sont associés sont divers et explorent plusieurs dimensions, mais j'ai souhaité les rassembler ici car, d'une manière ou d'une autre, ils concernent tous le champ des relations de discours. Les questions traitées sont donc à la fois plus générales et plus ciblées que celles présentées dans la partie précédente. Plus générales, car elles ne se cantonnent pas à un corpus déterminé, construit et clos, et ne cherchent pas à mettre en évidence un fonctionnement stylistique singulier. Plus ciblées, car elles traitent d'un phénomène (en apparence) plus circonscrit, les relations de discours, avec une insistance sur les relations sans marque lexicale. Si je reviens au titre de la présente synthèse, *Lexique et discours*, la balance penche assez nettement ici du côté du discours, mais sans s'interdire quelques contrepoids lexicaux.

Globalement, on peut définir les relations de discours, ou relations de cohérence, comme des relations sémantiques et pragmatiques établissant entre des segments discursifs des liens comme la cause, la condition, l'opposition, etc. Quand elles ne sont pas marquées lexicalement par un connecteur (par ex. *donc, pourtant, si, autrement dit*), la question est de savoir si elles peuvent être inférées sur la base d'autres propriétés, et s'il est possible de décrire les processus de choix ou de préférence en cas d'interprétations multiples. Par exemple, dans (1), les deux segments entre crochets sont reliés par une relation qui peut être conditionnelle, temporelle ou consécutive, que l'on pourrait matérialiser avec des connecteurs comme *si, quand/à chaque fois que* ou *donc*. Dans (2), l'énoncé n'est pas perçu comme incohérent et la relation est généralement décodée comme conditionnelle.

- (1) [j'ai un coup de cafard] [je mets une cassette vidéo jusque les trois quatre heures du matin] (ST I3, dans Darnat 2008b [9], ex. 8 p. 2469)
- (2) [Tu avances] [je te casse la gueule]

La section 2.1 est une synthèse de mes travaux sur les juxtapositions du type de (1) et (2). La section 2.2 présente des recherches collaboratives récentes dans le domaine de la fouille de données textuelles et de l'argumentation.

#### Références des travaux concernés

2008b [9]\*, 2008d [11], 2009 [12]\*, 2010b [14]\*, 2011 [15], 2012a [16]\*, 2012b [17],  
2012c [18], 2013 [20]\*, 2014b [22], 2015a [23]\*, 2016b [26]\*, 2019a [30], 2019c [32]\*,  
2020d [36], 2021a [39]\*

## 2.1 Constructions juxtaposées<sup>49</sup>

Pour décrire ce type de constructions et les relations de discours en jeu, j'ai utilisé plusieurs approches. J'ai également fait appel à des éléments permettant de décrire certaines de leurs propriétés prosodiques (les contours) (voir Jayez & Dargnat 2008d [11], 2011 [15], Dargnat et. al 2012a [16], Bartkova et al. 2012b [17], Bartkova & Dargnat 2013 [20], 2014b [22]). Enfin, j'ai eu recours à un cadre plus général, celui des Grammaires de construction (CxG) (Dargnat 2008b [9], Jayez & Dargnat 2009 [12], Dargnat & Jayez 2010b [14]<sup>50</sup>), d'une part, pour concevoir et représenter les phénomènes étudiés dans leur multidimensionnalité et, d'autre part, pour interroger leur caractère plus ou moins compositionnel (voir Dargnat 2015a [13]). Ces différents outils avaient pour vocation de répondre aux questions suivantes :

- Quelles sont les relations de discours utiles pour étudier l'organisation du discours ? Comment peut-on les décrire ? Et sur quels critères peut-on les attribuer ? Ces critères sont-ils exclusifs les uns des autres : est-il possible voire nécessaire de déclarer plusieurs relations de discours entre deux segments ?
- A l'oral, quand la relation de discours n'est pas marquée lexicalement par un connecteur, quelle est la contribution de la prosodie ? Existe-t-il des patrons associés à une ou plusieurs relations de discours ? Ou la contribution prosodique est-elle plus abstraite ?
- Comment rendre compte de la coalescence de propriétés diverses, par exemple morphosyntaxiques, prosodiques, sémantiques et pragmatiques qui caractérisent certaines configurations discursives, ici les juxtapositions ? Un problème important et récurrent est celui du caractère *compositionnel* ou *constructionnel* de ces assemblages de propriétés.

Les problèmes sur lesquels je me concentre ici sont : le type de dépendance entre les propositions juxtaposées (2.1.1) ; le statut des juxtapositions à interprétation conditionnelle, et en particulier les cas où la première proposition est au conditionnel ou à l'impératif (2.1.2) ; la fonction de la prosodie dans l'établissement du lien entre propositions (2.1.3).

### 2.1.1 Juxtapositions propositionnelles : propriétés générales

Admettons qu'une juxtaposition est composée de deux segments A et B. Ces segments sont mis entre crochets et identifiés dans les exemples. Pour décrire des juxtapositions de type AB, il est nécessaire de se poser la question de la catégorisation de A et de B pris séparément, du mode de dépendance entre A et B, et de la catégorisation du bloc AB.

49 J'ai d'abord parlé de *parataxe*, mais je préfère désormais utiliser le terme plus neutre de *juxtaposition*. Il désigne des agencements de propositions reliées sémantiquement mais sans que le lien ne soit marqué lexicalement. Le terme *parataxe* s'oppose à *hypotaxe* et n'implique pas chez tout le monde l'absence de marque. Pour le débat terminologique, voir Béguelin et al. (2010) et Glikman (2005).

50 Voir aussi la référence aux CxG dans certains travaux sur les marqueurs de discours présentés dans la partie 3 (sections 3.2.2.1, p. 134 et 3.2.4.2, p. 151), notamment Combettes & Dargnat (2016b [26]) et Dargnat (2022a [42]).

J'ai ciblé les juxtapositions où A et B sont des associations contiguës de propositions verbales, comme dans les exemples (1) et (2) ci-dessus, ou encore (3), (4) et (5) ci-dessous. J'ai laissé de côté le cas des phrases averbales telles les OM sentences<sup>51</sup> et les constructions corrélatives dites aussi siamoises<sup>52</sup>.

(3) [il est en troisième]<sub>A</sub> [il a pas le niveau qu'on avait à l'époque au certificat d'étude à notre époque]<sub>B</sub> (Fra80 fa17)

(4) [je m'occupais des personnes âgées]<sub>A</sub> [on les levait le matin]<sub>B</sub> (Allier mad2/6)

(5) bon [le soir elle arrivait à la maison pour lui faire faire ses devoirs]<sub>A</sub> [moi je comprenais rien]<sub>B</sub> (Fra80 fa17)

Il paraît difficile de construire des juxtapositions avec des propositions à l'infinitif, et cela se vérifie dans les corpus. On trouve de rares exemples dans des variations régionales, en particulier en français québécois, où des exemples comme (6) et (7) reçoivent une interprétation conditionnelle :

(6) Maintenant, je pense que [recommencer]<sub>A</sub> [je ferais une garde-malade]<sub>B</sub> (dans Mougeon 1995)

(7) [Jeanne avoir sorti son char]<sub>A</sub> [on serait allé à Toronto]<sub>B</sub> (dans Dostie & Léard 1985)

J'ai donc étudié des juxtapositions où A et B sont des propositions verbales finies. Dans de telles configurations, la question de la dépendance peut s'aborder au niveau syntaxique et au niveau sémantique, les deux ne coïncidant pas forcément (pour une synthèse, voir Dargnat 2021a [39], p. 1569-1572).

### 2.1.1.1 Syntaxe : dépendance et catégorisation

Syntaxiquement, on peut se demander s'il existe entre A et B une relation de type argument ou de type ajout. Dans (Dargnat 2008b [9], p. 2469-2471 et Dargnat & Jayez 2010a [14], p. 71-80), j'ai mis en évidence (i) que A est saturé, (ii) qu'il n'est pas clair que B constitue un ajout à A, et (iii) que le statut catégoriel de AB comme phrase n'est pas acquis. Je n'ai examiné que la dépendance éventuelle de B par rapport à A, mais les mêmes conclusions valent a priori pour la dépendance éventuelle de A par rapport à B.

(i) La plupart des exemples montrent en effet que le verbe de A n'attend plus de constituant, et donc que B n'est pas un argument du verbe de A, ni A un argument du verbe de B<sup>53</sup>. Certains cas peuvent cependant interroger, par exemple (8a) ou (9a), où la construction attributive de *se retrouver* fait de B un argument de A, avec une interprétation du type « se retrouver tout seuls / se retrouver deux dans la tranchée ». Si l'on change l'ordre des constituants, le verbe *se retrouver* perd sa construction prédicative (8b et 9b), et l'interprétation n'est plus la même.

(8) a. L'hiver tout le monde partait sur le continent [on se retrouvait]<sub>A</sub> [on était tout seuls]<sub>B</sub> (dans Deulofeu 1989, ex. 8b Perr 15)

⇒ on se retrouvait complètement seuls

b. on était tout seuls on se retrouvait

51 Constructions du type « One more step (and) P » (*un pas de plus (et) il tombait*). Voir Culicover (1972).

52 Constructions du type « plus P plus P » : *Plus il mange, plus il grossit*. Voir entre autres Abeillé & Borsley (2006) et Savelli (1993).

53 D'un point de vue normatif, la place de l'argument verbal est à droite du verbe, mais il est possible à l'oral de trouver des cas où l'argument intervient avant sa tête (ex. *Mille euros il lui a donné*). Cela semble plus difficile avec un argument phrastique (ex. *\*Marie est sympa Jean trouve*).

⇒ on se retrouvait complètement seuls

- (9) a. [Je me suis retrouvé]<sub>A</sub> [on était plus que 2 dans la tranchée]<sub>B</sub> (web, voir Dargnat & Jayez 2010b [14], 9a, p. 72)

⇒ on se retrouvait seulement deux dans la tranchée

- b. on était plus que 2 dans la tranchée on se retrouvait

⇒ on se retrouvait seulement deux dans la tranchée

Les précédentes manipulations font pencher ici pour une dépendance de type argumentale entre A et B. Parmi les verbes permettant cette construction prédicative non saturée, il semblerait que l'on trouve des verbes ayant pour point commun le fait que leur argument dénote le résultat du procès en cours. Ainsi, on trouve aussi *se trouver*, *se voir*, *en arriver*, *être rendu*<sup>54</sup>. Une étude complémentaire est nécessaire pour tester les contraintes. En effet, la comparaison avec d'autres verbes à complément prédicatif comme *se sentir*, *sembler*, *rester*, *avoir l'air* (ex. 10) ou des verbes sous-catégorisant des compléments, comme *se comporter*, *mesurer* (ex. 11) montre qu'en dehors des quelques exemples cités ci-dessus, il est en fait assez difficile d'avoir à gauche un syntagme non saturé.

- (10) \* [Jean semble]<sub>A</sub> [il est fatigué]<sub>B</sub>

- (11) \* [La table mesure]<sub>A</sub> [elle fait deux mètres]<sub>B</sub>

A ce stade, je dirai que la règle dans les juxtapositions de type AB est la saturation syntaxique de A (et de B), en négligeant quelques exceptions qui restent à approfondir<sup>55</sup>.

(ii) Si la relation entre A et B n'est pas argumentale, est-elle pour autant de type ajout ? Il y a une dépendance de type ajout entre A et B, si B est associé sémantiquement à A ou à un élément de A sans pour autant être catégorisé par A ou un élément de A<sup>56</sup>. Je me concentre sur les cas où B serait un candidat ajout à A. Typiquement, ce sont les ajouts dits circonstanciels, exprimant le but, la cause, le lieu, le temps, etc. L'ajout B modifie A ou un élément de A sémantiquement en apportant un certain nombre d'informations sur sa réalisation, mais, syntaxiquement, il est optionnel et ne modifie pas la catégorie syntaxique de A. Si B est ajout à A, alors AB est de la même catégorie syntaxique que A, c'est-à-dire une phrase. Dans le cas des juxtapositions ayant un correspondant sous forme de subordination ou de coordination explicites, on pourrait considérer qu'il y a simplement effacement du connecteur et que les propriétés syntaxiques sont les mêmes. Les juxtapositions sont en réalité plus contraintes et il n'est pas certain que l'on puisse se contenter de l'hypothèse de l'effacement de la conjonction ni même de celle d'une relation d'adjonction. Plusieurs problèmes émergent en effet : je rappelle celui du repérage locatif et temporel et renvoie à la section 2.1.2 (p. 63) pour un développement plus complet sur les juxtapositions à interprétation conditionnelle.

54 En français québécois, la construction avec « être rendu + proposition juxtaposée » est fréquente : *je suis rendu je ris comme Karl* (web, voir Dargnat & Jayez 2010b [14], p. 74, note 16).

55 C'est la raison pour laquelle, dans les structures de traits reproduites ici dans la Figure 5 (p. 60) et ici dans l'Annexe 1 (p. 193), la valeur du trait SLASH est donnée vide par défaut (/).

56 Chez Bonami, *ajout* désigne « le fonctionnement des dépendants qui ne sont pas sélectionnés syntaxiquement par le verbe. Les ajouts sélectionnent syntaxiquement et sémantiquement le verbe avec lequel ils se combinent. » (1999, p. 8). Voir aussi Abeillé & Godard (2021, p. xl, 31-36 et 1983-1993).

Les ajouts locatifs sont difficilement exprimables par une juxtaposition ainsi que le montre (12) qui n'est pas récupérable comme *il était assis à un endroit (d'où) on voit la rivière*. Dans ce cas, B est un ajout locatif à A, équivalent de *là*.

(12) \* [il était assis]<sub>A</sub> [on voit la rivière]<sub>B</sub>

Cela est peut-être dû au fait qu'une colocalisation d'événements n'est pas vraiment une relation de discours, au sens où il n'y a pas de relation entre le déroulement des événements eux-mêmes. D'ailleurs, en français, il n'existe pas de conjonction locative, il faut avoir recours à un ajout prépositionnel et un système anaphorique. Le problème semble plus général que celui des juxtapositions : dans les grammaires, il n'existe pas de subordinées ni de coordonnées de lieu (voir Abeillé & Godard 2021, p. 1541-1637, Riegel et al. 2009, p. 841-870) ; les inventaires des relations de discours (par ex. ceux proposés par l'école de Genève, la RST ou la SDRT<sup>57</sup>) et des connecteurs associés<sup>58</sup> ne mentionnent rien à propos de la colocalisation d'événements<sup>59</sup>. Pour construire un modèle de situation, une scène cohérente qui corresponde à la relation de deux événements ou deux situations, la colocalisation ne semble pas suffire, il semblerait qu'il faille a minima un ordonnancement temporel entre ces deux événements ou situations<sup>60</sup>.

Le cas des ajouts temporels est différent. Il est tout à fait possible d'exprimer une relation temporelle entre A et B sans marque lexicale, comme en (13) ou (14) pour lesquels *quand*, *au moment où*<sup>61</sup>, *alors* seraient des connecteurs envisageables.

(13) [L'explosion s'est produite]<sub>A</sub> [il était huit heures]<sub>B</sub>

(14) [Marie est arrivée]<sub>A</sub> [Paul partait]<sub>B</sub>

Si B est un ajout « normal » par rapport à A, il doit réagir à certains tests de la même manière que des ajouts avérés, c'est-à-dire être dans la portée d'une interrogation totale ou d'une négation. C'est le cas quand le connecteur est explicite (ex. 13a et c), mais ce n'est pas le cas dans les juxtapositions (ex. 13b et d).

- (13) a. Est-ce que [l'explosion s'est produite]<sub>A</sub> quand [il était huit heures]<sub>B</sub> ?  
 b. ? Est-ce que [l'explosion s'est produite]<sub>A</sub> [il était huit heures]<sub>B</sub> ?  
 c. [L'explosion ne s'est pas produite]<sub>A</sub> quand [il était huit heures]<sub>B</sub>, mais quand il était...  
 d. ? [L'explosion ne s'est pas produite]<sub>A</sub> [il était huit heures]<sub>B</sub>, mais il était...

57 Pour l'École de Genève, voir Fillietaz & Roulet (2002), pour la RST (*Rhetorical Structure Theory*), voir <https://www.sfu.ca/rst/01intro/definitions.html> et pour la SDRT (*Segmented Discourse Representation Theory*) voir Asher & Lascarides (2003, p. 459-471). Pour une comparaison de plusieurs théories, voir Sanders et al. (1992 et 2021).

58 Pour le français, voir par exemple Roze (2009).

59 Dans le détail, la question semble plus complexe. Certaines approches, bien que ne reconnaissant pas de relation de discours spatiale, attribuent cette valeur à quelques items relevés comme connecteurs, ou plus généralement comme organisateurs textuels (Riegel et al. 2009, p. 1047-1048). Si l'on compte comme une propriété des connecteurs leur capacité à articuler des structures prédicatives, le statut de connecteur me paraît discutable pour ces expressions spatiales.

60 Le cadre de la cognition des événements (*event cognition*, Radvansky & Zacks 2014) est un cadre pertinent pour exprimer cela.

61 On peut tout à fait considérer *au moment où* comme une forme figée à classer parmi les connecteurs. Dans sa typologie, Roze (2009) en fait même une conjonction de subordination. Il n'est pas possible de faire la même chose avec *à l'endroit où*.

Dans les juxtapositions que j'ai étudiées, il n'y a pas d'indices syntaxiques clairs que B constitue un dépendant de A. Le type de dépendance existant joue à un autre niveau, qui n'est pas celui de la *rection* au sens employé par Deulofeu (1989 et 1999 p. 430) quand il analyse ce type de constructions (qui sont chez lui des « couplages » asyndétiques de propositions avec verbe conjugué). B y est dit en *relation d'association* ou *relation énonciative* avec A, par opposition aux relations prédicatives (p. 114). Le couplage, comme relation d'association, peut donner l'impression d'une subordination, sans pour autant en avoir les propriétés syntaxiques. Je pense qu'il serait très intéressant de questionner ce qui est dit des relations d'association dans l'optique de la macro-syntaxe aixoise du point de vue de ce qui est avancé en général pour les relations de discours, même si les objectifs ne sont pas les mêmes. Cela impliquerait d'étiqueter sémantiquement ces relations d'association, et donc d'enrichir sémantiquement une approche qui se veut d'abord syntaxique.

(iii) Enfin, que dire de la catégorie du bloc AB ? On pourrait penser que la construction juxtaposée prise en bloc a le statut de phrase, au même titre qu'une phrase complexe par coordination ou subordination. On a vu que A avait le statut de constituant saturé, dans notre cas une phrase. B n'est pas un argument de A, ni du verbe de A, donc B n'a aucun effet sur la catégorie de A. Et même si l'on admettait que B était un ajout à A, il n'aurait pas non plus d'effet sur la catégorie de A, puisque que l'adjonction à un constituant ne modifie pas la catégorie de ce constituant. La première hypothèse est donc celle du statut phrastique de AB. Mais cette construction ne réagit pas systématiquement bien aux tests permettant habituellement de mettre en évidence qu'un constituant est phrastique. Les contraintes en jeu sont multiples, et c'est un point qui demeure encore confus. Je reprends quelques éléments développés dans Dargnat & Jayez (2010b [14], p. 77-80) montrant cette difficulté.

La manipulation de (13) ci-dessus a déjà montré que la juxtaposition n'était pas un simple effacement de la conjonction. Un autre test consiste à coordonner AB à une phrase, la coordination réclamant des éléments de même statut de part et d'autre du coordonnant (voir ex. 2a, 4a, 13e et 15). Les acceptabilités sont approximatives et varient fortement selon les exemples et selon les personnes, d'où le signe %.

- (2) a. % je perds rapidement patience et [tu avances]<sub>A</sub> [je te casse la gueule]<sub>B</sub>
- (4) a. % J'assurais différentes tâches dans l'établissement : je nettoiais les bureaux ou [je m'occupais des personnes âgées]<sub>A</sub> [on les levait le matin]<sub>B</sub> (corpus Allier)
- (13) e. % J'étais en télétravail ce matin-là car [l'explosion s'est produite]<sub>A</sub> [il était 8 heures]<sub>B</sub> (et les transports en commun étaient coupés)
- (15) J'étais très motivée mais [j'ai essayé d'y aller]<sub>A</sub> [j'ai pas trouvé]<sub>B</sub> (d'après corpus Allier march 11/9)

On peut aller plus loin en notant que la juxtaposition AB (ex. 16) ne semble pas avoir les propriétés d'une phrase complexe par coordination (ex. 16c), qui demande dans certaines configurations (enchâssement dans une complétive ou une interrogative indirecte) une distribution du complémenteur, ici *que* (16a et 16d) et *si* (16b et 16e).

- (16) [il est sorti]<sub>A</sub> [il a vu la voiture]<sub>B</sub>
- a. \*Marie dit qu'[il est sorti]<sub>A</sub> qu'[il a vu la voiture]<sub>B</sub>
- b. \*Marie demande s'[il est sorti]<sub>A</sub> s'[il a vu la voiture]<sub>B</sub>
- c. [il est sorti]<sub>A</sub> et [il a vu la voiture]<sub>B</sub>

- d. Marie dit qu'[il est sorti]<sub>A</sub> et qu'[il a vu la voiture]<sub>B</sub>  
 e. Marie demande s'[il est sorti]<sub>A</sub> et s'[il a vu la voiture]<sub>B</sub>

De plus, il n'est pas garanti que l'interprétation de la juxtaposition en (16) corresponde à celle d'une coordination en *et*<sup>62</sup>.

On peut se demander si AB est plus proche d'une subordination de type « (*subordonnant*) AB » ou « A (*subordonnant*) B » avec une conjonction implicite<sup>63</sup>. Si c'est le cas, AB doit avoir les mêmes propriétés que « *subordonnant* AB » ou « A *subordonnant* B ». Par exemple, on doit pouvoir enchâsser AB dans une structure introduisant une phrase complexe par subordination, par exemple une complétive (ex. 17b et 17c), une interrogative ou un système subordonné (ex. 17d et 17e), etc. A la différence des manipulations sous (16), il n'y a pas de distribution du complémenteur. Prenons l'exemple authentique (17), interprété préférentiellement comme (17a), à savoir où A est un ajout temporel de B, du type *quand* A, B ou *au moment où* A, B.

- (17) [j'ai rentré au théâtre]<sub>A</sub> [ça allait]<sub>B</sub> (Allier mad 4/2)  
 a. Quand [j'ai rentré au théâtre]<sub>A</sub>, [ça allait]<sub>B</sub>  
 b. %Je me souviens que [j'ai rentré au théâtre]<sub>A</sub> [ça allait]<sub>B</sub>  
 c. Je me souviens que quand [j'ai rentré au théâtre]<sub>A</sub> [ça allait]<sub>B</sub>  
 d. %Même si [j'ai rentré au théâtre]<sub>A</sub> [ça allait]<sub>B</sub>, j'avais encore des problèmes  
 e. Même si quand [j'ai rentré au théâtre]<sub>A</sub> [ça allait]<sub>B</sub>, j'avais encore des problèmes

Ces différents tests donnent des indications, mais les observations ne sont pas uniformes à cause de contraintes de natures différentes. Deux facteurs au moins influencent la possibilité de telles manipulations. D'une part, la combinaison sémantique des relations de discours en jeu, d'autre part, la structure informationnelle. Cette dernière peut être corrélée à des phénomènes prosodiques. Tout cela soulève un problème récurrent en linguistique, celui de la maîtrise de la cible des tests. Je n'ai pas forcément de solution sinon celle de la prudence et de la modestie des conclusions.

### 2.1.1.2 Sémantique : interprétation et (non)compositionnalité

Du point de vue sémantique, (i) je me suis intéressée à la signification de la relation entre A et B, (ii) j'ai questionné le caractère plus ou moins compositionnel de l'interprétation du bloc AB, ce qui m'a amenée (iii) à utiliser le cadre et les outils des CxG (voir Dargnat 2008b [8], p. 2469 sqq., Dargnat & Jayez 2010b [14], p. 62-65 et 80-81).

(i) La cohérence repose sur la perception d'une relation entre A et B qui peut faire l'objet de plusieurs interprétations. Même si elles peuvent parfois être ambiguës ou multiples, ces interprétations ne sont pas complètement libres. L'absence de connecteur dans les juxtaposées n'est pas synonyme d'absence de contraintes pour l'interprétation de la relation entre A et B. A priori, le spectre des relations de discours en jeu peut être assez large et ne se réduit pas aux subordonnées circonstancielle exposées dans les grammaires traditionnelles<sup>64</sup>. Les deux théories que j'ai utilisées pour certains points proposent chacune un inventaire des relations distribuées selon plusieurs propriétés,

62 La réflexion sur *and* proposée par Gomez-Txurruga (2003) peut être appliquée au français *et*.

63 La notion de connecteur implicite est utilisée pour l'annotation du *Penn Discourse Treebank* (PDTB) (Prasad et al. 2008). Elle désigne les relations de discours non marquées lexicalement entre deux phrases adjacentes. Elle est reprise dans le *French Discourse Treebank* (FDTB) (Braud 2015, Danlos et al. 2015).



majoritairement sémantiques et pragmatiques. La Rhetorical Structure Theory (RST) insiste sur le dialogisme et l'intentionnalité et la Segmented Discourse Representation Theory (SDRT)<sup>57</sup> sur les indices de calcul de la relation<sup>65</sup>, sa formalisation et les contraintes d'enchaînement. J'ai utilisé la SDRT ponctuellement pour l'intégration de phénomènes prosodiques (Jayez & Dargnat 2008d [11], p. 55-58, et 2012c [18], p. 121-131) et pour la définition de quelques relations (Combettes & Dargnat 2016b [26], p. 4 et 2021c [41], p. 58-59). Le cadre de la RST et l'utilisation que j'en ai faite sont rappelés à la section 2.2.2 (p. 83). Ici, je me concentre sur deux limites que montre l'examen des corpus : il ne paraît pas possible de construire une juxtaposition de type AB ayant une interprétation finale (B étant le but de A) (ex. 18), ni une juxtaposition ayant une interprétation de changement de topique (B introduit un nouveau thème à partir de A) (ex. 19).

- (18) [on leur a donné le Bled]<sub>A</sub> [ils font des progrès]<sub>B</sub><sup>66</sup>  
 ⇒ on leur a donné le Bled, du coup il font des progrès  
 ⇏ on leur a donné le Bled pour qu'il fassent des progrès

- (19) KE383 – CES on appelle  
 JR – ah bon au fait quelle est la différence entre euh ces sigles qui pour moi sont tout nouveaux enfin (ESLO1, ENT 089)

(18) est possible mais, en juxtaposition, l'interprétation préférentielle est consécutive. L'absence de relation finale en juxtaposition est probablement à corrélérer aux contraintes de mode des subordinées finales (subjonctif), mode qu'il est difficile d'avoir sans système conjonctif. En (19), la situation est dialogale, mais il y a bien une relation de discours entre « CES on appelle » et « quelle est la différence entre ces sigles ». Le locuteur JR rebondit sur les sigles employés par le locuteur KE383 et réoriente ainsi le thème. *Au fait* permet de faire le lien et d'assurer la continuité de la cohérence, en évitant ainsi l'effet de rupture, ce qui serait moins évident en cas de simple juxtaposition<sup>67</sup>.

(ii) J'ai abordé la compositionnalité à plusieurs reprises en questionnant différents phénomènes, et en premier lieu les juxtapositions. A un niveau très général, la compositionnalité peut être définie, comme un assemblage réglé d'unités indépendantes. Une expression ou une structure complexes soulèvent le problème de la possibilité de construire l'interprétation du tout à partir de l'interprétation de ses parties. La notion de compositionnalité est une notion historiquement marquée par la logique en référence aux travaux de Frege (voir Janssen 2012 pour une synthèse relativement récente). Elle a largement été exploitée en linguistique pour décrire l'interprétation de syntagmes complexes dans des configurations d'adjacence

64 En général, on y retrouve les « nuances » suivantes : temps, manière, condition, opposition, concession, cause, but, comparaison. Leur présentation et leur regroupement diffèrent d'une grammaire à l'autre. Certaines grammaires du XIX<sup>ème</sup> siècle proposent des listes beaucoup plus fournies (voir Chervel 1979, Colombat & Lahaussais 2019, p. 35). Concernant la notion plus large de relation de discours, voir par exemple Gruber & Redecker (2014), Renkema & Schubert (2018, chap. 6) ou Sanders et al. (1992 et 2021).

65 La SDRT est une approche *bottom-up*, c'est-à-dire que la signification de la relation est le résultat d'un calcul des contraintes extraites des deux segments mis en relation et des contraintes générales d'enchaînement.

66 D'après *hein je vois on leur a donné le Bled on leur apprend quand même beaucoup plus l'orthographe on leur apprend à former des lettres comme on faisait dans le temps* (Fra80 hfa17).

67 Sur ce type de relation dite « de digression » et les marqueurs associés, voir Charolles (2020).

(typiquement un nom suivi d'un adjectif). L'utiliser oblige à se poser des questions qui interrogent les limites de la compositionnalité dans les différents domaines empiriques et théoriques où elle a été invoquée (pour un développement, voir Dargnat 2015a [23]) :

- Comment construire la hiérarchie entre parties et tout ?
- Quel genre d'objet complexe vise-t-on : syntagmes en général, catégories particulières de syntagmes (par exemple syntagmes verbaux, phrases), discours et unités de discours, patrons prosodiques, etc. ?
- Quel est le mode dépendance entre les propriétés des composants et les propriétés de l'objet complexe ?

La première question correspond à la problématique du découpage des messages linguistiques. Une des réponses traditionnelles est celle des distributions, c'est-à-dire des ensembles structurés de substitutions possibles. L'idée centrale est que deux objets simples sont « comparables » (rapportables à une même catégorie) quand la substitution de l'un par l'autre laisse invariante une certaine propriété de l'objet complexe auquel ils appartiennent. Appliqué aux juxtapositions, cela veut dire que, si l'interprétation de AB est compositionnelle, on peut remplacer A ou B par des propositions dont les interprétations sont respectivement équivalentes à A et B sans que l'interprétation globale ne soit affectée<sup>68</sup>.

La deuxième question soulève le problème de la nature des tous. Si l'on s'éloigne des syntagmes, on trouve des assemblages plus difficiles à maîtriser et à catégoriser, car ils renvoient à des propriétés illocutoires ou informationnelles et mobilisent des notions moins standard comme le couplage, le détachement, la dislocation, etc. Les exemples (13), (16) et (17) ont montré qu'il n'est pas évident d'attribuer le statut de phrase complexe aux juxtapositions. De plus, à ce niveau-là, à l'oral, l'intervention des facteurs prosodiques est fréquente et contribue à l'effet « colle » entre A et B (voir section 2.1.3, p. 69). Le signe est en quelque sorte enrichi, car aux traits syntaxiques et sémantiques habituellement invoqués pour décrire la combinatoire au niveau syntagmatique sont intégrés des traits discursifs ou contextuels jugés pertinents et distinctifs.

La réponse à la troisième question est plus technique. Elle repose sur l'évolution de la définition de la compositionnalité, sur l'extension des objets auxquels il est possible de l'appliquer et sur le renouvellement des solutions pour l'exprimer. Ici, je retiens principalement que l'assouplissement de la notion a permis d'aborder des phénomènes discursifs et de questionner les contraintes d'enchaînement qui ne sont plus vues comme des règles rigides mais comme des préférences résultant de l'application de règles par défaut<sup>69</sup>. Pour les juxtapositions, cela revient à se demander si l'interprétation de AB est dérivable de l'interprétation de A et de B pris isolément et d'un ensemble de contraintes indépendantes de la juxtaposition elle-même.

68 Cette idée a été exploitée lors d'une expérience en TAL sur les enchâssements (voir section 2.2.2).

69 Cette idée d'application d'une règle par défaut est développée dans les approches dites non monotones, telles que la SDRT. La non-monotonie renvoie au fait que l'ajout d'une information peut modifier le résultat, c'est-à-dire l'interprétation finale. Dans les structures de traits proposées ici dans les Figures 5 et 8 (ici p. 60 et 65), l'application par défaut est représentée par le signe /.

Les juxtapositions que j'ai étudiées n'ont pas les mêmes propriétés. Certaines sont compositionnelles, comme en (20) :

(20) [Paul a poussé John]<sub>A</sub> [John est tombé]<sub>B</sub>

La relation entre les événements A et B est communément interprétée comme une relation causale ou consécutive (A est la cause de B et B est la conséquence de A). A et B ont la même interprétation pris isolément et en juxtaposition et les contraintes permettant d'interpréter leur connexion sont des contraintes générales communément admises, qui sont notamment que pousser produit une chute et pas l'inverse (ce qui permettrait de résoudre l'anaphore si la deuxième occurrence de John était remplacé par *il*). Les temps verbaux de l'exemple ne remettent pas en cause cette contrainte. D'autres contextes, d'autres jeux temporels rendraient possibles d'autres interprétations, mais qui seraient moins « pertinentes », au sens où elles auraient un coût cognitif plus important<sup>70</sup>. Cette « propension à l'interprétation causale » est fréquemment notée, de part le rôle de la causalité dans la rationalité et dans la catégorisation humaines<sup>71</sup>.

D'autres juxtapositions fonctionnent différemment, comme (21) :

(21) [Paul prend le tram]<sub>A</sub> [il sera à l'heure]<sub>B</sub>

Intéressons-nous au cas où l'interprétation globale de AB est conditionnelle et pas simplement consécutive, c'est-à-dire au cas où B vaut sous la condition A. Il s'agit d'une *prédictive* dans la typologie de Dancygier (1998)<sup>72</sup>. La réalisation de A conditionne/prédit la réalisation de B. Quand A est pris isolément, il est avant tout assertif et l'indicatif actualise le procès. Il n'est pas évident que l'on puisse dériver une interprétation conditionnelle de la valeur isolée de A. C'est seulement dans la configuration AB, telle que présentée en (21), que cette valeur apparaît. Dans ce cas-là, l'interprétation de AB comme conditionnelle n'est pas complètement compositionnelle, on la dira *constructionnelle*.

(iii) J'utilise les termes *construction* et *constructionnalité* en référence au cadre des Grammaires dites de construction. Globalement, les CxG sont présentées comme des grammaires qui formulent ce qu'un sujet doit savoir pour former et interpréter avec succès des énoncés. Leur objectif est de répertorier tous les patrons abstraits (*constructions*) possibles qui permettent la production et l'interprétation d'énoncés bien formés pour une langue donnée (*constructs*), mais sans faire d'hypothèse quant au caractère simple ou complexe, central ou périphérique, ni compositionnel de ces patrons. Le terme *construction* est utilisé de manière générale et depuis longtemps dans les grammaires traditionnelles et en linguistique, en particulier en syntaxe, mais il est présenté comme un concept central avec des propriétés précises dans le cadre des CxG. Sont considérées comme des constructions toutes les associations conventionnelles d'une forme et d'un sens. Les constructions syntaxiques sont la porte d'entrée privilégiée, mais, en théorie, la notion de construction est applicable au morphème, au lexème, au syntagme, à la phrase, voire plus. En effet, le champ de ces grammaires est

70 Par exemple *Paul a poussé John, John était tombé*, ou *Paul a poussé John, il (Paul) est tombé*.

71 Pour une approche linguistique, voir Nazarenko (2000, p. 45-48 et 148-150), Gross & Nazarenko (2004, p. 37-40) et pour une perspective cognitive, voir Radvansky & Zacks (2014).

72 « [Predictive conditionals] represent predictive reasonings, and they are therefore marked with *if*-backshift in the protases and have a predictive modal in their apodoses. » (Dancygier 1998, p. 48).

multiple, ce pourquoi on trouve de plus en plus le terme au pluriel et les phénomènes décrits dépassent aujourd'hui largement celui des constructions verbales argumentales « à la Goldberg »<sup>73</sup>. S'il existe des sensibilités différentes<sup>74</sup>, émergent tout de même des questionnements communs qui correspondent aux miens et des outils de représentation qui m'ont été utiles pour décrire et expliquer plusieurs des phénomènes que j'ai étudiés, dont les juxtapositions. A ce stade, je retiens : la multidimensionnalité des descriptions et leur formalisation en structures de traits ; le critère non-compositionnel (ou quantitatif) dans la définition d'une construction ; la mise en évidence d'héritages de propriétés et la réticulation des constructions, en synchronie et en diachronie.

Les constructions sont présentées comme des unités abstraites de la grammaire à l'interface entre plusieurs dimensions de la description linguistique. Dans une construction, la forme est généralement exprimée en termes de propriétés syntaxiques, morphologiques et phonétiques et le sens renvoie à une articulation de propriétés sémantico-lexicales, pragmatiques et discursives. Toutes les approches ne le font pas, mais certaines représentent les constructions au moyen de schémas, des matrices attribut-valeur (AVM)<sup>75</sup>, qui rendent compte de cette multidimensionnalité en imbriquant des boîtes donnant diverses informations. Les représentations diffèrent, ont évolué ou se sont enrichies. Elles n'ont pas la même « allure » chez tout le monde. J'ai adapté ce formalisme pour rendre compte des constructions juxtaposées en général et des juxtapositions conditionnelles en particulier en précisant leurs contraintes spécifiques.

La question de la compositionnalité ou non des constructions complexes n'est pas non plus traitée chez tous de la même manière et sa prise en compte a évolué. Chez Goldberg, qui reprend les débats antérieurs et contemporains à sa réflexion, les constructions sont par définition des structures non compositionnelles. Une structure abstraite n'a le statut de construction que quand au moins une de ses propriétés ne peut pas être prédite(s) à partir de la connaissance des autres constructions existant dans la grammaire de cette langue.

C is a construction iff<sub>def</sub> is a form-meaning pair  $\langle F_i, S_i \rangle$  such that some aspect of  $F_i$  or some aspect of  $S_i$  is not strictly predictable from C's component parts or from other previously established constructions. (Goldberg 1995, p. 4)

C'est cette contrainte définitoire qui m'a particulièrement intéressée dans mes premières utilisations des CxG sur les juxtapositions.

### **2.1.1.3 Contraintes sur l'enchaînement discursif**

Je propose ci-dessous une représentation générale faite à l'occasion de deux séminaires<sup>76</sup>, publiée en partie dans Darnat (2008b [9], p. 2476 et reprise ici dans

73 Construction du type *She baked him a cake* (Goldberg 1995, p. 9), que l'on peut traduire par « Elle lui a fait cuire un gâteau (dans l'intention qu'il l'ait) ».

74 Parmi les très nombreux travaux, voir par exemple Desagulier (2015), Fried & Östman (2004), Hilpert (2014), Hoffman & Trousdale (2013), Kay (1995, 2002), Legallois (2016), Legallois & Gréa (2003), Mathieu (2003), Sag (2012), Yoon & Gries (2016).

75 C'est aux travaux de Fillmore (1988) que l'on renvoie en général pour l'utilisation de matrices attribut-valeur (AVM) dans le cadre des CxG. Les constructions décrites sont notamment sujet-prédicat, les figements (*phraseological unit*) comme *do one's best* ou les corrélatives conditionnelles sur le modèle *the more..., the more...*

76 « Parataxes du français et Grammaires de Construction », Séminaire du Laboratoire Parole & Langage, 28 mars 2008, Université de Provence, Aix-en-Provence ; « Parataxes : contraintes

l'Annexe 1a, p. 193). Elle me sert ici de base pour exposer certaines contraintes de l'enchaînement discursif et pour cibler plus loin les particularités des juxtapositions à interprétation conditionnelle.

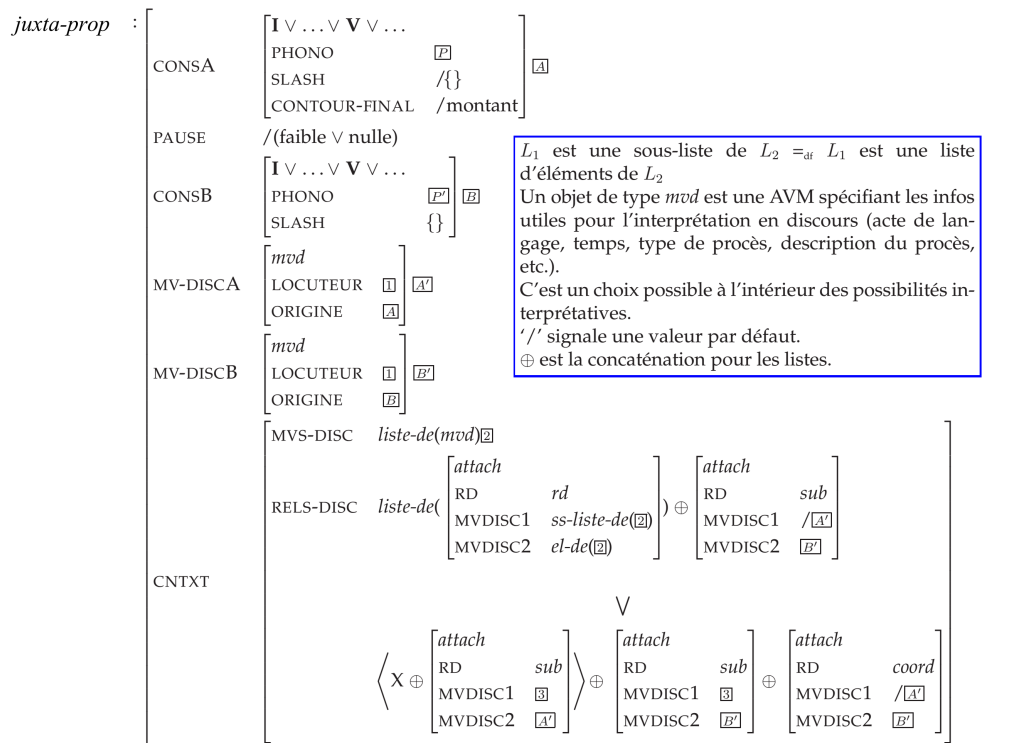


Figure 5. Structure de traits de la construction *juxta-prop*  
 Les attributs sont en petites capitales et les types en italiques.

Les valeurs des attributs CONSA, CONSB et PAUSE renvoient aux propriétés présentées ou mentionnées ci-dessus, à savoir :

- que chaque constituant est de forme phrastique avec un verbe conjugué (type I, II, ... ou V). A ce niveau-là, ces types peuvent correspondre à des déclaratives, des interrogatives, des impératives, des optatives, etc.
- que les deux constituants sont (par défaut pour A) saturés syntaxiquement, c'est-à-dire qu'ils sont présentés comme n'ayant aucune « place vide », aucun *gap* – c'est ce que signifie la valeur vide pour l'attribut SLASH<sup>77</sup>,
- que la pause entre les deux constituants est par défaut faible ou nulle,
- que le constituant A présente par défaut un contour prosodique final montant.

Je développe plus spécifiquement les attributs en lien avec les mouvements discursifs (i) MV-DISC A, MV-DISC B, et (ii) CNTXT.

d'interprétation et de représentation », séminaire SMS (Syntaxe, Morphologie, Sémantique), 29 janvier 2010, ATILE, Nancy.

77 La notation SLASH est empruntée à la modélisation de type HPSG. L'attribut GAP est aussi en usage (voir Gazdar 1981, Müller et al. 2021, Pollard & Sag 1994.). Pour la comparaison avec les CxG, voir Mathieu (2003) et Müller (2021 dans Müller et al. 2021, chap. 32). Pour les valeurs par défaut, voir Lascarides & Copestake (1999).

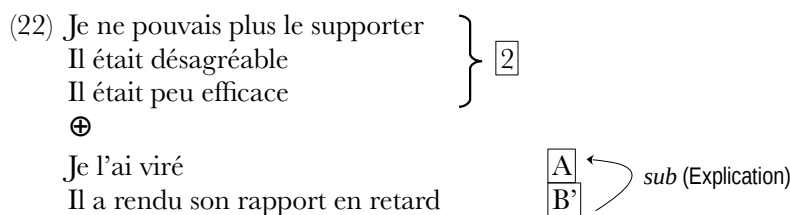
(i) Les deux constituants ne sont pas présentés comme simplement additionnés. Ils sont perçus du point de vue de leur propriétés discursives, ce pourquoi ils apparaissent sous les attributs MV-DISCA et MV-DISCB avec comme valeur un type mouvement discursif (*mv*). Pour simplifier la lecture, les constituants enrichis de leurs propriétés discursives sont renommés A' et B'. On y trouve des propriétés comme le locuteur et des propriétés héritées des constituants d'origine, respectivement CONSA et CONSB. La Figure 5 cible des énoncés monologiques où les deux constituants sont le fait d'un même locuteur (LOCUTEUR [1]), mais des enchaînements dialogaux sont théoriquement possibles.

(ii) L'attribut CNTXT concerne les propriétés et les contraintes de l'enchaînement discursif juxtaposé A'B' et de son insertion dans le flux du discours<sup>78</sup>. On y retrouve notamment les propriétés associées aux mouvements discursifs A' et B', des informations sur la relation de discours en jeu entre eux et sur les contraintes d'attachement associées. Dans cette représentation générale, deux scénarios d'attachement sont présentés comme possibles. On peut les paraphraser de la manière suivante.

– Scénario 1. Si on lit la Figure 6 ci-dessous de gauche à droite, on a une liste de paires de mouvements discursifs attachés (*attach*) par des relations de discours. On ajoute ( $\oplus$ ) à cette liste la paire correspondant à la juxtaposition de type A'B', avec une relation de discours de subordination discursive (*sub*) entre A' et B'. On peut illustrer ce phénomène par l'exemple (22), où la relation entre A' et B' est une relation **Explication**, relation considérée ici comme discursivement subordonnante.

$$\text{liste-de} \left( \begin{array}{cc} \text{attach} & \\ \text{RD} & rd \\ \text{MVDISC1} & \text{ss-liste-de}(\boxed{2}) \\ \text{MVDISC2} & \text{el-de}(\boxed{2}) \end{array} \right) \oplus \left[ \begin{array}{cc} \text{attach} & \\ \text{RD} & \text{sub} \\ \text{MVDISC1} & /[\text{A}'] \\ \text{MVDISC2} & [\text{B}'] \end{array} \right]$$

Figure 6. Contrainte d'attachement de type subordination discursive



– Scénario 2. Dans la Figure 7 ci-dessous, les choses se comprennent mieux de droite à gauche. Si l'on cherche à attacher B' à A' par une relation de coordination discursive, cela implique dans la structure que B' soit lui-même attaché par une relation de subordination au même mouvement discursif ([3]) que A'. Pour simplifier, [3] est lui-même présenté ici comme le dernier mouvement discursif d'une liste potentielle (X) de paires de mouvements discursifs. On peut illustrer ce phénomène par l'exemple (23), où

<sup>78</sup> Le flux discursif est envisagé comme une liste de mouvements discursifs. Dans la Figure 5, cela correspond à l'attribut MVS-DISC avec une valeur de type *liste-de(mv-disc)*.

A' et B' sont reliés par la relation *Énumération*, considérée comme discursivement coordonnante.

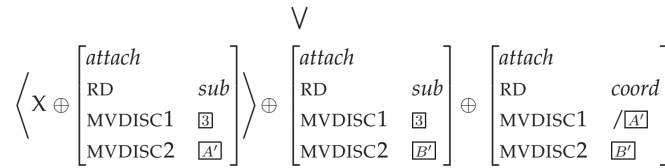
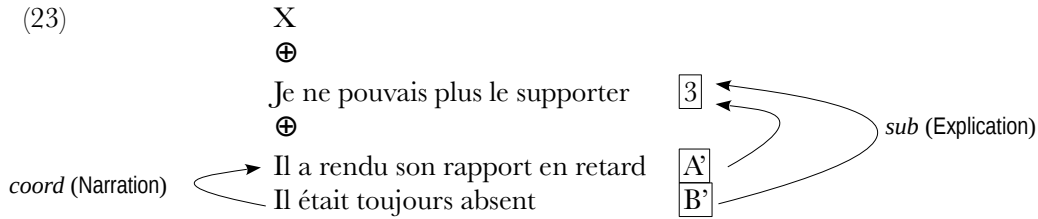
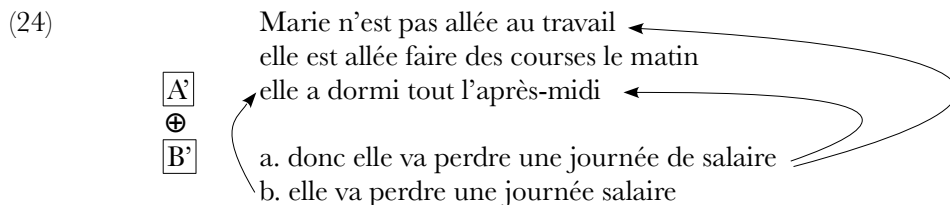


Figure 7. Contrainte d'attachement de type coordination discursive



La distinction coordination et subordination discursives correspond globalement à deux règles d'insertion dans le flux du discours, c'est-à-dire dans une liste de mouvements discursifs. On la doit initialement à Polanyi (1985, p. 308 et Polanyi et al. 2003, p. 3)<sup>79</sup>. Pour simplifier, il s'agit d'une analyse de la cohérence discursive basée sur les contraintes d'enchaînement linéaire de gauche à droite de constituants simples ou complexes. En résumé : (i) dans cette approche et celles qui suivent, les relations de discours sont soit subordonnantes (par ex. *Élaboration*, *Interruption*, *Digression*, etc.) soit coordonnantes (par ex. *Narration*, *Liste*, *Contraste*, etc.), mais avec un certain flottement dans le classement selon les critères retenus<sup>80</sup>; (ii) il existe une représentation arborescente des attachements sur la base de laquelle on peut visualiser les sites d'attachement possibles, lorsqu'on souhaite ajouter un constituant à un ensemble déjà formé et cohérent de plusieurs constituants. C'est à ce niveau-là que l'on parle de *contrainte de la frontière droite* pour désigner le fait que seuls les sites sur la branche droite de l'arbre sont accessibles pour l'attachement d'un nouveau constituant.

Sur la base de quelques exemples, j'ai remarqué qu'il semblait y avoir pour les juxtapositions étudiées une contrainte plus spécifique que celle de la frontière droite : le seul site accessible pour B' est A'. Autrement dit, on a un attachement purement local. Cela renforce l'idée d'une intégration forte de A' et B' dans un même bloc, sans pour autant que ce bloc acquière toutes les propriétés d'une phrase complexe. Un exemple comme (24) illustre ce phénomène.



79 Elle été largement exploitée ensuite par la SDRT (Asher 1993, p. 270 sq., Asher & Vieu 2005).  
 80 La relation consécutive n'existe pas telle quelle dans toutes les théories, il faut parfois la chercher dans la causalité ou le résultat (voir Nazarenko 2000, p. 15-26, Gross & Nazarenko 2004, p. 17). Présentée comme discursivement coordonnante, notamment pour des raisons temporelles, elle peut aussi dans certains contextes être subordonnante (voir Asher & Vieu 2005, p. 609, Delort 2004).

Dans le cas où B' est relié à ce qui précède par une relation de conséquence marquée par *donc* (cas a), l'attachement peut se faire assez naturellement avec le tout premier constituant (*Marie n'est pas allée au travail*), ou avec A' (*elle a dormi tout l'après-midi*). Dans le cas b, où l'attachement n'est pas marqué lexicalement, et où les propriétés de la construction *juxta-prop* sont réalisées (notamment le contour prosodique et l'absence de pause), seul A' semble accessible pour l'attachement de B'. L'idée d'une restriction locale de la contrainte de la frontière droite dans les constructions juxtapositions n'est qu'une piste de réflexion. Elle mériterait une analyse plus systématique. J'y reviens rapidement ci-dessous à propos des conditionnelles.

### 2.1.2 Juxtapositions conditionnelles : contraintes et constructionnalité

Mes premières recherches sur ce phénomène (Darnat 2008b [9]) ont servi de base à une présentation récente plus synthétique dans le cadre de la *Grande Grammaire du Français* (Darnat 2021c [39]). Le paradigme des juxtapositions conditionnelles est assez diversifié. Je me suis focalisée sur les constructions du type de *juxta-prop*<sup>81</sup>, comme dans les exemples suivants, où A et B présentent toutes les propriétés de CONSA et CONSB de la Figure 5 présentée précédemment.

- (25) [J'ai un coup de cafard]<sub>A</sub> [je mets une cassette vidéo jusque les trois quatre heures du matin]<sub>B</sub> [ST I3 33'54]
- (26) [Il prendrait le tram]<sub>A</sub> [il serait à l'heure]<sub>B</sub>
- (27) [Tu écris tes mémoires]<sub>A</sub> [tu leur donnes quel titre ?]<sub>B</sub> (Canal +, 5/1/08, 13h20)
- (28) [Tu as faim]<sub>A</sub> [il y a des biscuits dans le placard]<sub>B</sub> (d'après Austin 1970 [1961])
- (29) [Tombe en panne]<sub>A</sub> [on est morts]<sub>B</sub>

De tels énoncés soulèvent quelques questions dont :

- (i) ce que l'on entend précisément par « interprétation conditionnelle »,
- (ii) le recensement des schémas constructionnels juxtaposés susceptibles de recevoir une telle interprétation,
- (iii) leur correspondance ou non avec les conditionnelles en *si*, marqueur prototypique de la relation,
- (iv) les propriétés héritées du schéma constructionnel général (*juxta-prop*) et les propriétés spécifiques aux juxtapositions conditionnelles,
- (v) le caractère plus ou moins compositionnel de tels agencements, en particulier du point de vue des valeurs modales en A.

#### 2.1.2.1 La relation de condition

(i), (ii) et (iii) soulèvent la question de l'interprétation conditionnelle. Globalement, cela renvoie au fait que l'on associe (conditionne) la réalisation d'une situation (B) à la réalisation d'une situation (A). Une théorie comme la RST inscrit cette dépendance dans la définition de la relation **Condition** et précise que la proposition conditionnante (A dans les exemples) doit être non-réalisée (hypothétique, future ou autre)<sup>82</sup>. L'intention

81 Certaines études prennent en compte également les conditionnelles marquées par la conjonction *et*, mais je les ai laissées de côté (voir Corminboeuf 2008, 2009).

82 Je renvoie également à la typologie proposée par Corblin (1999 et 2002), à la réflexion de Vogüé (1999) sur les deux sens de *condition*, aux remarques de Gross & Nazarenko (2004, p. 21-22) et à



du locuteur est que son interlocuteur reconnaisse « en quoi la réalisation de B dépend de celle de A »<sup>83</sup>, ce qui explique pourquoi cette relation de discours acquiert facilement un relief argumentatif. Quelques remarques générales ont émergé dans mes analyses.

– *Remarque 1* : la portée. Les exemples du type de (28), fréquemment appelés phrases austiniennes, ne fonctionnent pas au niveau du contenu des propositions mais des actes de langage (⇒ « dans le cas où/si tu as faim, je te signale qu'il y a des biscuits dans le placard »). Ce n'est pas la faim de l'interlocuteur qui conditionne la présence de biscuits. Ces configurations fonctionnent aussi en juxtaposition. En revanche, un autre type de conditionnelle proche, parfois appelée métatextuelle (voir ex. 30), n'est pas récupérable en juxtaposition. Comme dans (28), la phrase en *si* ne conditionne pas le contenu de l'autre proposition mais son énonciation. On remarquera que ces formules (*si je peux me permettre, si l'on peut dire, si tu veux*, etc.) ont tendance au figement et ne sont pas loin de constituer des adverbes modalisateurs voire des marqueurs discursifs.

(30) Ce livre est, si j'ose dire, un livre de piété. Je l'ai écrit en souvenir de mon enfance et des heures de lecture inoubliables [...] (Caradec, *Christophe Colomb*, 1956, p. 13)

– *Remarque 2* : les conditionnelles négatives. Le plus souvent, elles sont marquées par le connecteur *sinon* comme dans (31). *Ou* et *ou sinon* sont aussi fréquents et montrent que l'interprétation repose sur une alternative ou un contraste implicite (si A alors non B / si non A alors B). Cette interprétation n'est pas possible en juxtaposition.

(31) Dès octobre, [il fallait entretenir le feu, le laisser couvrir toute la nuit]<sub>A</sub>, sinon [la maison en profitait pour s'ensauvager]<sub>B</sub> : chaque pierre des murs redevenait un morceau de roche glaciale et sans pitié. (Frantext, C. Barreau, *La confiture de morts*, 2020, p. 130)

– *Remarque 3* : la condition inverse. Elle est prototypiquement marquée par *à moins que, sauf si, excepté si* (voir ex. 32). Ce sont les cas où B ne se réalise que quand A ne se réalise pas. Cette interprétation n'est pas non plus récupérable en juxtaposition.

(32) Sauf si [Marie vient]<sub>A</sub>, [Paul sera triste]<sub>B</sub>.

– *Remarque 4* : les constructions en *si* non conditionnelles. Elles reçoivent des interprétations explicative (33), concessive (34) et dialectique (35) (de Vogüé 1999, p. 2, pour les catégories et les exemples associés).

(33) S'[il est riche]<sub>A</sub>, c'est parce qu'[il a gagné au loto]<sub>B</sub>.

(34) S'[il est riche]<sub>A</sub>, [il n'est pas milliardaire]<sub>B</sub>.

(35) S'[il est riche]<sub>A</sub>, [cela ne se voit pas]<sub>B</sub>.

Elles sont moins naturellement récupérables en juxtaposition avec A non réalisé, mais peut-être pas impossibles, en particulier, à l'oral, avec une prosodie (contour continuatif et absence de pause) et des mimiques marquées (par ex. haussement de sourcils, moue).

### 2.1.2.2 Héritage et spécification

Les questions (iv) et (v) ci-dessus (p. 63) peuvent être développées sur la base de la spécification de certaines valeurs de la structure générale *juxta-prop*. Par exemple, au niveau du typage de la construction et de ses composants et au niveau de l'enchaînement discursif. Quelles sont les propriétés notables des juxtapositions

l'analyse de Dancygier (1998), dont les études sur l'anglais sont transposables au français.

83 <https://www.sfu.ca/rst/07french/definitions.html>, voir « condition ». Une définition complète est donnée plus loin (section 2.2.2, exemple (50)).

conditionnelles ? On pourrait construire une structure de traits spécifique aux juxtapositions conditionnelles (nommée *juxta-cond*) sur la base de la Figure 5, en spécifiant certains éléments. C’est ce que font ci-dessous les Figures 8 et 9<sup>84</sup>. La structure complète est reprise ici dans l’Annexe 1b (p. 193).

L’intérêt de la perspective constructionnelle est de représenter les ressources d’une langue non seulement sous la forme de constructions mais aussi sous la forme d’un réseau hiérarchisé de constructions<sup>85</sup>. Dans la Figure 8, les propriétés des deux constituants en jeu sont héritées de la construction juxtaposition propositionnelle en général, mais certaines valeurs sont plus spécifiques du point de vue des types de constituant.

<i>juxta-cond</i> :	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 5px;">CONSA</td> <td style="padding: 5px;"> <math>\left[ \begin{array}{l} \text{decl-hd-su-cl} \vee \text{imp-hd-cl} \vee \text{inv-hd-su-cl} \vee \text{opt-hd-subj-cl} \\ \text{SLASH} \\ \text{CONTOUR-FINAL} \end{array} \right]</math> </td> <td style="padding: 5px; vertical-align: middle;">/{} <span style="border: 1px solid black; padding: 2px;">A</span></td> </tr> <tr> <td style="padding: 5px;">PAUSE</td> <td style="padding: 5px;">/(faible <math>\vee</math> nulle)</td> <td></td> </tr> <tr> <td style="padding: 5px;">CONSB</td> <td style="padding: 5px;"> <math>\left[ \begin{array}{l} \text{decl-hd-su-cl} \vee \text{int-hd-su-cl} \\ \text{SLASH} \end{array} \right]</math> </td> <td style="padding: 5px; vertical-align: middle;">/{} <span style="border: 1px solid black; padding: 2px;">B</span></td> </tr> <tr> <td style="padding: 5px;">[...]</td> <td></td> <td></td> </tr> </table>	CONSA	$\left[ \begin{array}{l} \text{decl-hd-su-cl} \vee \text{imp-hd-cl} \vee \text{inv-hd-su-cl} \vee \text{opt-hd-subj-cl} \\ \text{SLASH} \\ \text{CONTOUR-FINAL} \end{array} \right]$	/{} <span style="border: 1px solid black; padding: 2px;">A</span>	PAUSE	/(faible $\vee$ nulle)		CONSB	$\left[ \begin{array}{l} \text{decl-hd-su-cl} \vee \text{int-hd-su-cl} \\ \text{SLASH} \end{array} \right]$	/{} <span style="border: 1px solid black; padding: 2px;">B</span>	[...]		
CONSA	$\left[ \begin{array}{l} \text{decl-hd-su-cl} \vee \text{imp-hd-cl} \vee \text{inv-hd-su-cl} \vee \text{opt-hd-subj-cl} \\ \text{SLASH} \\ \text{CONTOUR-FINAL} \end{array} \right]$	/{} <span style="border: 1px solid black; padding: 2px;">A</span>											
PAUSE	/(faible $\vee$ nulle)												
CONSB	$\left[ \begin{array}{l} \text{decl-hd-su-cl} \vee \text{int-hd-su-cl} \\ \text{SLASH} \end{array} \right]$	/{} <span style="border: 1px solid black; padding: 2px;">B</span>											
[...]													

Figure 8. Spécification des attributs CONSA, CONSB et PAUSE de la construction *juxta-cond*

En A, on peut avoir une proposition déclarative (*decl-hd-su-cl*) (ex. (25) plus haut ou encore (36)), impérative (*imp-hd-cl*) (ex. (29) plus haut ou encore (37)), optative (*opt-hd-subj-cl*) (ex. (38)) ou une proposition avec inversion du sujet (*inv-hd-su-cl*) (ex. 39)<sup>86</sup>.

- (36) Tu continues, je te vires pour la paix et la tranquillité de la communauté. C’est mon rôle de la protéger des trublions. (web, jeuxvideo.com, 20-07-2012, dans Dargnat 2021a [39], ex. 65c)
- (37) Donne, il te sera donné : retour sur les rallyes JA15. (web, auvergne.adventiste.org, 21-05-2015, dans Dargnat 2021a [39], ex. 66c)
- (38) Qu’il vienne et je lui montrerai (dans Jayez & Dargnat 2009 [12], ex. 2c)
- (39) M. Astley nous rencontre souvent à la promenade, il se découvre et passe, bien qu’il meure, cela va sans dire, du désir de se joindre à nous. L’invite-t-on, il refuse aussitôt. (Dostoïevski, *Le joueur*, dans Corminboeuf 2017, p. 11, ex. 28b)

Il aurait fallu ajouter la possibilité d’avoir une proposition interrogative en A, comme l’illustre l’exemple (40) :

84 La structure de traits complète est proposée dans Dargnat (2008b [9], p. 2476). Des éléments plus spécifiques aux constructions juxtaposées conditionnelles avec A à l’impératif sont repris de Jayez & Dargnat (2009 [12], p. 258).

85 Certains travaux utilisent les termes de *macro-*, *méso-* et *micro-constructions* pour désigner cette hiérarchisation (Traugott 2007, 2008, Trousdale 2012, 2015, Hoffman 2013). Dans mes travaux sur les juxtapositions, je n’ai pas utilisé cette terminologie. Elle est apparue lors de réflexions diachroniques plus récentes sur le passage de la comparaison à la concession (Combettes & Dargnat 2016b [26], p. 3, et section 3.2.2.1 ci-dessous). A posteriori, il me semble que cela serait tout à fait applicable aux différentes constructions juxtaposées. Par exemple « juxtaposition propositionnelle » serait une macro-construction, « juxtaposition propositionnelle conditionnelle » une méso-construction et « juxtaposition conditionnelle à l’impératif en A » (voir pseudo-impératif) une micro-construction.

86 *decl-hd-su-cl* : proposition (clause) de type déclaratif avec une tête verbale et un sujet ; *imp-hd-cl* : proposition avec une tête verbale à l’impératif ; *opt-hd-subj-cl* : proposition avec une tête verbale au subjonctif exprimant le souhait ; *inv-hd-su-cl* : proposition avec tête verbale et un sujet inversé (ex. *viendrait-il, je serais heureuse*). Je n’ai pas traité cette dernière construction.

(40) Votre voisin se lève tard ? Dénoncez-le sur delation-gouv.fr (presse écrite, dans Corminboeuf 2017, ex. 28a p. 11)

En B, ne sont a priori présentés comme possibles que les cas où la proposition est déclarative (*decl-hd-su-cl*) (ex. (25), (26), (36), etc.) ou interrogative (*int-hd-su-cl*) (ex. (27)).

### 2.1.2.3 Enchaînement discursif

La relation *Condition* relève a priori du mode d'insertion de type subordination discursive<sup>87</sup>. On est donc dans une configuration du type de celle représentée dans la Figure 6 plus haut. Là aussi, l'analyse des données oblige à spécifier une contrainte supplémentaire.

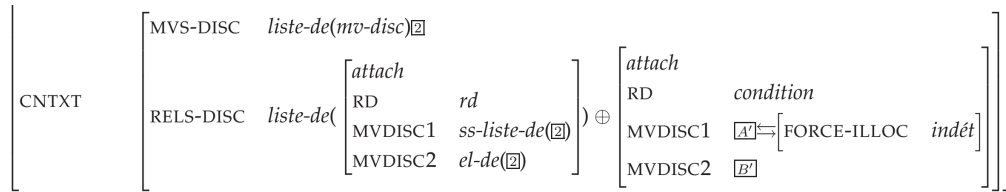


Figure 9. Spécification de l'attribut CNTXT de la construction *juxta-cond*

Tout se passe dans la « boîte » la plus à droite, où l'attachement de B' à A' est de type *condition*, et où la valeur de la force illocutoire de A' peut être modifiée par la construction ( $A' \Leftrightarrow \text{FORCE-ILLOC}$ ). J'y reviens ci-dessous, en prenant l'exemple du conditionnel et de l'impératif.

### 2.1.2.4 Valeurs modales et non-compositionnalité

Dans certains des exemples commentés, et dans la Figure 9, apparaît un processus de « relecture » de la valeur modale de la forme verbale du constituant A. Tout se passe comme si certaines des propriétés du constituant A pris isolément étaient modifiées dès lors qu'il est compris comme un mouvement discursif (A') pris dans le schéma constructionnel décrit. J'illustre cela par deux exemples : (i) le conditionnel et (ii) l'impératif.

(i) La comparaison des juxtapositions conditionnelles et de leur correspondantes en *si* montre que l'usage du conditionnel en A n'est pas uniforme. Un exemple comme (26) ( $[\text{il prendrait le tram}]_A [\text{il serait à l'heure}]_B$ ) est tout à fait naturel en juxtaposition. La présence du conditionnel en A contraint la présence du conditionnel en B. L'équivalent avec *si* (*si il prendrait le tram, il serait à l'heure*) n'est pas bon, ou alors si l'on admet une variation de type sociale, acquisitionnelle ou géographique. Il s'agit donc d'une contrainte spécifique à la construction *juxta-cond* qui devrait être complétée par une règle formulable ainsi :

$$\boxed{\text{A}} \left[ \text{HEAD} \quad \left[ \text{MODE} \quad \text{cond} \right] \right] \Rightarrow \boxed{\text{B}} \left[ \begin{array}{l} \text{decl-hd-su-cl} \\ \text{HEAD} \quad \left[ \text{MODE} \quad \text{cond} \right] \end{array} \right]$$

Figure 10. Exemple de contrainte de la construction *juxta-cond* (d'après Darnat 2008b [9], p. 2477)

<sup>87</sup> La relation conditionnelle n'est pas présentée partout de la même manière. Elle n'est pas directement définie chez Polanyi (1985), elle relève de la conséquence chez Asher et Lascarides (2003), elle est explicitement ajoutée chez Roze (1999, 2013) comme relation subordonnante.

A y regarder d'un peu plus près, c'est le fonctionnement modal habituel du conditionnel qui est modifié par la construction. Prenons un énoncé isolé comme (41a), où le conditionnel demande une accommodation avec une autre situation, ce que propose (41b). En (41c), le même énoncé pris dans une construction juxtaposée à interprétation conditionnelle ne présente pas les mêmes propriétés modales. Le conditionnel du constituant A (*rentrait*) ne réclame aucune accommodation avec une autre situation, c'est lui qui sert de situation accommodante pour le conditionnel du constituant B (*mangerait*). En (41d), le statut du conditionnel du segment B (*rentrerait*) semble être celui de (41b), c'est-à-dire qu'il est modalement subordonné à la proposition en *si* qui la précède (*s'il n'y avait pas d'embouteillage*). Son statut de situation accommodante pour le conditionnel de B (*mangerait*) n'est pas clair.

- (41) a. Il rentrerait plus tôt  
 b. S'il n'y avait pas d'embouteillage, il rentrerait plus tôt  
 c. [Il rentrerait plus tôt]<sub>A</sub> [il mangerait avec ses enfants]<sub>B</sub>  
 d. S'il n'y avait pas d'embouteillage, [il rentrerait plus tôt]<sub>A</sub> [il mangerait avec ses enfants]<sub>B</sub>

Pour résumer, dans les constructions juxtaposées conditionnelles avec un verbe au conditionnel en A, le conditionnel perd la valeur d'accommodation modale<sup>88</sup> qu'il a habituellement.

**(ii)** Une analyse des cas où le verbe de A est à l'impératif montre également une relecture de la valeur modale habituelle d'ordre ou de conseil. Il s'agit d'exemples comme :

- (29) [Tombe en panne]<sub>A</sub> [on est morts]<sub>B</sub>  
 (42) [Avance]<sub>A</sub> [je te casse la figure]<sub>B</sub>  
 (43) [Ouvre le Guardian]<sub>A</sub> [tu trouveras trois coquilles par page]<sub>B</sub> (d'après Corminboeuf 2009)

Ce phénomène est relativement étudié et porte parfois le nom de pseudo-impératif<sup>89</sup>. Dans Dargnat (2008b [9], p. 2471) et plus spécifiquement dans Jayez & Dargnat (2009 [12]) et Dargnat & Jayez (2010b [14], p. 81-86), nous nous sommes penchés sur le cas de ces conditionnelles à l'impératif juxtaposées ou coordonnées par *et*<sup>90</sup>. Ceci nous a obligés à préciser le type de dépendance sémantique et sa représentation. Dans les exemples du type de (29) ou (42), il est exclu que l'impératif de A corresponde à un ordre ou à un conseil, bien au contraire. Il n'a pas de valeur qu'hypothétique. En (43), c'est un peu différent, car il ne serait pas contradictoire que le locuteur invite son interlocuteur à ouvrir le journal en question. Dans les deux cas, se pose la question de savoir si l'on peut dériver la valeur hypothétique de A d'après les valeurs habituelles de l'impératif. C'est la compositionnalité ou la constructionnalité de la juxtaposition AB

88 Je reprends cette idée à Roberts lorsqu'elle parle de subordination modale : « Modal subordination is a phenomenon wherein the interpretation of a clause is taken to involve a modal operator whose force is relativized to some set of contextually given proposition. » (1989, p. 718).

89 Voir par exemple Clark (1993), Culicover (1972), Culicover & Jackendoff (1997), Franke (2008) et Russel (2007).

90 Les mêmes remarques s'appliquent aussi aux cas où le verbe de A est à l'indicatif (*tu tombes en panne (et) on est morts*, etc.). On parle parfois de pseudo-déclaratives pour signifier la modification de la valeur modale.

qui sont en jeu. Est-il possible pour A d'avoir une valeur hypothétique en dehors de son insertion dans une construction comportant un constituant B et une relation conditionnelle entre A et B ? Notre traitement est constructionnel, c'est-à-dire que la valeur hypothétique de A n'est possible qu'à partir du moment où A est appréhendé dans sa relation à B et où cet agencement répond à certain nombre de contraintes. Pour dire les choses en les simplifiant beaucoup, la relation conditionnelle paraît plus « forte » entre A et B en juxtaposition que dans les phrases équivalentes non juxtaposées (par exemple en *si*). En termes de probabilité, la probabilité de « B sous condition A » est plus forte en juxtaposition que dans les autres cas. Pour exprimer cette idée, nous avons parlé d'*automaticité* (Jayez & Dargnat 2009 [12]) en référence à des notions comme la conséquence intrinsèque (*intrinsic consequence*) ou la construction modale suffisante (*Sufficiency Modal Construction* ou SMC).

Bolinger (1977, p. 162 sq.) parle de *conséquence intrinsèque* pour des cas qui relèvent des juxtaposées conditionnelles à l'impératif :

(44) Eat too much (and) you will get a stomach ache

(45) Eat your spinach (and) you'll be strong (d'après Bolinger 1977, ex. 101 et 105)

Dans ce cas, il remarque que la condition porte sa conséquence intrinsèque, dans le sens où le locuteur la présente comme inévitable, comme un résultat automatique.

La construction modale suffisante (SMC) de von Stechow & Iatridou (2007) ne traite pas explicitement des impératifs conditionnels du type de (44) et (45), mais de cas comme (46) correspondant à une conditionnelle anankastique (*anankastic conditional*) et (47) correspondant à une conjonction causale (*causal conjunction*).

(46) If you want good cheese, you only have to go to the North End.

(47) You only have to go to the North End and you will get good cheese. (von Stechow & Iatridou 2007, ex. 12 et 13, p. 447-448)

Leur analyse insiste sur le fait que la vérité d'une des deux propositions est suffisante pour déclencher la vérité de l'autre.

Nous avons récupéré ces deux idées et les avons appliquées aux juxtapositions conditionnelles à l'impératif en formulant une contrainte d'*automaticité*, qui prend en compte l'interlocuteur :

Given an agent *a* and a couple of eventualities *e*, *e'*, in which *a* participates, we say that *e'* is an automatic consequence of *e* with respect to *a*, if *e* causes *e'* and *e'* is not an action by *a*. (reformulation d'après Jayez & Dargnat 2009 [12], p. 252-253)

Cela tend à montrer que la relation entre la condition (événement *e*) et sa conséquence (événement *e'*) est automatique, quoi que fasse l'agent (*a*) à qui le locuteur s'adresse. Nous sommes allés plus loin dans la définition sémantique de la dimension causale de la condition en nous appuyant sur les travaux de Lewis<sup>91</sup>, mais je ne reprends pas cette partie du travail car ma contribution n'a été que ponctuelle.

Avant de rappeler brièvement quelques éléments concernant la connexion discursive et la prosodie, je me permets une remarque plus générale sur l'utilisation

91 Pour un panorama de son approche, voir D. Lewis. 2004. « Causation as influence ». In Collins J., Hall N., Paul L. A., *Causation and Counterfactuals*. Cambridge : MIT Press, p. 74-106.

que je fais de la notion de construction dans des travaux plus récents. J'ai abordé la question des constructions dans deux directions, qui correspondent à des exploitations aujourd'hui bien installées de la notion.

Premièrement, la direction diachronique. Les dernières années, la linguistique diachronique s'est emparée de la notion de construction pour mettre en évidence notamment le rôle du contexte (linguistique et situationnel) dans les processus de grammaticalisation et de pragmatization, et pour proposer une typologie des familles de construction sur la base d'héritage de traits (voir Traugott 2007, 2008, Trousdale 2012, 2015, Hoffmann 2013). Dans des travaux en collaboration (voir Combettes & Dargnat 2016b [26] et 2021d [42]), nous avons par exemple mis en évidence le rôle de certains contextes et la rémanence de quelques-unes de leurs valeurs pour l'interprétation de constructions concessives et causales, et des marqueurs discursifs associés (*d'autant que, si... que, dans la mesure où, en principe*, etc.) (voir section 3.2.2.1, p. 134).

Deuxièmement, la direction quantitative. En association avec la linguistique de corpus, un critère quantitatif a été avancé pour la recherche des constructions, en particulier des lexèmes complexes, à travers la notion de *collostruction* (voir par exemple Desagulier 2015, Gries 2019, Hilpert 2014). J'utilise cette perspective quantitative dans des recherches plus récentes sur les marqueurs de discours complexes (voir Dargnat & Jayez 2021b [40] et Dargnat 2021c [41]) (voir section 3.2.4.2, p. 151).

### 2.1.3 Connexion discursive et prosodie

La dimension prosodique intervient à différents moments de ma recherche : d'abord à propos des constructions juxtaposées et plus tard à propos des marqueurs discursifs (voir section 3.2.3, p. 142 et section 4.1.1, p. 159). Dans cette section, je synthétise quelques travaux qui cherchent à associer les propriétés prosodiques observées dans les juxtapositions à une configuration prosodique montante parfois décrite comme un contour continuatif majeur (section 2.1.3.1). Je présente aussi une expérience interrogeant en pratique et de manière systématique ces propriétés et la signification qu'on peut leur associer (section 2.1.3.2).

#### 2.1.3.1 Contour montant

Les propriétés prosodiques des juxtapositions ont été abordées dès mes premiers travaux (Dargnat 2008b [9]), p. 2468-2469, Dargnat & Jayez 2010b [14], p. 65-71) et elles ont été intégrées aux structures de traits comme un élément des constructions *juxta-prop* et *juxta-cond* (voir ici Figure 5, p. 60 et Annexe 1, p. 193). Deux questions ont émergé : celle de leur valeur en termes de fréquence fondamentale (F0) et celle d'une signification qu'on pourrait leur associer indépendamment de la construction.

– Les observations faites à la volée lors de ces premières études correspondaient globalement aux observations de Choi-Jonin & Delais-Roussarie (2006), à savoir une montée mélodique finale montante sur le premier constituant (A) et une absence de pause ou une pause minimale entre les deux constituants (A et B). C'est la raison pour laquelle, dans les structures de traits, la valeur de l'attribut CONTOUR-FINAL de CONS A a été notée *montant* par défaut (/).

– La signification associée à ce type de patron prosodique est dans plusieurs travaux celle d'une incomplétude discursive, c'est-à-dire que le segment se terminant par une

montée mélodique (ici une proposition) est présenté comme non terminé (voir Beyssade et al. 2006, Delais-Roussarie 2005, p. 99-121, Marandin 2006, Portes et al. 2007). Cette association forme prosodique/signification rejoint la notion de contour et dans le cas présent celle de *contour continuatif majeur*. On le trouve défini dès les travaux de Delattre (1966) et ensuite dans la plupart des travaux décrivant la prosodie du français (Chen 2007, Delais-Roussarie 2011, Di Cristo 2010, p. 5-23, Lacheret-Dujour & Beaugendre 1999, Martin 2015, p. 81, Morel & Danon-Boileau 1998, Rossi 1999, etc.).

Parler de *contour* prosodique implique, d'une part, qu'on se situe au niveau de groupes intonatifs et pas au niveau de valeurs atomiques et, d'autre part, qu'on leur associe une signification globale (voir l'idée des contours comme constructions, Marandin 2006). Je ne développe pas les différents modèles disponibles pour l'analyse de phénomènes prosodiques ni les différentes conceptions des contours comme séquences continues ou discontinues de variations tonales, car l'inventaire et le statut des contours comme « signifiants » sont assez variables<sup>92</sup>. Je reprends simplement quelques enjeux de l'association d'un contour montant continuatif considéré comme majeur pour l'étude des juxtapositions propositionnelles.

Pour le français, c'est effectivement à Delattre (1966) que l'on renvoie généralement pour l'idée de patrons prosodiques associés à des significations générales, qui marquent par exemple la complétude ou l'incomplétude du discours, des frontières de groupes comme le syntagme ou la proposition ou des modalités. Parmi les dix intonations qu'il met en évidence pour le français se trouvent trois configurations montantes. Sur une échelle à quatre niveaux, il représente ainsi les différences :

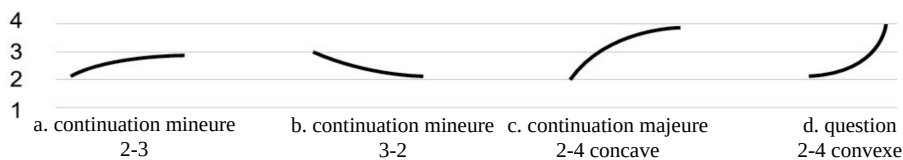


Figure 11. Les contours du français, d'après Delattre (1966, p. 4)

Le célèbre exemple repris en (48) expose la différence entre un contour montant dit mineur (*a* dans la figure ci-dessus et ↑ dans l'exemple ci-dessous) et le contour continuatif dit majeur (*c* dans la figure et ↑↑ dans l'exemple) :

(48) Si ces œufs ↑ étaient frais ↑↑, j'en prendrais. (Delattre 1966, p. 5)

La représentation et l'exemple appellent quelques questions qui peuvent être mises en rapport avec les relations de discours et la juxtaposition.

### (i) La frontière

Les contours mineurs et majeurs semblent s'opposer sur la nature du groupe dont ils marquent la frontière : un syntagme dans le premier cas, une proposition dans le second. Dans notre cas, on peut se demander si la frontière « majeure » correspond à un mode de dépendance (syntaxique, sémantique ou discursif) particulier entre les deux propositions.

92 Voir, en plus des travaux déjà cités, Beyssade & Marandin (2007), Beyssade et al. (2004), Di Cristo (2013), Jun & Fougeron (2000), Ladd (2008), Lacheret (2007), Portes et al. (2007), Portes et al. (2014).

**(ii) Caractéristique et spécificité des courbes**

La forme des courbes montre que les continuations mineures montent moins haut que les continuations majeures et qu'elles peuvent aussi être descendantes. Les configurations correspondant en théorie à des continuations majeures sont-elles en pratique toujours montantes ? D'autre part, une question et une continuation majeure se différencient-elles uniquement par le caractère concave/convexe de la courbe ?

**(iii) Montée mélodique, incomplétude et continuation**

Du point de vue de la signification, la continuation est associée à la perception d'une incomplétude (on attend une suite). Est-il possible de préciser cette incomplétude ? Autrement dit, le contour en lui-même est-il porteur d'une signification plus fine ?

Dans les juxtapositions, la spécificité du contour final du premier constituant et son assimilation au contour continuatif majeur à la Delattre ont été spécifiquement interrogées et testées pour le français dans Jayez & Dargnat (2011 [15], 2012c [18]), Dargnat et al. (2012a [16]) et Bartkova et al. (2012b [17]). La section ci-dessous expose le cadre expérimental et quelques résultats.

**2.1.3.2 Expérience**

Dans le cadre du projet *Intonale*<sup>93</sup>, que j'ai initié et coordonné, nous avons construit des expériences dans lesquelles nous provoquons la production de certaines configurations par des locuteurs de langue maternelle française. Nous souhaitons entre autres<sup>94</sup> tester systématiquement (i) si les propriétés prosodiques observées dans certaines de ces configurations correspondaient aux cas de continuations mineures et majeures telles qu'elles sont décrites par Delattre ; (ii) si, dans plusieurs configurations de fin de proposition, les différences prosodiques observées correspondaient à la distinction entre le contour de continuation majeure (théoriquement montant), le contour de question (théoriquement plus montant) et le contour de finalité (théoriquement descendant) ; (iii) si, dans les configurations prévues à cet effet (phrases complexes comportant deux propositions), on pouvait observer une différence au niveau prosodique en fonction du marquage ou non de la relation de discours, en d'autres termes si la juxtaposition avait un impact sur les propriétés prosodiques de la fin du premier constituant.

Pour la description de l'expérience et les principales conclusions relativement aux trois points ci-dessus, je m'appuie sur Dargnat et al. (2012a [16]).

- **Présentation du protocole**

Huit configurations ont été construites et cinq stimuli expérimentaux ont été élaborés pour chaque configuration. Elles sont reprises dans le tableau ci-dessous avec un exemple de base pour chaque cas.

Nom	Type de configuration visée	Stimuli
CIS	Configuration continuative sur SN sujet simple dans une phrase simple	Les <u>élèves</u> vont lire Malraux

93 *Intonale* (Production, perception et apprentissage des contours prosodiques en L1 et L2), financement CCOSL et ATILF (2009-2012).

94 Le projet *Intonale* était plus large que les questions retenues ici et comportait aussi un volet contrastif avec l'anglais L1 et L2. Je renvoie à Bartkova et al. (2012b [17]), Bartkova & Dargnat (2013 [20]) et (2014b [22]) pour ce dernier aspect.



<b>CIA</b>	Configuration continuative sur SN sujet simple dans la première proposition d'une phrase complexe	Les <u>élèves</u> vont lire Malraux c'est au programme du bac
<b>DIS</b>	Configuration déclarative conclusive sur SN simple (réponse à une question)	Les <u>élèves</u>
<b>QIS</b>	Configuration de question sur SN simple	Les <u>élèves</u> ?
<b>CAO</b>	Configuration continuative à la fin de la première proposition d'une phrase complexe avec mot coordonnant	Il relit <u>Maria</u> car elle fait trop de fautes
<b>CAP</b>	Configuration continuative à la fin de la première proposition d'une phrase complexe sans mot coordonnant (parataxe/juxtaposition)	Il relit <u>Maria</u> elle fait trop de fautes
<b>DAS</b>	Configuration déclarative conclusive à la fin d'une phrase simple	Il relit <u>Maria</u>
<b>QAS</b>	Configuration de question à la fin d'une phrase simple	Il a vu <u>Maria</u> ?

Tableau 6. Configurations et exemples de stimuli pour l'analyse des continuations prosodiques (d'après Darnat et al. 2012a [16], tableau 1 p. 1474)

L'expérience a été passée par 32 participants au moyen du logiciel *Corpus Recorder*<sup>95</sup> adapté pour l'occasion. Chaque stimulus était présenté dans un contexte mettant chaque participant en situation de prononcer un énoncé d'une des huit configurations avec la possibilité de s'écouter et de recommencer s'il trouvait sa production peu naturelle (voir ci-dessous la Figure 12)<sup>96</sup>. Après élimination des enregistrements inexploitable, nous avons récupéré 160 fichiers. Ces fichiers ont été segmentés et annotés manuellement en phonèmes et en mots avec le logiciel *Winsnoori*<sup>97</sup>. Les valeurs de la fréquence fondamentale (en semi-tons) ont été récupérées automatiquement toutes les 10 ms, ce qui a rendu possible le calcul de la pente pour les syllabes dans les positions ciblées pour chaque phrase enregistrée (mot souligné dans la colonne stimuli du Tableau 6).

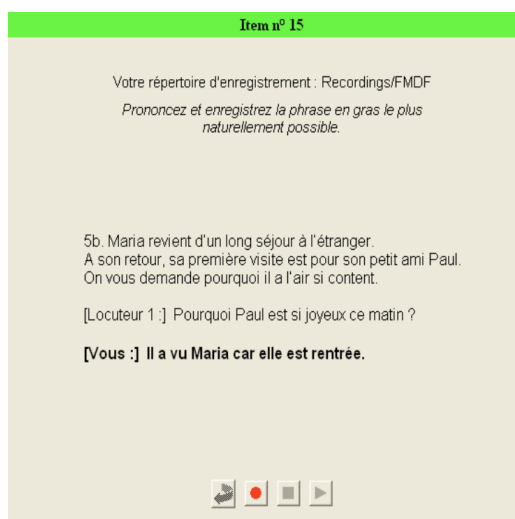


Figure 12. *Corpus Recorder*, exemple d'écran proposé aux participants

95 Ce logiciel a été développé au LORIA au sein de l'équipe Parole (Multispeech), mais il n'est pas diffusé. Pour l'expérience, il avait été adapté par Vincent Colotte.

96 Un processus de randomisaiton a été effectué pour éviter qu'un même sujet soit confronté deux fois à des énoncés trop semblables. Des distracteurs ont également été ajoutés.

97 *Win.Snoori*. Free Software for Speech Analysis. LORIA, Nancy. Version utilisée : 1.34. <https://members.loria.fr/YLaprie/WinSnoori/index.html> Pour les travaux ultérieurs, nous avons utilisé *Praat*.

Cette étape d'élaboration du protocole et de mise en place de l'expérience a été très formatrice. Elle m'a permis d'évaluer des intuitions linguistiques et de développer des collaborations avec des spécialistes de la prosodie dans le domaine du TAL. Les contraintes expérimentales qui visent à maîtriser le plus de paramètres possibles m'ont amenée vers un autre type de corpus que celui que j'avais pratiqué auparavant et ont été un moteur de réflexion important. La fabrication des stimuli s'apparente à un bricolage très concret et très local, mais soulève des questions fondamentales sur l'objet à tester et sur le but de la démarche.

- **Quelques résultats pertinents**

La construction de l'expérience permet de confronter les patrons prosodiques des configurations réellement prononcées avec les caractéristiques des contours de Delattre. Elle permet aussi un certain nombre de contrastes pour mettre en évidence des différences ou des ressemblances de réalisation entre les différentes configurations. Nous ne les avons pas toutes exploitées.

Tout d'abord, la hiérarchie globale des valeurs des pentes mesurées sur les mots-cibles (exemple souligné dans le Tableau 6) est la suivante. J'utilise le signe > pour indiquer que les valeurs d'une configuration sont le plus souvent supérieures à celle d'une autre configuration :

$$\text{QAS} > \text{CIA} > \text{CIS} > \text{CAO} > \text{CAP} > \text{DAS}^{98}$$

Je commente brièvement les résultats de contrastes qui apportent des éléments de réponse pour les trois questions mentionnées ci-dessus. Les différences effectivement observées ont été analysées avec un test de significativité<sup>99</sup> dont on peut tirer quelques conclusions.

- (i) **Contraste configurations CIS-CIA vs CAO-CAP** (voir ici Annexe 2a, p. 195)

L'hypothèse est que les configurations CIS-CIA correspondent à des continuations mineures (contour final sur un SN sujet du type *les élèves vont lire Malraux*) et que les configurations CAO-CAP correspondent à des continuations majeures (contour final de la première proposition dans une phrase complexe du type *Il relit Marie (car) elle fait trop de fautes*). On s'attendrait à ce que les éléments à la place de *élèves* et de *Marie* aient un contour majoritairement montant et que la montée sur les seconds soit plus prononcée que sur les premiers.

On remarque que les configurations pressenties comme continuatives majeures (CAO-CAP) sont hétérogènes et qu'elles sont loin d'être toutes montantes, à la différence des configurations pressenties comme continuatives mineures (CIAS-CIA) qui sont dans leur quasi-totalité montantes. De plus, dans les profils montants, ce sont les configurations pressenties comme mineures qui montent le plus.

- (ii) **Contraste CAO-CAP vs QAS vs DAS** (voir ici Annexe 2b, p. 195)

Dans les configurations de fin de proposition, l'hypothèse est que les différences prosodiques observées correspondent à la distinction entre le contour de continuation majeure (CAO-CAP, ex. *Il relit Marie (car) elle fait trop de fautes*) théoriquement concave

98 Les résultats sont présentés sous forme de diagrammes et de tableaux (Dargnat et al. 2012a [16], p. 1478-1479). J'en profite pour corriger une erreur dans la figure 4, p. 1478, DAS et QAS ont été malencontreusement intervertis dans la publication. Les graphes et les résultats complets corrigés sont reproduits ici dans l'Annexe 2, p. 195.

99 Il s'agit d'un modèle linéaire mixte paramétré sous R (package *lme4*).

montant, le contour de question (QAS, ex. *Il relit Marie ?*) théoriquement convexe et ‘plus’ montant et le contour de finalité (DAS, ex. *Il relit Marie*) théoriquement descendant. Cela conduit à interroger la sémantique intrinsèque des contours définis comme continuatifs majeurs (portent-ils en eux-mêmes l’idée d’incomplétude ?).

Dans le détail, on observe qu’il n’y pas de différence significative entre les configurations de type CAP (juxtapositions) et celles de type DAS (les assertions associées au contour dit de finalité chez Delattre), mais qu’il y a une différence significative entre les configurations de type CAP et les configurations de question (QAS). De leur côté, les configurations de type CAO (phrase complexe avec conjonction) ne diffèrent pas significativement de la configuration de question, mais diffèrent significativement de la configuration assertive<sup>100</sup>.

(iii) Contraste CAO-CAP (voir ici Annexe 2a ou 2b , p. 195)

Les configurations CAO et CAP ont été pensées pour pouvoir tester si la présence ou non d’un marqueur de discours – en l’occurrence la conjonction causale *car* – a un impact sur la réalisation du contour final de la première proposition. Autrement dit, si le contour de l’élément souligné change selon la présence ou non de *car* dans des phrases du type *Il relit Marie (car) elle fait trop de fautes*. L’hypothèse est que le contour continuatif serait plus « marqué » dans la construction juxtaposée.

Ce contraste montre bien une différence jugée significative entre la réalisation de la configuration avec conjonction (CAO) et la construction juxtaposée (CAP). Les configurations avec conjonction sont plus souvent montantes, ce qui n’était pas attendu, en tout cas pas dans ce sens, et les valeurs des pentes des configurations avec conjonction sont systématiquement supérieures aux valeurs pour les juxtapositions.

Pour conclure, il subsiste plusieurs questions. Premièrement, le contour déclaré dans la structure de traits *juxta-prop* est par défaut présenté comme montant, ce qui ne se reflète pas dans l’expérience. Toutefois, cela peut indiquer qu’un paramètre nous a échappé ou que nous ne l’avons pas étudié : par exemple la durée, l’intensité, la présence ou non d’une pause, la structure informationnelle des stimuli. Deuxièmement, le rapprochement entre les configurations CAP et QAS (différences jugées non significatives) pourrait correspondre à un élément commun qui serait le marquage prosodique d’une incomplétude. Pour les juxtapositions, cela a déjà été commenté, pour les questions, on peut tout à fait considérer que l’acte lui-même de question projette sa réponse comme une suite.

## 2.2 Relief argumentatif

Mes travaux et mes enseignements sur l’interprétation des textes, sur les relations de discours et sur les connecteurs m’ont fait rencontrer à différents niveaux la question de l’argumentation. C’est un domaine que j’aborde aussi régulièrement à travers la direction de travaux d’étudiants fonctionnaires stagiaires (mémoires de Master 2 MEEF). Mais ce n’est finalement que dans des collaborations récentes (Huber et al.

<sup>100</sup> La proximité des contours dits de continuation et des contours de question a aussi été observée dans une expérience de perception menée dans le cadre de Jayez & Dargnat (2011 [15], p. 226-229).

2019a [30], 2019c [32], 2020d [36]) et des co-directions de travaux d'étudiants que la question de la spécificité de l'argumentation s'est posée plus franchement dans ma réflexion<sup>101</sup>. Je mets en lumière les aspects qui m'ont été le plus directement utiles, en ouvrant sur des réflexions en cours (section 2.2.1), avant de présenter quelques points soulevés dans les travaux mentionnés (2.2.2).

## 2.2.1 Quelques paramètres de l'argumentation

Intuitivement, l'argumentation est une activité qui repose sur l'interprétation d'une proposition comme *orientée* en faveur ou en défaveur d'une autre proposition. La première sera perçue comme un argument pour ou contre le bien fondé de la seconde, appelée la thèse, la conclusion ou le *claim* selon les théories<sup>102</sup>. Le processus argumentatif renvoie généralement à l'intention qu'a le locuteur de modifier la croyance d'un interlocuteur vis-à-vis de cette thèse. Cela peut se faire de manière positive en défendant la thèse ou de manière négative (soit en la combattant soit en la suspendant).

### 2.2.1.1 Repères

L'argumentation est traditionnellement liée aux domaines de la rhétorique et de la logique et à ceux plus récents de l'analyse du discours, de la pragmatique et de la communication ; domaines dont les contours ne sont pas uniformes<sup>103</sup>. Je laisse de côté les débats épistémologiques qui ont parfois opposé les approches logiques formelles et les approches pragmatiques<sup>104</sup> (informelles) associées à un certain renouveau de la rhétorique (Perelman & Olbrecht-Tyteca 1958 et Toulmin 1958). Des travaux plus récents précisent ou enrichissent le cadre logique en proposant des outils pour décrire les réseaux et les types d'arguments. Par exemple, la perspective coalescente de Gilbert (1997, chap. 8) permet d'articuler plusieurs approches/dimensions ; la réflexion de Pollock (1995, chap. 3) intègre la non-monotonie et la comparaison de la force des arguments ; Freeman (1991, 2011) propose une typologie des fonctions des arguments dans la continuité de Toulmin ; les travaux autour de Walton se concentrent en particulier sur les situations dialogales et les discours juridiques (Walton & Krabbe 1995) et mettent en évidence des schémas argumentatifs stéréotypiques (Walton, Reed et Macagno 2008, Walton 2013). Enfin, les travaux de Stede et collègues se situent explicitement dans le domaine du traitement automatique et de la fouille d'arguments. Ils comparent notamment des annotations et des graphes issus de théories du discours

101 En particulier la co-direction de thèse de Laurine Huber et les stages de M2 TAL de Yi Ting Tsai et de Chaker Memmadi.

102 Les termes *argumentation* et *argument* ne sont pas uniformément définis selon les travaux et les traditions (voir Plantin 2002, p. 64-72). On distingue en général le processus et le produit argumentatifs. Ici, j'ai tendance à utiliser *argumentation* pour désigner l'activité discursive d'argumenter et le terme *argument* pour désigner le contenu propositionnel d'un (dans certains cas plusieurs) segments de discours en tant qu'ils sont perçus comme en faveur ou en défaveur d'une thèse.

103 Voir Amossy (2009), Gardes-Tamine (1996), Plantin (1990, 2003) et, pour un panorama historique et une confrontation des principales approches, van Eemeren et al. (2013) et Gilbert (1997, p. 3-27).

104 J'utilise le terme *pragmatique* à défaut d'une étiquette plus précise. J'entends par là que ces approches ne cherchent pas à mettre en évidence des relations vériconditionnelles entre propositions, mais envisagent un ensemble de facteurs qui interviennent dans un processus de persuasion, que ce soit au niveau des croyances ou des actions. Cela inclut par exemple le facteur émotionnel et la place de l'auditoire.

(RST et SDRT) avec une annotation en arguments ouvertement inspirée des travaux de Freeman (Peldszus & Stede 2013, 2016, Stede & Schneider 2018, Accuosto 2021).

L'étude de l'argumentation demeure donc un domaine actuel et productif qui suit l'évolution des méthodes et des outils utilisés en sciences du langage. Je relève quelques paramètres qui font sens par rapport à mes collaborations récentes ou qui font écho à des phénomènes abordés antérieurement, en particulier la dimension dialectique et dialogique et l'identification des arguments.

- **Les dimensions dialectique et dialogique**<sup>105</sup>

L'idée fondamentale est que le phénomène argumentatif sous-entend une confrontation de points de vue autour d'une proposition et que cette confrontation est prototypiquement associée à une situation interactionnelle entre un locuteur qui cherche à influencer son interlocuteur en faveur de la thèse qu'il défend et les réactions de ce dernier. Cette orientation argumentative ne se développe pas dans le discours en suivant des chemins nécessairement « tranquilles » et rectilignes. Le locuteur peut faire mine de se contrer lui-même en introduisant des arguments en défaveur de la proposition qu'il défend pour mieux les dépasser et ainsi renforcer ladite proposition. Cette technique peut être vue comme une anticipation de contre-arguments que l'interlocuteur pourrait avancer et comme une sorte de dialogue implicite intégré. Dans les situations explicitement dialogales, où locuteur et interlocuteur échangent réellement, des arguments en faveur ou en défaveur de la proposition défendue peuvent être apportés directement par l'interlocuteur. Des arguments positifs montreront son accord et sa coopération, des arguments négatifs le fait qu'il est en désaccord ou qu'il ne croit pas encore suffisamment à la proposition défendue pour l'accepter ou agir dans son sens. Pour résumer, l'étude des étapes et des circonvolutions d'un discours reconnu comme argumentatif et l'identification des arguments et de leur orientation reposent sur la perception de la structure dialectique – conflictuelle – qui suppose parfois une dimension dialogique – polyphonique – de ce discours. Envisager ces deux composantes, qui rendent possibles et manifestent en même temps l'argumentation, conduit souvent à privilégier les questions suivantes : (i) la spécificité de certains genres, (ii) le critère de l'intentionnalité et (iii) le rôle de l'auditoire.

(i) *Genres de discours*. Certains genres apparaissent comme plus argumentatifs que d'autres car ils exposent des thèses et des arguments de manière transparente. Ce sont toutes les situations délibératives, plus ou moins institutionnalisées, où l'échange et la confrontation d'arguments doivent aboutir à un jugement ou à une prise de position par rapport à des faits passés ou à venir<sup>106</sup>. Il en résulte un rapport de force ou une

105 J'utilise le terme *dialectique* dans un sens large pour désigner une pratique discursive mettant en jeu une confrontation de valeurs représentées voire incarnées par un défenseur et un adversaire relativement à une thèse. J'utilise le terme *dialogique* dans un sens assez large également pour renvoyer à la dimension intersubjective de tout discours, qu'il soit dialogal ou monologal, en particulier en référence aux travaux de Bakhtine (1984 [1979], par exemple la distinction dialogisme interlocutif *vs* dialogisme interdiscursif). Pour une synthèse des définitions, voir Bres (2005).

106 Le panel des situations prototypiquement argumentatives est aujourd'hui plus large que ce que la rhétorique désignait initialement par les genres judiciaire, délibératif et épideictique. Ces situations peuvent être publiques ou privées et le discours produit le fait d'un ou de plusieurs locuteurs. Mais l'idée de fond reste la même : il s'agit de tous les discours qui visent l'établissement d'un jugement, d'une position à tenir ou d'une action à mener suite à un échange

négociation plus ou moins marqués. Dans le domaine récent de la fouille d'arguments, Accuosto (2021, p. 18) distingue une douzaine de genres et de données associées particulièrement exploités : les échanges en classe, les essais, les textes juridiques, les discours politiques, les articles scientifiques, etc. Dans mes travaux, j'ai été confrontée à plusieurs types de données que l'on peut dire explicitement argumentatives, soit totalement, soit en partie : un extrait du corpus DÉCLICS consistant en un échange patient-médecin exposant un désaccord à propos de l'interprétation d'un résultat d'analyse (Dargnat 2022b [43] et ici section 3.2.2.2, p. 140, sur concession et interaction) ; le corpus *arg-microtexts*<sup>107</sup> de Peldszus et Stede (2015) dont le principe est de demander à des sujets une réponse courte et argumentée à une question polémique<sup>108</sup> (Huber et al. 2019a [30] et 2019c [32]) ; le corpus *Sci-DTB*<sup>109</sup> de Yang & Li (2018) qui est une banque de données de résumés scientifiques extraits d'articles publiés dans des conférences (Huber et al. 2020d [36]). J'y reviens de manière plus détaillée dans la section 2.2.2. (p. 83).

(ii) *L'intentionnalité*. Dans ce qui vient d'être rappelé, l'argumentation comme activité langagière suppose une intention chez le locuteur. Certaines situations prédisposent le locuteur à avancer à découvert, la forme que prend son discours posant explicitement et littéralement son intention. Ce sont par exemple les formules du type « dans cet article, nous souhaitons montrer que... », l'utilisation de relations de discours comme la justification ou la concession, ou les schémas dialectiques classiques des dissertations (thèse, antithèse, synthèse) devant répondre à une problématique formulée comme une question. Mais, pour être interprété comme argumentatif, un texte doit-il nécessairement déclarer son intention ? Et doit-il formuler le but qu'il vise littéralement ? Cela soulève pour moi quelques interrogations générales, auxquelles je ne prétends pas répondre mais qui me font relire certains de mes travaux plus anciens avec un relief argumentatif, en particulier certaines dimensions de mes analyses des textes de Tremblay.

- A un niveau général, toutes les productions langagières n'ont-elles pas un but ? Et, dans ce sens-là, ne comportent-elles pas une dimension argumentative de base ?
- La dimension argumentative d'un texte est-elle nécessairement consciente ? Autrement dit, est-il possible qu'un locuteur fasse de l'argumentation sans le savoir ? Ou qu'on lui attribue a posteriori une intention qu'il n'avait pas forcément, consciemment en tout cas ?
- La dimension argumentative d'un texte peut-elle être implicite ? Le but visé peut ne pas être aussi clairement affiché que dans les exemples mentionnés ci-dessus, et pourtant être bien réel. Dans ce cas-là, quels en sont les indices ?
- Peut-on parler d'argumentation *figurée* ? Je pense spécifiquement aux cas où l'on cherche à donner une valeur argumentative à un texte littéraire fictionnel dans sa globalité. En tant qu'évènement et forme faisant sens, le texte entier peut constituer un argument non littéral<sup>110</sup> en faveur ou en défaveur d'une thèse sociale, politique,

---

d'arguments en faveur ou en défaveur de ce jugement, de cette position ou de cette action.

107 Arg-microtexts : <https://github.com/peldszus/arg-microtexts>

108 Liste des 23 sujets abordés et des questions : lien précédent + [/blob/master/topics\\_triggers.md](#)

109 SciDTB : <https://github.com/PKU-TANGENT/SciDTB>

110 Certains passages des textes, certaines répliques des personnages peuvent énoncer assez littéralement la position de l'auteur ou des positions adverses. Le genre dramatique est d'ailleurs un choix judicieux pour matérialiser les dimensions dialectique et dialogique du discours de

culturelle ou esthétique, etc. par rapport à laquelle l'écrivain se positionne. Cette extension du cadre argumentatif peut être discutée, car elle semble dissoudre la spécificité de l'argumentation dans la discursivité (voir Amossy 2000, 2005)<sup>111</sup>.

(iii) *L'auditoire*. L'auditoire occupe une place importante dans la perspective argumentative. C'est l'efficacité de l'argumentation qui est en jeu. Le locuteur, dès lors qu'il souhaite accroître ou orienter la croyance de son interlocuteur en faveur de la thèse qu'il défend ou de la réaction qu'il cherche à provoquer chez lui, a tout intérêt à choisir ses arguments et à les mettre en forme en fonction de la situation d'interlocution et en fonction de son interlocuteur. L'idée d'une nécessaire adaptation à l'auditoire n'est pas une donnée nouvelle<sup>112</sup> et elle est récurrente dans les différents travaux que j'ai consultés. Les situations dans lesquelles sont produits des discours argumentatifs sont diverses et les auditoires potentiellement très variés et plus ou moins spécialisés (voir Tindale 2009). Les connaissances et les valeurs qui peuvent servir de ressources pour l'argumentation peuvent être plus ou moins étendues et plus ou moins partagées entre le locuteur et son auditoire, auquel il a accès directement ou qu'il projette. Elles relèvent de connaissances générales ou spécialisées sur le monde qui permettent d'obtenir l'adhésion rationnellement, mais aussi de connaissances plus personnelles qui pourront servir à cerner et à mieux toucher émotionnellement cet auditoire<sup>113</sup>. Perelman distingue plusieurs cas, avec l'hypothèse que plus l'auditoire est restreint, plus la tâche est facile pour le locuteur (1969, chap. 2). Ces réflexions montrent que la capacité du locuteur à se projeter à la place de son interlocuteur détermine en grande partie la recevabilité des énoncés formés et des arguments sélectionnés (les discours de meetings politiques sont à ce titre des exemples caricaturaux). Je retiens le fait qu'un discours argumentatif efficace rend compte non seulement d'une intention du locuteur mais aussi, sous un certain angle, de contraintes concernant l'auditoire, de ce qui est jugé « recevable » par lui du point de vue du locuteur. Je n'ai pas travaillé spécifiquement cet aspect pragmatique de l'argumentation, mais je l'ai retrouvé ponctuellement et il est à corrélérer à certaines approches ou notions que j'ai utilisées. J'insiste sur quatre points : (a) l'auditoire comme composante de la définition du genre discursif, (b) la place réservée à l'auditoire dans les approches cognitives comme la théorie de la pertinence et la théorie de l'esprit, (c) l'idée d'un interlocuteur comme prétexte structurant, (d) la place de l'interlocuteur (Reader) dans l'approche de la RST. (c) et (d) sont développées dans les ouvertures (section 2.2.1.2, p. 81) et dans la présentation d'expériences (section 2.2.2, p. 83).

(a) La prise en compte de l'auditoire dans sa diversité a des répercussions sur la constitution du message, dans sa forme et son contenu. Ici, le choix des arguments, de leur formulation et de leur agencement en discours sera un symptôme non seulement de l'intention du locuteur mais aussi de la recevabilité de ces arguments mis en forme et articulés dans un contexte donné. La projection de l'auditoire participe donc à la définition du genre discursif et à sa variation. Je peux corrélérer cela à plusieurs de mes

---

l'auteur.

111 Il me semble toutefois qu'elle est pertinente pour l'analyse que j'ai proposée de l'utilisation du jocal dans les textes de Tremblay et, de manière plus générale, parce qu'elle fait écho à la démarche d'une sémiotique/stylistique de la réception évoquée dans la section 1.1.

112 On la trouve déjà dans le concept d'*aptum* chez Aristote.

113 L'idée n'est pas nouvelle non plus, elle est à rapprocher des arguments relevant du *logos* et du *pathos*.

travaux : l'évolution des fonctions du jocal chez Tremblay concerne à la fois la pratique individuelle, la réception du phénomène et le renouveau du genre ; la constitution de corpus pour la fouille d'arguments nous a confrontés à la fois aux pratiques individuelles des auteurs, mais aussi au cadre plus ou moins imposé par les *templates* des revues et conférences ciblées, eux-mêmes variant en fonction des disciplines et du lectorat associé.

**(b)** Dans une théorie comme celle de la pertinence, l'efficacité est calculée « à réception » en fonction du coût cognitif pour l'interlocuteur. La projection et l'usage que fait le locuteur de la connaissance et des capacités cognitives et émotionnelles de son interlocuteur relèvent de ce qu'on appelle dans le domaine de la psychologie de la théorie de l'esprit<sup>114</sup>. Chercher à planifier les comportements de l'autre permet de s'exprimer en conséquence et, éventuellement, de mieux le convaincre.

**(c)** La présence d'un interlocuteur ou énonciateur implicite est utilisée dans plusieurs approches pour expliquer la structure des discours monologiques. Les travaux auxquels je pense ne visent pas spécifiquement l'argumentation, mais ils peuvent s'y appliquer de manière pertinente. Ils relèvent d'époques et de perspectives différentes, ne s'appellent pas forcément les uns les autres et n'ont pas la même granularité. Mais ils se croisent sur le besoin méthodologique de formuler une/des questions pour contextualiser, expliquer voire formaliser le statut informationnel des discours qu'ils analysent. Il s'agit du dialogisme de Bakhtine, déjà évoqué, de la problématique formulée par Meyer (1995), de la question en discussion (QUD) que l'on retrouve dans les travaux de Roberts (2012 [1996]) et Ginzburg (2012) et des questions potentielles proposées par Onea (2016).

**(d)** La RST prévoit dans les définitions des relations non seulement un rôle pour le locuteur (ou *Writer*) mais également un rôle pour un interlocuteur (ou *Reader*). Ce dernier n'intervient en fait que comme une projection du locuteur et pas comme acteur d'une interaction réelle. L'auditoire est une composante de l'effet produit/recherché par le choix de tel ou tel enchaînement discursif de la part du locuteur. On peut se demander si ce composant pragmatique (contextualisation interactionnelle) et cognitif (projection des connaissances partagées et des croyances de R) est pertinent pour toutes les relations de discours<sup>115</sup>. Et si la dimension rhétorique affichée est systématiquement argumentative.

#### • **Identification des arguments**

Une fois admis que l'on a affaire à un discours de type argumentatif, reste encore pour l'analyser à identifier les arguments (au sens de proposition en faveur ou en défaveur de la thèse défendue). Il peut y avoir de grandes variations d'une théorie à l'autre, d'une tradition à l'autre. Dans tous les cas, cela renvoie à une nécessaire segmentation du discours, à un typage des segments du point de vue de leur fonction argumentative et à une structuration de l'ensemble basée sur l'identification et le typage des relations entre ces segments. Ces différents paramètres peuvent être plus ou moins formalisés et

114 Certains travaux parlent de *vigilance épistémique* pour désigner la faculté d'un sujet (humain) à évaluer la qualité d'une information en fonction de la confiance qu'il a envers le sujet qui lui communique cette information. Voir Sperber et al. 2010. « Epistemic vigilance », *Mind and Language* 25 (4), p. 359-393.

115 Chez Renkema (2009), que l'on peut rapprocher explicitement de la RST, chaque définition de relation comporte un volet discursif et un volet dialogique (chap. 2). De plus, un certain nombre de relations de discours relèvent d'une catégorie nommée *Interjunction* (chap. 5).



schématisés. Le cadre retenu pour la description de la structure argumentative, décrit à la section 2.2.2, correspond à des choix qui pour moi soulèvent en amont plusieurs questions du point de vue linguistique sur la nature des arguments, leur orientation et leur qualité.

Dans la note 102 (p. 75), j'ai parlé des arguments en termes de contenu propositionnel associé à un segment de discours (un énoncé) ayant une fonction argumentative en faveur ou en défaveur d'une thèse, qui est l'acception assez commune du terme en français. L'analyse de Plantin (2002) montre que l'usage du terme dans la tradition anglo-saxonne diffère. Le fait que des terminologies ne soient pas tout à fait superposables ou transposables n'est pas exceptionnel, mais, dans mon cas, un éclaircissement s'est avéré nécessaire, dans la mesure où l'arrière-plan théorique retenu pour la fouille d'arguments renvoie à une définition qui n'est pas celle que j'ai spécifiée. Chez Freeman (1991), repris par Peldszus & Stede (2013), l'argument est défini comme un ensemble « prémisses(s) + thèse », où la ou les prémisses est/sont en faveur de la thèse<sup>116</sup>. De ce point de vue, un argument est minimalement binaire. Par exemple, dans un texte comme (49), l'argument désigne les deux propositions, la thèse (T) et la prémisses (P).

(49) [Ses parents sont des gens responsables]<sub>T</sub>. [Ils n'ont jamais baissé les bras face à ses difficultés scolaires]<sub>P</sub>.

L'argument peut être plus complexe (49') et comporter plusieurs prémisses (P1, P2, P3), liées ou non, en plus de la thèse (T).

(49') [Dans la mesure où ils ont toujours répondu présents]<sub>P1</sub>, [ses parents sont des gens responsables]<sub>T</sub>. [Ils n'ont jamais baissé les bras face à ses difficultés scolaires]<sub>P2</sub> et [ils l'ont soutenu financièrement après son divorce]<sub>P3</sub>.

Dans la littérature en français, le terme argument désigne plutôt le contenu propositionnel associé à un segment comme P en (49), P1, P2 ou P3 en (49') et ne renvoie pas à la thèse. Cela peut donc porter à confusion. Probablement par tradition et par sémantisme du terme lui-même en français, j'ai tendance à utiliser argument pour cibler le contenu propositionnel d'une prémisses.

Reste à savoir comment l'on perçoit et comment l'on rend compte de l'orientation des arguments ainsi définis. Deux directions sont généralement avancées, une orientation en faveur de la thèse et une orientation en défaveur de la thèse. Dans le deuxième cas, on parle parfois de contre-argument. L'identification de cette orientation est associée à la fois à l'identification d'une thèse défendue et à la perception de relations de discours, par exemple les relations **Condition**, **Justification** ou **Contraste** et **Concession**. Dans les faits, lorsqu'on est confronté à des textes, et de surcroît à leur annotation, il n'est pas toujours facile d'isoler au premier abord la thèse défendue, ni de percevoir les orientations et les niveaux où ces orientations agissent. Cela est beaucoup plus simple lorsque la thèse prend la forme d'une assertion liminaire et que tous les arguments sont introduits par un connecteur. Mais c'est loin d'être toujours le cas. Ajoutons à cela que les orientations s'imbriquent et que les valeurs en faveur et en défaveur peuvent s'inverser en fonction de ce sur quoi elles portent : par exemple, il est

116 « We define an argument to consist of a non-empty set of premises supporting some conclusion. We thus use the term 'argument' not for premises, but for the complex of one or more premises put forward in favor of a claim ». (Peldszus & Stede 2013, p. 6)

possible de formuler un contre-argument à un contre-argument, etc. Les structures concessives sont en cela de bons exemples de renforcement d'une orientation en faveur d'une thèse par l'anticipation d'un argument en sa défaveur. Ces problèmes d'identification et de récursivité ressortent avec évidence lorsqu'on cherche à représenter la structure argumentative d'un texte.

Enfin, dernier point, celui de la forme linguistique des unités argumentatives. Jusqu'à présent, il a plutôt été question du contenu propositionnel. Mais ce contenu s'instancie dans des énoncés réels et contraints par le code linguistique en jeu. Concrètement, ce sont les énoncés qui livrent le contenu propositionnel et guident vers la structuration discursive et argumentative. Cela est d'autant plus vrai dans les approches automatiques, qui partent des chaînes de caractères. Certains éléments de la langue sont de bons indicateurs, comme les connecteurs, des formes verbales, certains noms, certains adjectifs, parfois le registre ou le niveau de langage, certaines constructions, etc. Mais leur présence n'est pas systématique ou pas toujours facile à calibrer sémantiquement et pragmatiquement. Je ne développe pas cet aspect, je voulais juste le noter comme une interrogation et un champ qui m'intéressent : en quoi la « qualité » de certains éléments linguistiques, simples ou complexes, est-elle corrélée à la « qualité » argumentative de l'énoncé ? Ces réflexions renvoient à des études bien connues sur l'idée d'une pragmatique/rhétorique intégrée à la langue (Anscombe & Ducrot 1976) et, dans une certaine mesure, sur la contribution de la résonance émotionnelle des certains phénomènes linguistiques au raisonnement argumentatif (voir Plantin 2011, Villata 2018, Walton 1992). Les études sur la reconnaissance automatique des opinions et des sentiments sont aussi un lieu privilégié de traque des phénomènes linguistiques, notamment lexicaux, liés à l'expression des attitudes, de la subjectivité et l'argumentation (voir Bahtia et al. 2020, Breck & Cardie 2022, Eensoo & Valette 2014, Eensoo et al. 2016, etc.).

### 2.2.1.2 Ouvertures

Mes réflexions et travaux sur la juxtaposition, la fonction de la prosodie et la structure argumentative relèvent tous à un niveau plus général de l'étude de la cohérence textuelle et des conditions linguistiques et pragmatiques de sa bonne formulation. Les renvois ponctuels au dialogisme et à la problématique (Dargnat 2002 [3], introduction) et l'appel à la notion de *Question en Discussion* (QUD) (Dargnat & Jayez 2020e [37], Dargnat 2020f [38]) ont soulevé l'intérêt de penser la cohérence textuelle en termes de structure informationnelle pertinente par rapport à un contexte formulable comme une question. L'idée n'est pas nouvelle<sup>117</sup>. Elle n'est pas utilisée ni présentée de la même manière dans les trois approches mentionnées, ni même appliquée aux mêmes unités, mais elle est relativement claire : il s'agit de penser la valeur informationnelle du texte en ce qu'il se pose dans sa forme et son contenu comme une réponse pertinente (en quantité, qualité, etc.) à une question préalable.

Chez Meyer (1992), la perspective mêle philosophie, sociologie, théorie du langage et s'applique en particulier à la réception des textes littéraires. Le texte entier est appréhendé comme réponse à une question implicite que sa forme et son contenu

117 C'est une technique classique utilisée pour identifier la structure informationnelle à l'échelle de la phrase (en termes de thème et de rhème, par exemple) et pour mettre en évidence les enchaînements possibles. Voir par exemple Combettes (1983, p. 40-44) qui s'appuie sur les travaux de l'approche fonctionnelle de la phrase de Daneš et Firbas (voir Daneš 1974).

dessinent en creux. La question correspond à une problématique idéologique, esthétique, politique, etc. présente dans le monde contemporain de la production du texte. La réflexion se situe donc à un niveau très macro, le plus souvent figuré et vise une sorte d'herméneutique.

Chez Bakhtine (Volochinov), la réflexion est énonciative, elle articule remarques linguistiques et théorie littéraire. Le schéma question/réponse est à la base de la conception du dialogisme, mais ne le résume pas, car ce sont les concepts plus généraux de polyphonie et d'intertextualité qui lui permettent de faire de la langue en usage un ensemble de ressources poreuses aux différentes voix qui y ont recours. La reconstruction de questions implicites est un moyen concret proposé comme une méthodologie pour expliquer le découpage des textes en paragraphes. « Il s'agit, écrit-il, de dialogues affaiblis et transformés en énonciations-monologues. A la base de la division en parties, dénommées paragraphes dans leur forme écrite, on trouve l'ajustement aux réactions prévues de l'auditeur ou lecteur. Plus l'ajustement à l'auditeur et la prise en compte de ses réactions sont faibles, plus le discours sera d'un bloc et moins il y aura de paragraphes. » (Bakhtine 1977 [1929], p. 58-159). Je retiens de cette remarque l'insistance sur la prise en compte de l'auditoire et son intégration dans la forme-même des enchaînements des parties d'un énoncé complexe. On peut se demander s'il est possible d'étendre cette réflexion au domaine des relations de discours et de l'argumentation, en assimilant les segments de discours et les arguments à des paragraphes et les questions à des requêtes d'enchaînements d'un certain type appelant un certain contenu.

La théorie de la QUD n'émane pas des mêmes traditions, elle est plus récente et relève explicitement du domaine de la sémantique et de la pragmatique. Elle est d'abord associée aux travaux de Roberts et connaît un certain retentissement dans l'analyse de phénomènes discursifs comme le rôle de la prosodie dans la structuration informationnelle, la sémantique du dialogue et plus récemment les relations de discours (voir Benz & Jasinskaja 2017, Ginzburg 2012, Onea 2016 et Roberts 2012). Elle systématise et formalise le principe d'un jeu question-réponse comme un outil pour analyser ce qui fait qu'un énoncé est approprié à (*fit*) son contexte. Le principe est que, dans un discours, chaque énoncé est présenté comme répondant à une question, souvent implicite, soit en lui apportant une réponse, c'est-à-dire en comblant l'incertitude soulevée dans la question, soit en soulevant une autre question pouvant aider à combler cette incertitude. Rapportée au domaine de l'argumentation, la QUD correspond alors à un ensemble de thèses alternatives entre lesquelles le locuteur défend un choix. Je n'ai pas (encore) étudié ni appliqué spécifiquement cette théorie à l'analyse des textes argumentatifs, mais il me semble qu'il serait intéressant de s'interroger d'un point de vue théorique sur l'utilité de penser la spécificité du processus argumentatif dans ce cadre. Pour simplifier, on peut attribuer à un discours une qualité argumentative quand on pense que le locuteur cherche à diminuer l'incertitude soulevée par une QUD tout en estimant que son interlocuteur est susceptible de ne pas accepter la résolution de la QUD qu'il propose.

## 2.2.2 Analyses et représentations : vers le TAL

Cette section rend compte de travaux collaboratifs récents qui s'inscrivent dans le domaine du traitement automatique du langage et plus spécifiquement de la fouille de textes et de la fouille d'arguments (Huber et al. 2019a [30], 2019c [32] et 2020d [36], Tsai 2020). L'enjeu pour moi est d'abord celui d'une comparaison des structures discursive et argumentative de textes et des questions et problèmes soulevés au regard des données et de la méthodologie. L'une de ces questions porte sur la possibilité de la prédiction la structure argumentative d'un texte à partir de sa structure discursive<sup>118</sup>.

### 2.2.2.1 Discours, argumentation et représentation de la structure textuelle

Une comparaison automatique de structures repose sur des prérequis concernant les données, notamment leur quantité, leur accessibilité, la cohérence et la comparabilité des annotations et de leur représentation graphique. Nous avons utilisé des données disponibles déjà annotées et représentées pour les deux types de structures (discours, argumentation). L'enrichissement des données brutes a supposé dans les deux cas une segmentation en unités, une identification des attachements entre ces unités et un typage de ces attachements. L'approche discursive est basée sur le système de la *Rhetorical Structure Theory* (RST) et l'approche argumentative sur le système de Peldszus et Stede. Je rappelle les éléments qui nous ont été utiles, sans discuter les entreprises générales, car elles sont bien documentées.

- **Le système de la RST**

Depuis ses débuts (Mann & Thompson 1988), la RST a subi plusieurs développements (Taboada & Mann 2006, Stede & Taboada 2017). Il s'agit d'une communauté toujours très active. Son objectif est la description de la cohérence des textes de tout genre sur la base de leur structuration en unités et des liens sémantiques et pragmatiques entre ces unités, qui sont appelés relations rhétoriques ou relations de cohérence. La connexion entre les unités est hiérarchisée et est représentée sous forme de graphes arborescents. Des données déjà annotées et représentées sont disponibles et il est possible également de structurer, d'annoter et de représenter des données brutes au moyen d'outils comme *RSTTool* ou *rstWeb*. La Figure 13 a été réalisée avec ces outils. Elle schématise la structure textuelle<sup>119</sup> de cinq unités numérotées, dont le titre.

---

118 La programmation et le traitement des données sous Python a essentiellement été fait par L. Huber, doctorante en informatique, et deux stagiaires de M2 TAL, C. Memmadi et Y. T. Tsai, co-encadrés avec Y. Toussaint.

119 Il s'agit d'un exemple traduit de l'anglais qui sert d'illustration sur le site de la RST.

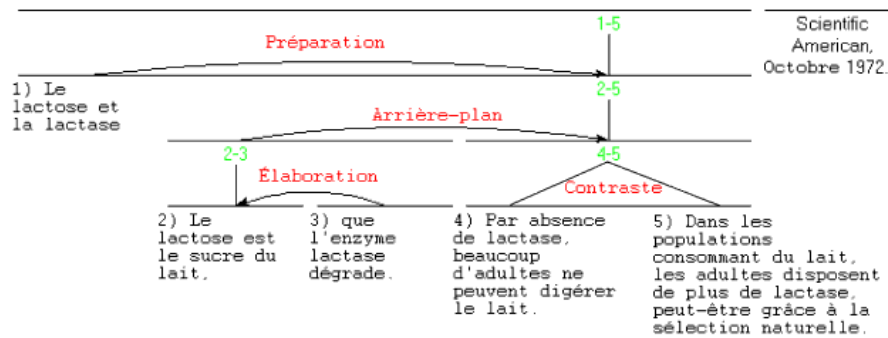


Figure 13. Exemple de graphe RST (repris du site)

*Légende*

Les couples de chiffres en vert correspondent à des blocs textuels composés des unités indiquées (2-5 signifie : bloc comportant les unités 2 à 5), il s'agit d'unités complexes. Les flèches sont orientées de l'unité satellite vers l'unité noyau, sauf dans le cas d'une relation entre deux unités noyaux. Les labels en rouge sont les noms des relations entre les unités indiquées.

En plus de son accessibilité et de son caractère relativement intuitif, l'intérêt de cette approche pour moi a principalement été (i) de poser la question de la segmentation du texte et de la nature des segments reliés et (ii) d'interroger les critères de définition des relations et le mode de dépendance entre les segments reliés.

(i) *Segmentation*. Les unités peuvent être simples (*elementary discourse units*, EDU), par exemple [1], [2], etc. dans la Figure 13, ou complexes (*text spans*), par exemple le bloc [4-5]. Le plus souvent, une EDU correspond à une structure prédicative unique<sup>120</sup> et une unité complexe à un ensemble d'EDU contiguës. La RST distingue des unités *nucléaires* (N) et des unités *satellites* (S). Elle s'appuie pour cela sur un critère informationnel : la conservation ou non du message principal, en cas de suppression de l'unité testée. Si l'on supprime le noyau, l'information principale n'est pas récupérable, alors que si l'on supprime le satellite, l'information principale est toujours accessible. Cette configuration noyau-satellite (*schéma*) répond à des contraintes qui ont des conséquences sur la manière d'appréhender la cohérence et la structure textuelles. Dans la Figure 13, le schéma reliant les unités [2] et [3] présente l'unité [3] comme le satellite de l'unité [2], selon les contraintes de la relation *Élaboration*. Il existe également des schémas ne mettant en œuvre que des unités nucléaires, c'est le cas de l'unité complexe [4-5], où les deux unités minimales constituent chacune un noyau, une information essentielle à la communication du message principal, en suivant les contraintes de la relation *Contraste*.

(ii) *Relations, hiérarchie et récursivité*. Pour sélectionner la relation, la théorie propose une typologie et des labels sur la base de critères sémantiques et pragmatiques. La liste, en apparence finie, est présentée comme ouverte et non prescriptive, et elle est adaptable et modifiable pour correspondre aux besoins de chaque utilisateur. La version

<sup>120</sup> Des discussions existent sur le statut des propositions compléments (subordonnées complétives) – la dépendance argumentale n'étant pas considérée comme une relation de discours –, sur les propositions relatives, qui peuvent être considérées comme relevant de la relation *Élaboration*, sur les propositions constituant du discours rapporté, qui ne sont pas considérées comme des EDU, et sur les ajouts propositionnels (les subordonnées circonstancielles), qui sont de bons candidats pour être des EDU, avec toutefois quelques restrictions (voir Stede et al. 2017, p. 5-7).

« de base » labellise 32 relations et les classe en trois types : des relations de présentation (par ex. *Arrière-plan*, *Concession*, *Justification*), des relations de contenu (par ex. *Condition*, *Élaboration*, *But*) et des relations multinucléaires (par ex. *Contraste*, *Liste*, *Disjonction*). Les deux premières reposent sur un schéma noyau-satellite, la troisième sur un schéma où chaque unité est un noyau. Leur définition consiste en une liste systématique de contraintes sémantiques et pragmatiques sur chaque unité simple ou complexe en jeu, sur la combinaison des unités et sur l'effet recherché par le locuteur sélectionnant cette relation. L'expression des contraintes touche en particulier à l'intentionnalité, aux états de croyances du locuteur et à l'attribution par le locuteur d'états de croyances de l'interlocuteur. La structure du texte est donc envisagée en RST du point de vue de la communication dans une perspective orientée dès le départ vers les états mentaux et l'argumentation. Le manuel d'annotation de Stede et al. (2017) abonde dans ce sens puisqu'il y est question d'une perspective « à réception », introduisant une part de subjectivité et des variations d'interprétation, et donc la possibilité d'annotations concurrentes. On retrouve chez ces derniers les grands principes des tableaux proposés sur le site de la RST<sup>121</sup>, avec quelques aménagements et des discussions sur les difficultés d'une annotation concrète. Les labels et le classement diffèrent sur quelques points. Les auteurs proposent deux critères non exclusifs pour regrouper les relations : un critère structurel autour de la nucléarité (schéma noyau-satellite dit mono-nucléaire ou schéma multi-nucléaire) et un critère prenant en compte le contenu et la portée de la relation, qui peuvent être sujets à préférence (relations avant-tout (*primarily*) pragmatiques, sémantiques et relations textuelles). Le format de définition rejoint celui de la version de base quant aux contraintes mais propose en plus d'indiquer le connecteur prototypiquement utilisé pour marquer la relation définie, un exemple décomposé avec, le cas échéant, un contexte d'interprétation. Par exemple, la relation conditionnelle est définie ainsi :

(50) Condition

**S:** A hypothetical, future, or in other ways unreal situation.

**N/S:** The realization of N depends on the realization of S.

**Effect:** R (the reader) recognizes how the realization of N depends on the realization of S.

**Typical connectives:** *if... , then ; in case ; ...*

**Example:** [When more and more communities are in financial trouble,]S  
[regularly replacing the school textbooks becomes just impossible.]N  
(maz00002, PCC) (Stede et al. 2017, p. 15)

D'autres travaux s'inspirant ouvertement de la RST, comme ceux de Renkema (2009), proposent une vision plus englobante et plus systémique. L'intérêt de son modèle dit de la *connectivité* est de penser l'effet-texte dans sa globalité et d'articuler une perspective textuelle et une perspective dialogique (*discourse and dialogic principles*). La cohérence d'un texte y repose sur des phénomènes dits de Conjonction (*Conjunction*), d'Adjonction (*Adjunction*) et d'Interjonction (*Interjunction*). Les relations de discours sont listées et traitées au niveau de l'Adjonction et de l'Interjonction. Dans le premier cas, c'est le contenu sémantique des propositions qui est jeu, dans le second cas, ce sont les propriétés interactionnelles qui sont visées. Bien que différant de la RST dans sa

121 <https://www.sfu.ca/rst/07french/definitions.html>

présentation et son empan, on retrouve les grandes catégories des relations, leurs composantes sémantiques et pragmatiques et l'orientation argumentative pour celles qui relèvent de l'Interjonction. Ce modèle m'a été utile pour caractériser les relations de discours que j'ai étudiées dans une taxinomie « serrée », pour en questionner les dimensions interactionnelles (notamment dialogiques) et subjectives, et pour situer la question de la présence d'un connecteur parmi d'autres phénomènes de cohésion servant la cohérence.

- **Structure argumentative : le système de Peldszus & Stede**

Ce système est explicitement basé sur les travaux de Freeman (1991 et 2011). Il a été développé en particulier pour l'annotation de données authentiques recueillies dans le cadre du projet *arg-microtexts*<sup>122</sup>, et a fait l'objet de plusieurs présentations et d'un guide pour les annotateurs (Peldszus & Stede 2013, 2015 et 2016, Stede et al. 2016). La visée applicative et la nature des données ont nécessité quelques adaptations et ont soulevé des questions d'interprétation. La représentation sous forme de graphes suppose ici aussi plusieurs étapes : la segmentation en unités possédant une fonction argumentative, l'attachement de ces unités et le typage des relations d'attachement. Je résume ci-dessous les éléments théoriques essentiels pour permettre la lecture des graphes et comprendre l'intérêt et les limites d'une comparaison avec le système de la RST. Il s'agit de (i) de l'opposition promoteur et adversaire, (ii) des relations de support et d'attaque et (iii) de la portée et de la combinaison des relations.

(i) *Promoteur et adversaire*. La structuration argumentative se fait en supposant un modèle dialectique prototypique où il est possible de distinguer dans le discours des positions et des propos associés relevant du promoteur (*proponent*) et de l'adversaire (*challenger*) d'une thèse. Il s'agit de rôles énonciatifs entrant dans le moule dialectique et dialogique évoqué plus haut (voir note 105, p. 76), qui ne s'instancient pas forcément dans un échange réel entre un locuteur et son ou ses interlocuteur(s). Du point de vue de l'argumentation, l'identification de ces rôles est importante, ce qui explique qu'elle soit matérialisée dans les représentations graphiques. L'unité argumentative considérée est cerclée quand elle est associée au rôle de promoteur, et elle est encadrée quand elle est associée au rôle d'adversaire.

(ii) *Relations de support et d'attaque*. Concrètement, la structuration consiste à expliciter et à placer les relations entre les segments de discours énonçant ces positions et la thèse. Ces relations sont soit en faveur soit en défaveur de cette thèse ou conclusion. A un premier niveau, on parlera d'une relation Support lorsque le segment est en faveur de la thèse, et d'une relation Attack lorsque le segment est en défaveur de la thèse. Ces deux orientations vis-à-vis de la thèse sont rendues différemment dans les graphes : une flèche avec une pointe triangulaire marque une relation Support entre l'unité d'où elle part et l'unité où elle pointe, une flèche avec un point en forme de cercle marque une relation Attack entre l'unité d'où part le trait et l'unité que touche le cercle. Par exemple, le texte de l'exemple (51) pourra être représenté par la Figure (14).

---

122 <https://github.com/peldszus/arg-microtexts>

(51) [We should tear the building down.]<sub>1</sub>  
 [It is full of asbestos.]<sub>2</sub>  
 [On the other hand, many people liked the view from the roof.]<sub>3</sub>

*Il vaudrait mieux démolir le bâtiment. Il est bourré d'amiante. D'un autre côté, beaucoup de gens aimaient la vue depuis le toit.*

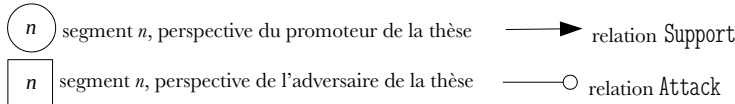
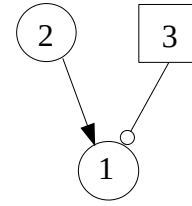


Figure 14. Schématisation de la structure argumentative de l'exemple (51)

**(iii) Portée, point d'attachement et configurations complexes.** Les discours argumentatifs sont généralement plus complexes que celui présenté dans l'exemple (51). Les positions en faveur et en défaveur de la thèse peuvent prendre des formes plus élaborées et moins directes. Il existe plusieurs configurations pour les relations Support et les relations Attack. Peldzsus & Stede (2013) en distinguent cinq dans le premier cas et quatre dans le second<sup>123</sup>. Dans les graphes, cela correspond à des lieux d'attachement différents. La Figure 15 en illustre deux de chaque.

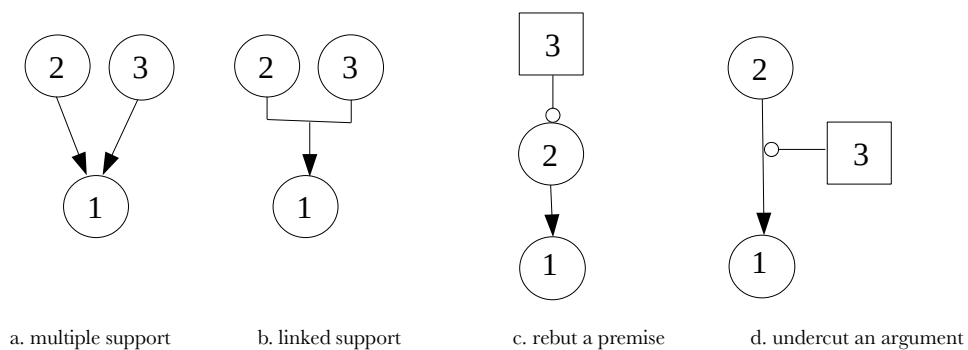


Figure 15. Exemples de configurations de type Support et Attack

J'insiste sur la différence entre les Figures 15c et 15d. La Figure 15c montre une relation Rebut (réfuter) qui porte sur une unité, une prémisses ici, correspondant au segment n° 2<sup>124</sup>. La Figure 15d montre une relation Undercut (que l'on peut traduire par « casser ») qui consiste ici à attaquer la relation de support entre une unité (segment n° 2) et la thèse principale (segment n° 1).

Les relations d'attaque ne sont pas réservées à l'adversaire de la thèse. Le promoteur peut aussi les utiliser pour répondre aux attaques elles-mêmes et, du même coup, renforcer la thèse défendue<sup>125</sup>. La relation Counter-attack (contre-attaque) est alors utilisée. On se trouve donc dans des situations où les relations Rebut et Undercut sont en

123 Configurations de support : *basic argument*, *linked support*, *multiple support*, *serial support* et *example support*. Configurations d'attaque : *rebut a conclusion*, *rebut a premise*, *undercut an argument* et *supporting a rebutter*. (Peldzsus & Stede 2013, p. 7-9)

124 Dans la Figure 14 précédente, la relation Rebut portait directement sur la thèse (segment n° 1).

125 Configurations de contre-attaque : *rebut a rebutter*, *rebut an undercutter*, *undercut a rebutter* et *undercut an undercutter* (Peldzsus & Stede 2013, p. 10-11).



quelque sorte positivement orientées par rapport à la thèse défendue. Pour illustrer cela, on peut reprendre la Figure 15d, en imaginant que le promoteur va chercher à casser l’attaque adverse en formulant une nouvelle unité argumentative. Appliqué à l’exemple (51), cela donne l’enchaînement (52) et la Figure (16) ci-dessous. La nouvelle unité ajoutée correspondant au segment [4].

- (52) [We should tear the building down.]<sub>1</sub>  
 [It is full of asbestos.]<sub>2</sub>  
 [In principle it is possible to clean it up.]<sub>3</sub>  
 [but according to the mayor that would be forbiddingly expensive]<sub>4</sub>

*Il vaudrait mieux démolir le bâtiment. Il est bourré d’amiante. On devrait pouvoir le nettoyer, mais d’après le maire ce serait excessivement coûteux.*

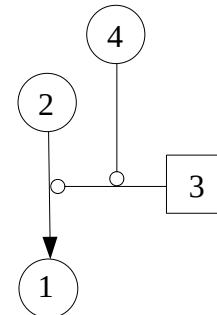


Figure 16. Schématisation de la structure argumentative de l’exemple (52)

Dans certains cas, des attaques associées à l’adversaire peuvent être laissées en suspens, définitivement ou momentanément. Cette configuration est étiquetée *Counter-consideration* (une non prise en compte de l’unité attaquante).

La combinaison de ces configurations permet de rendre compte d’un grand nombre de structures argumentatives. L’expérience d’annotation de données authentiques a cependant rendu nécessaires des adaptations du système. Les variations stylistiques souvent à des fins expressives, la dilution ou répétition de la thèse dans plusieurs unités et l’absence de fonction argumentative de certaines unités peuvent compliquer les tâches de segmentation et d’attachement. Trois opérateurs ont été ajoutés pour régler ces problèmes. Je ne les détaille pas mais note tout de même leur intérêt pour la catégorisation de l’argumentation dans le langage ordinaire<sup>126</sup>.

#### • Intérêts de ces approches pour ma réflexion

La présentation même rapide de ces approches ne correspond pas (complètement) au souci scolaire de leur exposition. Elle m’a permis, dans le cadre de cette habilitation, de pointer plus précisément des aspects jusqu’alors demeurés un peu vagues ou sous-exploités dans mes travaux et donc d’avancer dans ma réflexion, mes collaborations et mes projets.

Assez longtemps, j’ai associé argumentation et structure discursive de manière un peu floue. Plus exactement, j’avais tendance à attribuer une intention argumentative à toute structuration discursive, à teinter argumentativement toutes les relations de discours. Cela est probablement dû au cadre de la RST, dont la définition des relations met en évidence l’intention du locuteur et les contraintes de réception, et aux relations sur lesquelles je me suis penchée, qui sont des ressources précieuses pour construire des arguments (relations de type *Cause*, *Justification*, *But*, *Concession*, *Condition*, etc.). Toutes les relations de discours ne sont pas nécessairement argumentatives – certaines sont plus prédisposées que d’autres (voir Azar 1999) – et les relations argumentatives sont moins

<sup>126</sup> L’opérateur *glue* permet littéralement de coller des segments adjacents formant une seule unité argumentative (on y retrouve les ajouts après le point). L’opérateur *restatement* (réaffirmation) permet d’associer deux segments non-adjacents exprimant le même contenu propositionnel et ayant la même fonction argumentative (par exemple, la répétition de la thèse). Enfin, l’opérateur *skip* permet de sauter un segment sans fonction argumentative.

nombreuses que les relations de discours. L'analyse séparée et comparée plus systématique et plus fine des deux perspectives montre au moins deux différences majeures au niveau des relations et de la structuration. La structuration discursive diffère de la structuration argumentative notamment en ce qui concerne la taille des segments pris en compte. Les segments argumentatifs peuvent comporter plusieurs segments discursifs, rarement l'inverse. Elles diffèrent aussi en ce qui concerne la linéarité de l'analyse. La structuration discursive repose sur la contiguïté des segments et sur la prise en compte de tous les segments, alors que ce n'est pas le cas pour la structuration argumentative (certains segments peuvent être laissés de côté, ou deux segments non contigus peuvent être réunis).

### 2.2.2.2 *Expériences*

Je présente maintenant deux analyses relevant de la fouille de données textuelles en relation avec les deux approches qui viennent d'être rappelées. La première est une comparaison de graphes, la seconde utilise la technique des plongements (*embeddings*).

- **Comparaison de graphes discours-argumentation**

Dans Huber et al. (2019a [30], 2019c [32]) et (Tsai 2020), nous avons proposé une comparaison de la structure discursive et de la structure argumentative, en repartant des études et des données annotées par Peldszus & Stede (2013, 2016). Nous avons utilisé la version en anglais des 112 textes du corpus *arg-microtexts* déjà mentionné. Ces textes, produits à des fins d'analyse, sont des réponses à un ensemble de questions volontairement polémiques<sup>127</sup>. Ils constituent donc un ensemble assez homogène et propice à la double analyse discursive et argumentative. Dans les deux cas, l'annotation et la structuration sont représentables et disponibles sous forme de graphes. Nous avons retenu les structures et graphes obtenus dans le cadre de la RST<sup>128</sup> et les structures et graphes obtenus dans le cadre de la théorie de l'argumentation des auteurs. Même si elle est contrainte par le cadre choisi, le type d'annotation et de représentation, la comparaison des structures discursives et argumentatives peut se faire à différents niveaux et peut servir plusieurs types d'hypothèses. Par exemple, Peldszus & Stede (2016) se sont concentrés sur l'alignement des relations deux à deux. Il s'est avéré que la correspondance n'est pas systématique et que cela est parfois dû aux différences d'expressivité des deux approches et des représentations associées. Notre démarche a consisté à interroger non pas la correspondance entre relations de discours (par ex. **Condition**, **Contraste**, **Élaboration**, etc.) et relations argumentatives (**Support**, **Rebut**, **Undercut**, etc.), mais la correspondance entre sous-graphes de discours et sous-graphes argumentatifs, c'est-à-dire entre « morceaux » de structures plus complexes mettant en jeu des combinaisons de relations<sup>129</sup>. Le processus d'analyse est décomposable en trois étapes : (i) l'encodage des structures discursives et argumentatives dans un

127 Rappel : [https://github.com/peldszus/arg-microtexts/blob/master/topics\\_triggers.md](https://github.com/peldszus/arg-microtexts/blob/master/topics_triggers.md)

128 Le corpus est également annoté dans le cadre de la SDRT, mais, comme nous ne voulions pour le moment pas « descendre » dans une sémantique plus fine des segments, nous nous sommes concentrés sur la segmentation, les relations de discours et la représentation hiérarchique proposée par la RST.

129 Dans le corpus *arg-microtexts*, les relations de la RST les plus fréquentes sont, par ordre décroissant : **Reason**, **Concession**, **Conjunction**, **Antithesis**, **Elaboration** et **Cause**. Les relations argumentatives les plus fréquentes sont, par ordre décroissant : **Support**, **Rebut** et **Undercut**.

format comparable, (ii) l'extraction des sous-graphes fréquents et (iii) la fouille de redescriptions.

(i) *Production de graphes RST et ARG « arrangés »*

Dans le corpus utilisé, chacun des 112 textes produits est initialement associé à deux structures représentées par deux graphes. L'exemple (53) servira d'illustration. Les segments qui servent de base à la structuration discursive et la structure argumentative sont entre crochets et numérotés.

(53) arg-microtexts, micro\_b\_006

*Topic* : Should Germany introduce the death penalty?

The death penalty is a legal means that as such is not practicable in Germany[1]. For one thing, inviolable human dignity is anchored in our constitution[2], and furthermore no one may have the right to adjudicate upon the death of another human being[3]. Even if many people think that a murderer has already decided on the life or death of another person[4], this is precisely the crime that we should not repay with the same[5].

*Thème/question* : L'Allemagne devrait-elle introduire la peine capitale ?

La peine de mort est un moyen légal qui en tant que tel n'est pas applicable en Allemagne[1]. D'une part, l'invulnérabilité de la dignité humaine est un pilier de notre constitution[2] et, d'autre part, personne ne peut avoir le droit de décider de la mort d'un autre humain[3]. Même si beaucoup de gens estiment qu'un meurtrier a lui déjà décidé de la vie ou de la mort d'une autre personne[4], c'est précisément le crime qui ne devrait pas être payé du même prix[5]. (trad. personnelle)

Les graphes initiaux sont reproduits en annexe (voir ici Annexe 3, p. 196). Même si la segmentation de base en unités élémentaires est la même, les deux démarches comportent un certain nombre de différences (labels, portée des relations, regroupement, contiguïté ou discontinuité des segments, nécessaire prise en compte de tous les segments ou possibilité de laisser des segments de côté, etc.). Une première étape a donc consisté à rendre les graphes initiaux compatibles pour pouvoir les comparer. Pour cela, deux transformations ont été nécessaires. Les Figures 17a et 17b en rendent compte<sup>130</sup>. Les numéros correspondent aux segments textuels.

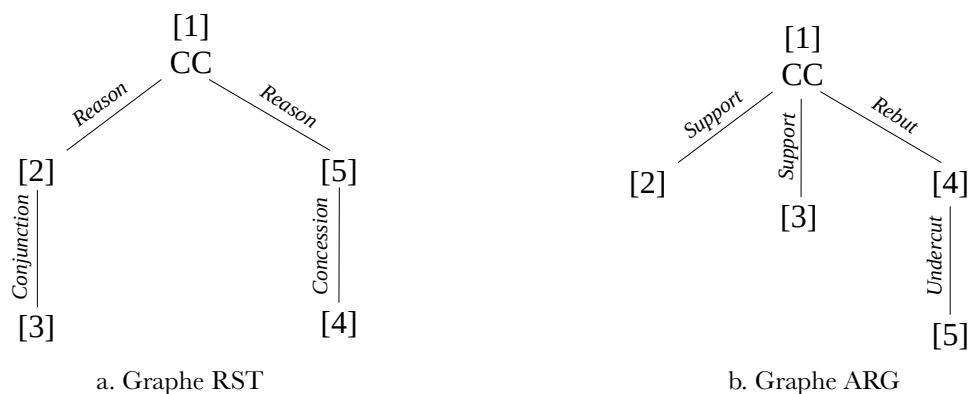


Figure 17. Compatibilité des graphes RST et ARG de l'exemple (53)

<sup>130</sup> Les figures sont reprises à Huber et al. (2019a [30], p. 85 et 2019c [32], p. 36) avec quelques modifications et compléments.

La première transformation a consisté en l’association du segment noyau principal dans le graphe représentant la structure discursive et de la thèse dans le graphe représentant la structure argumentative (renommés CC pour *Central Claim*). Ceci soulève le problème de partir d’une racine commune pour construire l’arbre. Dans ce corpus, Peldszus & Stede ont remarqué qu’il y avait une correspondance du segment noyau principal et de la thèse dans 85 % des cas, ce qui est un très bon score (2016, p. 106). Cependant, du point de vue linguistique, les 15 % restants posent question et il serait intéressant de sonder l’aspect plus ou moins aléatoire de cette non-correspondance.

Deuxièmement, on a modifié la portée de la relation argumentative *Undercut*. Dans la théorie de l’argumentation convoquée, comme cela a été dit plus haut, la relation *Undercut* est une relation d’attaque portant sur la relation *Support* entre un segment et la thèse (voir le graphe initial repris ici dans l’Annexe 3b). Dans le graphe ARG « arrangé » (Figure 17b), la relation *Undercut* porte sur le nœud [4] et pas sur l’arc entre [4] et [1]. On comprend bien la démarche qui a pour but de faciliter la comparaison et les redescriptions – on a suivi en cela ce qui est préconisé par d’autres (Wachmuth et al. 2017) – mais, du point de vue sémantique, ce n’est plus aussi précis et cohérent. Pour dire les choses simplement, en cherchant à régler un problème à un niveau, celui de la représentation pour l’automatisation, on en crée un au niveau de l’interprétation des données. Il me semble que c’est, sinon un écueil, au moins une tension fréquente en TAL, entre contraintes cognitives (interprétabilité humaine) et contraintes méthodologiques.

### (ii) *Extraction de sous-graphes fréquents*

La deuxième étape a consisté à extraire, pour chaque ensemble de graphes RST et ARG « arrangés »<sup>131</sup>, les sous-graphes les plus fréquents, sans les mettre en correspondance. L’algorithme utilisé (*gSpan*<sup>132</sup>) a permis de récupérer tous les sous-graphes RST et ARG apparaissant dans au moins 2 des 112 textes. Ainsi, on a obtenu 311 sous-graphes de type RST et 98 sous-graphes ARG, que nous avons numérotés. Il n’est pas possible de reproduire la matrice complète, car elle serait trop grande et illisible. Elle permet d’attribuer à chaque texte un certain nombre d’attributs  $r$ , désignant les sous-graphes parmi les 311 sous-graphes RST préalablement obtenus qui sont présents dans ce texte, et un certain nombre d’attributs  $a$ , désignant les sous-graphes parmi les 98 sous-graphes ARG préalablement obtenus qui sont présents dans ce texte. Par exemple, dire que le texte 46 a les attributs  $r27, r85, r211$  et les attributs  $a34, a66$  veut dire que dans ce texte, on trouve les sous-graphes RST numérotés 27, 85 et 211 dans la liste des 311 sous-graphes RST fréquents et les sous-graphes ARG numérotés 34 et 66 dans la liste des 98 sous-graphes ARG fréquents.

### (iii) *La fouille de redescriptions*

Nous avons ensuite cherché à savoir s’il était possible de faire correspondre des sous-graphes RST et des sous-graphes ARG. C’est la méthode de la fouille de redescriptions

131 Soit 112 graphes RST sur le modèle de la Figure 17a et 112 graphes ARG sur le modèle de la Figure 17b.

132 *gSpan* = *Graph-Based Substructure Pattern Mining*. Il s’agit d’un algorithme capable d’extraire les sous-graphes les plus fréquents dans un ensemble de graphes. Le seuil de fréquence a été fixé pour nous à 2, c’est-à-dire qu’un sous-graphe a été jugé fréquent s’il apparaissait au moins dans deux textes (voir Yan & Han 2002).

qui a été utilisée, je ne la détaille pas<sup>133</sup>. Elle consiste à trouver deux caractérisations différentes d'un même ensemble d'objets. Dans notre cas, les objets sont les textes et les deux caractérisations différentes sont les sous-graphes ARG et sous-graphes RST que les textes ont simultanément en commun. En guise d'illustration, je reprends les données et les commentaires de la redescription ayant obtenu le meilleur score c'est-à-dire regroupant le plus de textes. Il s'agit de la redescription 1 correspondant ici à la Figure 18 et au Tableau 7.

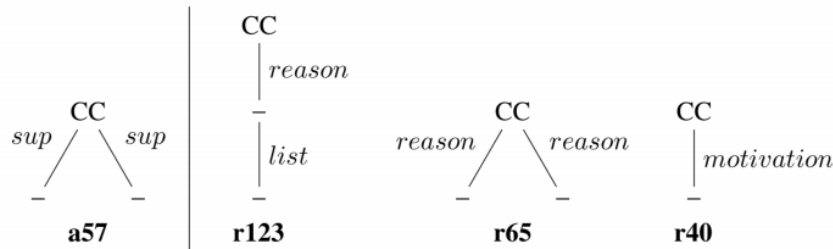


Figure 18. *Arg-microtexts* : sous-graphes de la redescription 1

	q1	q2	J(q1,q2)	nbre de textes
Redescription 1	a57	r40 V r65 V r123	0.691	54

Tableau 7. *Arg-microtexts* : propriétés de la redescription 1

Concrètement, cela revient à dire que dans 54 textes sur les 112 du corpus, le sous-graphe argumentatif a57, qui montre deux relations Support entre deux segments (représentés par -) et la thèse (CC), est associé à trois sous-graphes discursifs : soit le sous-graphe r123 où les deux segments s'enchaînent, soit le sous-graphe r65 où chaque segment est en relation de Justification avec la thèse, soit le sous-graphe r40 où un seul segment est en relation de Motivation avec la thèse. Cette première approche aurait mérité d'être approfondie, en particulier du point de vue de l'analyse des résultats. Au niveau linguistique, un retour systématique aux données permettrait, peut-être, de montrer la prédisposition de certaines formes ou formules pour marquer certaines configurations. Et, inversement, il aurait été intéressant de creuser les impossibilités d'équivalence entre des configurations argumentatives et discursives.

#### • L'utilisation des plongements

Dans une perspective différente, nous avons exploré d'autres méthodes de fouille de données textuelles, qui, bien que ne ciblant pas ouvertement la structure argumentative des textes, peuvent y contribuer. Dans Huber et al. (2020d [36])<sup>134</sup>, nous avons cherché à tester s'il était possible de capturer des propriétés discursives potentiellement argumentatives au moyen de modèles dits de plongements. Il s'agit d'une méthodologie

<sup>133</sup> Elle repose sur l'utilisation de l'algorithme ReReMi et sur le calcul de l'indice de Jaccard, qui est un indice de similarité. Rapporté à notre problématique, plus l'indice est élevé, plus il y a de textes qui partagent les mêmes sous-graphes RST et les mêmes sous-graphes ARG. (voir Galbrun & Miettinen 2012 et 2017).

<sup>134</sup> Ce travail a été effectué dans le cadre de la formation de Master 2 de Chaker Memmadi.

représentant le sens des mots sous la forme de vecteurs numériques de grande dimension. Notre démarche consistait à comparer quatre modèles de plongements, deux non contextuels (GloVe et Google USE) et deux contextuels (BERT et SciBERT) pour montrer lesquels prenaient le mieux en compte les propriétés discursives qui nous intéressaient<sup>135</sup>. Nous avons voulu tester si les modèles étaient sensibles à la cohérence entre deux propositions à l'intérieur d'une même phrase et à la valeur sémantique de la relation de cohérence associée. Pour cela, nous avons dû sélectionner des données pertinentes, et nous avons construit quatre expériences différentes (deux pour la cohérence entre propositions, et deux pour la valeur sémantique de la relation entre propositions). Je ne reprends pas la description des tâches ni le détail des résultats, car la cible est plus informatique que linguistique (il s'agit de tester des modèles) et car les résultats se sont finalement avérés peu satisfaisants. Mais les réflexions que la collaboration a suscitées ne sont pas pour autant dénuées d'intérêt du point de vue linguistique. J'en présente quelques-unes sous forme de remarques concernant (i) les données et l'unité de base, (ii) la cohérence et la structure informationnelle, (iii) le « poids » des relations de discours.

**(i) Les données et l'unité de base**

Pour être « testables » au niveau de leur propriétés discursives, les données devaient être segmentées, ponctuées et comporter une annotation sémantique des relations de discours. Nous avons utilisé un corpus existant et annoté, le corpus SciDTB<sup>136</sup>, composé de 798 résumés d'articles scientifiques issus de l'anthologie de l'*Association for Computational Linguistics* (ACL) (voir Yang & Li 2018 pour un descriptif). Il s'agit de données relativement calibrées et homogènes du point de vue de la taille et du genre de discours. Le genre « résumé scientifique » relève de la communication académique et renvoie à des conventions propres à une communauté. Un résumé scientifique comporte notamment des parties identifiables, une perspective à la fois descriptive et argumentative et parfois des formules attendues (voir Bhatia 2013, Swales 2013 etc. et ici même la section 1.1.3 sur le genre, p. 25). L'annotation discursive a été faite sur la base théorique de la RST, mais dans une perspective de dépendance. Les labels des relations sont similaires mais la représentation est différente des graphes présentés précédemment (voir Li et al. 2014). L'unité qui nous a servi de référence pour étudier les propriétés discursives est la phrase (*sentence*), comprise basiquement comme un segment textuel entre deux points<sup>137</sup>.

La plupart des 798 documents comportent entre 5 et 7 phrases, chaque phrase étant composée d'une ou plusieurs propositions (*clause*). L'ensemble de base est composé de 4196 phrases, dont 787 qui ne comportaient qu'une proposition et qui ont donc été exclues de notre analyse. La Figure 19 représente la distribution des relations de discours observables entre propositions à l'intérieur d'une phrase.

135 BERT = Bidirectional Encoder Representations from Transformers (voir Devlin et al. 2019) ; Word2Vec = WORD To VECTOR, GloVe = GLOBAL VECTOR (for Word Representations). Voir la plateforme [huggingface](https://huggingface.co), qui contient un certain nombre de modèles et qui permet de les exploiter en partie. Le détail des calculs et la comparaison précise des modèles sont hors de mon champ de compétence. Je renvoie à la présentation générale et accessible de Pilehvar & Camacho-Collados (2021).

136 SciDTB = Discourse Dependency Treebank for Scientific Abstracts.

137 Ce qui situe nos expériences dans le champ des plongements de phrases et non des plongements de mots (*sentence embedding*, voir Pilehvar & Camacho-Collados 2021, p. 97-102).

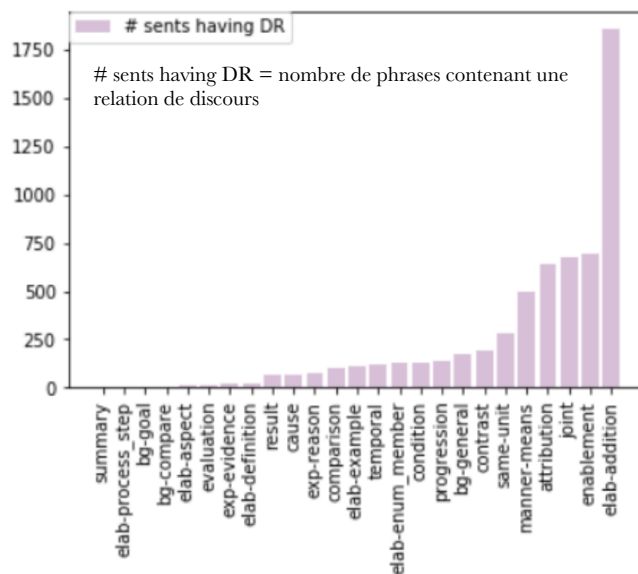


Figure 19. Distribution des relations de discours<sup>138</sup> entre propositions dans Sci-DTB (repris à Huber et al. 2020d [36], p. 4)

L'unité phrase comme unité maximale n'est pas sans poser problème, pour plusieurs raisons : le critère est essentiellement graphique, ce qui est insuffisant sur le plan sémantique et même parfois sur le plan syntaxique<sup>139</sup> ; pour étudier les relations de discours, elle oblige à sélectionner des phrases nécessairement complexes, et dans certains cas nécessairement binaires (ne contenant que deux propositions). Du point de vue textuel, cela est limité, car des éléments de structure passent « sous les radars », notamment des relations entre phrases, la constitution de blocs plus complexes et leur organisation hiérarchique, autrement dit tout ce qui fait qu'un texte n'est pas qu'une suite linéaire de phrases. Dans une telle approche limitée à l'analyse des relations de discours à l'intérieur de la phrase graphique, ce qui pose problème du point de vue linguistique, c'est la réduction systématique de la portée des relations.

**(ii) La cohérence et la structure informationnelle**

Afin de tester les modèles de plongement du point de vue de leur sensibilité à des propriétés discursives, nous avons construits quatre tâches. Deux d'entre elles sont censées tester si les modèles prennent en compte la cohérence entre les propositions à l'intérieur d'une phrase (Huber et al. 2020d [36], p. 2-3). La manipulation des données pour tester cette hypothèse nous a conduits à comparer des phrases telles qu'elles apparaissaient dans le corpus et des phrases construites à partir du corpus, mais dont la cohérence a été modifiée, soit en changeant l'ordre des propositions (*swapped units*

<sup>138</sup> Il existe deux niveaux de catégorisation/labellisation. Un premier niveau comporte 17 labels, et un niveau plus fin comporte 25 labels. Cela est dû à la prise en compte ou non des sous-catégories de la relation Élaboration (ex. Éla-addition, Élab-definition, etc.), Arrière-plan (ex. Bg-compare, Bg-general), etc.

<sup>139</sup> Je pense à des structures comme les ajouts après le point, qui ne sont pas si rares que cela, y compris à l'écrit, et peuvent servir l'argumentation. Par exemple : *Nous serons toutes des oies blanches, des blanches colombes, voire des Blanche-Neige à qui on ferait avaler n'importe qu'elle pomme empoisonnée. Même si le blanc est, en Asie et en Afrique, associé à la mort et au deuil.* (NO 02 03 06, dans Combettes 2007, ex. n° 38). Mais il est vrai que ce type de construction vise souvent des effets énonciatifs auquel le genre « résumé scientifique » du corpus SciDTB ne fait pas vraiment appel.

*detection*)<sup>140</sup> soit en mélangeant des propositions extraites de phrases différentes (*scrambled sentence detection*)<sup>141</sup>.

Dans le premier cas, au niveau linguistique, cela correspond intuitivement au fait que l'ordre des propositions dans une phrase complexe contribue à la cohérence et que le modifier va perturber ou mettre à mal cette cohérence, pour des raisons temporelles, de résolution d'anaphores, etc. Ceci est particulièrement vrai pour les juxtapositions, mais ne semble pas généralisable. En effet, passer d'une phrase de type AB à une phrase de type BA ne produit pas forcément de l'incohérence. D'une part, dans le cas où la relation de discours est marquée par certains connecteurs, l'inversion peut jouer à la marge sur la cohérence. Par exemple, une phrase de type *siA B* (ex. *Si Marie vient, Paul sera content*) sera équivalente à sa correspondante *B siA* (ex. *Paul sera content, si Marie vient*). D'autre part, même si l'inversion des deux propositions dans la phrase change l'interprétation, elle ne produit pas forcément de l'incohérence, mais crée une autre cohérence. Par exemple, une phrase de type AB comme *Marie va au restaurant, le serveur est charmant* est interprétable sur la base d'une relation de cause, mais cette cause peut être A ou B (« A parce que B » ou « Comme A, alors B »). Le fait de modifier l'ordre des propositions (*Le serveur est charmant, Marie va au restaurant*) dans la phrase ne produira pas d'incohérence, mais fera privilégier, peut-être, une interprétation plutôt qu'une autre (« Comme A, alors B » ou « A donc B »). L'association de la perte ou de la modification de la cohérence par l'inversion des propositions dans une phrase complexe est certainement une hypothèse trop forte, ou trop grossière, et la manipulation automatique des données ne permet pas, à ce stade, de l'affiner.

La deuxième tâche imaginée repose sur l'association de la cohérence à la structure informationnelle de la phrase, en particulier la présence d'un thème commun entre les deux propositions qui composent cette phrase. La manipulation des données a consisté à comparer des phrases authentiques de type AB extraites d'un texte du corpus et des phrases « bricolées » de type AF où A est une proposition de la phrase d'origine et F une proposition extraite aléatoirement d'un autre texte du corpus. L'hypothèse est qu'en procédant ainsi on devrait provoquer une rupture thématique qui devrait faire des phrases bricolées de type AF des phrases incohérentes. Là aussi, si l'incohérence recherchée est souvent au rendez-vous, elle n'est pas totalement garantie ou pas forcément pour les raisons souhaitées. D'une part, la difficulté d'interprétation peut résulter de problèmes d'ordre syntaxique et pas thématique (par exemple A annonce une coordination et F a la forme d'une subordonnée). D'autre part, le corpus étant relativement homogène et correspondant à un domaine scientifique ciblé (les résumés scientifiques d'une conférence en linguistique informatique), certains textes auront des thèmes proches et potentiellement des propositions interchangeable sans trop de dommages pour la cohérence<sup>142</sup>.

Plus encore, dans les cas où A et F ne présenteraient pas de problèmes de jonction syntaxique, mais pas non plus de points communs thématiques, cela ne voudrait pas dire qu'un esprit humain ne reconstruirait pas de cohérence. Il est même assez certain qu'il essaierait d'en tirer quelque chose, au prix d'un effort plus ou moins important,

140 Par exemple A-B devient B-A.

141 Par exemple A-B (résumé 25) → A-F (A étant une proposition extraite d'une phrase d'un autre document).

142 D'autant plus que la rhétorique des résumés scientifiques est une rhétorique assez conventionnelle et bien rodée qui peut produire des formules passe-partout.



variable selon les individus, qui rendrait la phrase plus ou moins pertinente dans son contexte, mais pas complètement inacceptable (selon le principe bien connu du jeu dit du *cadavre exquis*).

(iii) *Le « poids » des relations*

Un autre point qu'ont soulevé pour moi les expériences menées concerne les types de relation de discours. Deux autres tâches ont été construites pour tester la sensibilité des modèles de plongement à la présence et à la sémantique de relations de discours entre les deux propositions d'une phrase complexe (tâches 3.3 et 3.4, dans Huber et al. 2020d [36], p. 3). La remarque suivante ne porte que sur la tâche de détection de la présence d'une relation (3.3. *relation detection*) et sur les contraintes de l'analyse, qui sont autant de limites pour l'application des résultats. Nous avons donc fait l'hypothèse que les modèles prendraient en compte la présence d'une relation de discours entre deux propositions d'une même phrase. Pour cela, nous avons été obligés d'opérer une sélection parmi les 4196 phrases du corpus. Les phrases ne contenant qu'une proposition ont assez logiquement été exclues et, parmi les 3409 qui restaient, nous avons fait une distinction entre relations de type additif ou continuatif et relations logiques pour ne travailler que sur les relations présentées comme logiques, soit 1447 phrases seulement. Parmi les relations présentes dans le corpus et quantifiées dans la Figure 19 ci-dessus, les relations que nous avons qualifiées de logiques correspondent aux relations : Attribution, Background, Cause, Condition, Contrast, Comparison, Enablement, Evaluation, Explanation (Exp-evidence et Exp-raison), Manner-mean, Result, Summary, Temporal. La liste des relations qualifiées d'additives ou continuatives est moins fournie, elle comporte les relations Elaboration (Elab-definition, Elab-example, etc.), Progression, Joint et Same-unit.

Cette distinction pose question pour plusieurs raisons et n'est en cela pas dénuée d'intérêt sur le plan linguistique. De mon point de vue, elle est notamment problématique sur le plan sémantique et sur le plan quantitatif. Du point de vue sémantique, la distinction proposée n'est pas vraiment expliquée et a suscité quelques discussions. On infère que les relations dites logiques « pèsent » plus lourd relativement au but des textes (présenter et justifier les idées défendues dans l'article complet). Il faut peut-être envisager cette liste comme une grille de lecture liée à un genre particulier et au relief argumentatif qu'on lui reconnaît. Mais, l'argumentation se nourrit de définitions et d'exemples, qui correspondent aussi à des sous-catégories de la relation Elaboration, considérée comme additive et continuative et donc non traitée ici. Sur le plan quantitatif, la Figure 19 expose assez clairement le problème : les relations dites logiques ne constituent pas la majorité des cas. Là encore, c'est la relation Elaboration (toutes sous-catégories confondues) qui est de loin la plus représentée. Ne pas la prendre compte oblige à exclure beaucoup de données et fait passer l'analyse à côté d'un fait que l'on peut supposer pertinent, à savoir la place de l'Elaboration dans le profil générique des résumés scientifiques. Je ne veux pas dire que la présence de la relation Elaboration dans le corpus Sci-DTB est symptomatique du genre résumé scientifique (on sait par ailleurs qu'elle fait partie des relations les plus fréquentes en général), mais certains paramètres pourraient être spécifiques, par exemple la proportion de sa présence, la répartition de sous-catégories et leur position dans les textes par rapport à des textes relevant d'autres genres.

Pour résumer, l'intérêt de ces expériences a été certes de me familiariser avec des méthodes expérimentales utiles, mais surtout, à ce stade, de soulever des problèmes linguistiques, théoriques et pratiques à l'interface entre le TAL et l'analyse de la cohérence textuelle. Dans les deux cas, fouille de redescriptions et méthode des plongements, un retour à la « matière première » que constituent les énoncés concrets paraît nécessaire pour interroger et affiner les distinctions et les résultats qu'une approche purement automatique oblige à faire ou rend.

## Synthèse 2

Dans cette partie, j'ai exposé des recherches plus diversifiées que dans la première partie et ayant comme point commun l'organisation du discours à travers des constructions particulières, les juxtapositions, et des schémas logico-argumentatifs.

L'analyse des juxtapositions montre l'utilité pratique d'un cadre comme celui des CxG, qui permet de penser et de représenter plusieurs dimensions descriptives (en particulier prosodique, syntaxique, sémantique et pragmatique) et de révéler le caractère plus ou moins compositionnel de la structure d'ensemble (voir le traitement des juxtapositions conditionnelles). De plus, la souplesse des CxG permet de dépasser le cadre des unités syntaxiques traditionnelles (syntagme, phrase, relations de dépendances argumentales ou adjonctives) et de déployer des structures hiérarchiques d'héritage entre constructions d'une même famille (constructions juxtaposées propositionnelles, constructions juxtaposées propositionnelles conditionnelles, constructions juxtaposées propositionnelles conditionnelles à l'impératif). Ces hiérarchies sont révisables en fonction des constructions décrites et des dimensions pertinentes. Par exemple, les expériences menées sur la prosodie des contours dits continuatifs ont montré que la valeur initialement déclarée pour le premier constituant des juxtapositions (contour final montant) n'était pas vérifiée chez tous les locuteurs.

La perspective argumentative a soulevé plusieurs questions. D'abord celle de la différenciation des structures discursive et argumentative représentables sous forme de graphes. Leur comparaison a nécessité quelques aménagements qui ont du même coup souligné ces différences (nature des relations de type RST ou ARG, identification d'une thèse (*claim*) pour l'argumentation, etc.). Par ailleurs, les travaux menés en collaboration dans le domaine du TAL ont été l'occasion pour moi d'aborder de nouveaux genres de discours (réponses à des questions polémiques et articles scientifiques) et d'interroger les données sans être guidée par une perception qualitative globale des textes. Il y avait donc une part de recherche « en aveugle », à partir de données déjà enrichies par annotation, mais dont la manipulation et la comparaison se sont faites non pas sur le matériau linguistique lui-même mais sur sa représentation graphique ou vectorielle.

Je retiens de ces deux approches des connexions sur la recherche d'explication de la cohérence et sur l'utilité de sa représentation, mais une différence majeure, qui est, pour la deuxième – fouille de textes et d'arguments – un usage limité de l'intuition du sens à travers les annotations.



# 3

## Marqueurs discursifs

*... ces mots, images des mouvements intérieurs, se présentent comme des accompagnements, ou s'il m'est permis de me servir de cette figure, comme des assaisonnements du discours, par le moyen desquels on ajoute à la peinture de la pensée celle de la situation, soit de l'âme qui sent, soit de l'esprit qui peint.*

Abbé Gabriel Girard, *Les vrais principes de la Langue Française*, vol. 2, Le Breton. Chapitre XIII Discours, Des particules, p. 313. [Gallica]



Cette troisième partie est une synthèse des travaux qui ont en commun de s'intéresser aux ressources lexicales marquant l'organisation du discours. L'idée de marquage renvoie autant à une fonction qu'à une trace. Les phénomènes lexicaux étudiés, que j'appelle à la suite de beaucoup d'autres *marqueurs discursifs*, permettent tout autant la structuration cognitive et « physique » du flux discursif qu'ils en sont les témoins. Ils sont « une certaine façon d'impliquer l'énonciation dans le sens » (Ducrot et al. 1980, p. 34) et c'est en cela que je les trouve intéressants.

Ma réflexion dans le domaine a commencé assez tôt puisque je les ai utilisés comme des indices de réalisme langagier des personnages dans les pièces de Michel Tremblay. Certains sont effectivement associés à la représentation de l'OPQ (par ex. *ça fait que (fak), câlisse, coudonc, maudit*, etc.) (Dagnat 2006 [6], p. 251-298). Mais leur description (prosodique, sémantique et pragmatique) systématique n'a débuté qu'à partir de 2012 avec le projet PARDI<sup>143</sup>, que j'ai initié et coordonné et qui m'a permis de développer plusieurs collaborations. Aujourd'hui, à travers quelques publications récentes (par ex. Dagnat 2022a [42]) et un autre projet collaboratif plus ambitieux que je coordonne (le projet CODIM<sup>144</sup>), je cherche à expliciter et à représenter les propriétés des combinaisons de marqueurs.

Cette partie correspond donc à une troisième étape dans mon parcours, mais une étape qui chevauche largement la seconde, tant du point de vue temporel que des objets étudiés, des arrière-plans théoriques et des méthodes utilisées.

La rédaction des sections ci-dessous a été l'occasion pour moi de clarifier des aspects terminologiques et d'approfondir des points théoriques que je continue de trouver complexes. Cela m'a aussi permis d'explorer davantage certains points traités rapidement dans *Grande Grammaire du Français* (Dagnat 2021a [39], p. 1569-1572) et de fournir une première version de la notice de l'*Encyclopédie grammaticale* qui m'a été demandée sur les particules énonciatives<sup>145</sup>. Il s'agit donc à la fois d'un apport nouveau et d'un rappel de mes travaux dans le domaine.

La section 3.1 expose et discute quelques caractéristiques fréquemment mentionnées dans la littérature. J'insiste sur l'indexicalité, le sens procédural, la portée et le niveau de signification. La section 3.2 est consacrée à des réflexions et des travaux plus ciblés sur les connecteurs et les particules énonciatives, avec une ouverture sur des considérations diachroniques et quantitatives.

#### Références des travaux concernés

2006 [6], 2014a [21], 2015b [24]\*, 2016a [25]\*, 2017a [27], 2017b [28]\*,  
2018 [29], 2019b [31], 2020a [33]\*, 2020b [34], 2020c [35], 2020e [37], 2020f [38]\*,  
2021a [39]\*, 2021b [40]\*, 2021c [41]\*, 2022a [42]\*, 2022b [43]\*

143 Projet PARDI : Sens et corrélats prosodiques des particules discursives. Projet financé par l'ATILF et par l'Université de Lorraine : 2014-2016.

144 Projet CODIM : COmpositionality of DIscourse Markers. Projet financé par l'ANR : 2023-2027. Pour un bref descriptif, voir la section 4.1.1.

145 Cette notice ne figure pas dans le document III, car elle n'était pas terminée au moment du dépôt de la présente synthèse. Je dois prendre en compte un certain nombre de remarques suite à l'envoi d'une première version, et compléter quelques sections encore lacunaires. Elle paraîtra en 2023 sur le site de l'*Encyclopédie Grammaticale*. J'y fais cependant référence, car elle reprend des travaux déjà publiés et des éléments développés ici.

## 3.1 Que marquent les marqueurs discursifs ?

*Marqueurs discursifs* (désormais MD) est une étiquette largement répandue pour désigner des items et des usages divers (section 3.1.1). En suivant Hansen (1997, p. 162), je les envisage comme des items lexicaux utilisés pour donner des instructions d'interprétation dans un contexte donné. Ce caractère instructionnel a au moins deux conséquences : leur mode de dépendance au contexte fait d'eux des expressions indexicales (section 3.1.2) et la signification des MD est en grande partie procédurale (section 3.1.3). Leur portée (section 3.1.4) et leur mode d'intégration sémantique (section 3.1.5) sont aussi des questions récurrentes.

### 3.1.1 Des repères dans la jungle

Le domaine des MD est très documenté et il n'est pas toujours facile d'être clairvoyant dans cette « jungle » (Fisher 2006, p. 1). Je pointe quelques éléments définitoires et terminologiques saillants concernant la multiplicité des termes et les propriétés générales.

#### 3.1.1.1 Multiplicité des approches et des étiquettes

Les différences terminologiques sont un écueil car elles ne se laissent pas réduire à un simple problème d'école ou de tradition. Les approches et les étiquettes ne sont pas toutes superposables ni même articulables. Un rapide sondage des publications sur le sujet fait apparaître la multiplicité et la multiplication des approches depuis les années 1970, avec une accélération les trente dernières années qui peut produire un effet de « saturation » (Dostie 2004, p. 15)<sup>146</sup>. Le corollaire de cette diffusion est d'abord l'explosion des termes utilisés. En parcourant les différents travaux, on recense facilement une vingtaine de termes en français. Dér (2010, p. 5) en relève jusqu'à 42 en anglais. Le tableau suivant, bien qu'incomplet, donne une idée de l'empan terminologique autour des trois labels principaux que sont *marqueurs*, *particules* et *connecteurs*<sup>147</sup>.

<b>Marqueurs</b>	<b>discursifs</b>	Aijmer 2002, Blakemore 2002, Chanet 2003, Dostie 2004, Fraser 1999, 2009, Hansen 1998a, Kahane & Mazziotta 2015, Schiffrin 1987, Shourup 2001, Rodriguez Somolinos 2011, Traugott 2007, Waltereit 2001, 2007, Winterstein 2010
------------------	-------------------	--

<sup>146</sup> On trouve par exemple les formules suivantes : « rapidly expanding body of research » (Shourup 1999, p. 227), « domaine prolifique qui a alimenté maints débats », « surenchère » ou « profusion de travaux » (Dostie 2004, p. 15), « the jungle of publications and different approaches » (Fischer 2006, p. 1), « a quite large body of literature on discourse particles » (Zimmermann 2011, p. 2013), « great deal of attention » (Aijmer & Simon-Vandenberg 2011, p. 223), « a considerable body of recent research » (Heine 2013, p. 1205), « Dans la jungle des mots du discours » (Vázquez Molina 2016, p. 1), « a growth industry » (Brinton 2017, p. 1).

<sup>147</sup> Je me concentre sur les publications en anglais et en français portant sur l'anglais et le français et occasionnellement sur l'allemand. Je ne reprends ici que les termes les plus fréquents. Pour un tour d'horizon plus complet, voir en particulier les synthèses de Aijmer (2013, p. 1-19), Brinton (2017, p. 1-13), Dostie (2004, p. 40-41), Fischer (2006, p. 1-20), Foolen (1996, p. 4-6), Hansen (1998a, p. 37-85), Métrich et al. (1998, p. 10-36) et Shourup (1999, p. 227-242).

	<b>pragmatiques</b>	Aijmer 2013, Aijmer & Simon-Vandenberg 2011, Beeching 2016, Brinton 1996, 2017, Dostie 2004, Fraser 1999, 2009
	<b>de la structuration de la conversation</b>	Auchlin 1981
<b>Particules</b>	<b>discursives</b>	Abraham 1991, Aijmer 2002, Fischer & Drescher 1996, Fischer 2000, 2006, Hansen 1998a, Wilson & Sperber 1993, Stede & Schmitz 2000, Vincent 1993
	<b>de la modulation</b>	Koch & Österreicher 2001
	<b>énonciatives</b>	Fernandez-Vest 1994, Paillard 2011
	<b>pragmatiques</b>	Abraham 1991, Beeching 2002, Foolen 1997, Östman 1981, 1995
	<b>métacommunicatives</b>	Vincent 1993
	<b>connectives</b>	Gülich 1970, Métrich et al. 1998-2002
	<b>modales</b>	Detges & Waltereit 2016, Diewald 2013, Hansen 1998a, Métrich et al. 1998-2002, Traugott 2007, Waltereit 2001
<b>Connecteurs</b>	<b>textuels</b>	Dostie 2004, Riegel et al. 2009
	<b>de discours</b>	Wilson & Sperber 1993, 2012, Blakemore 1987, 1992 ( <i>discourse connectives</i> )
	<b>pragmatiques</b>	Berrendonner 1983, Roulet 1987, Roulet et al. 2001, Ducrot & Anscombe 1997 [1983], Vázquez Molina 2016
	<b>argumentatifs</b>	Ducrot 1983, Anscombe & Ducrot 1997 [1983]
	<b>phatiques</b>	Davoine 1980, Bazzanella 1990
	<b>adverbes</b>	Métrich et al. 1998-2002 ( <i>adverbes connecteurs</i> )

Tableau 8. Liste indicative d'étiquettes désignant des MD

Une deuxième conséquence de la multiplication des travaux est la difficulté de comparer les catégorisations<sup>148</sup>, ce que souligne Diewald en plaçant pour une catégorisation flexible et l'identification de MD prototypiques (2013, p. 41). Les différentes propositions que l'on trouve dans la littérature ne mobilisent en fait pas forcément les mêmes dimensions parce qu'elles visent des ensembles de mots et d'expressions distincts. De plus, elles demeurent parfois assez vagues sur des points cruciaux. Bien que discutée, la liste des propriétés proposée par Brinton (1996, 2017, reprise ici dans l'Annexe 4, p. 197) constitue un bon point de départ. Je m'arrête sur l'emploi des termes *pragmatique* (i), *non-compositionnel* (i) et *procédural* (iii).

(i) On comprend intuitivement que des auteurs comme Brinton ou Heine (2013), en utilisant le terme *pragmatique*, ciblent essentiellement des mots ou des expressions étiquetés comme interjections (*Hwoet* en vieil anglais), tournures exclamatives (*if only*<sup>149</sup>), parenthétiques (*I think*), etc. et pas les items habituellement

148 Voir les parties définitives dans Aijmer & Simon-Vandenberg (2011), Andersen & Fretheim (2000), Beeching (2016), Crible & Degand (2019), Dér (2010), Denturck (2008), Diewald (2013), Fraser (2009), Ghezzi & Molinelli (2014), Ghezzi (2014), Heine (2013), Jucker & Ziv (1998), Kahane & Mazzotta (2015), Lopez Villegas (2019), Petit (2010), Stede & Schmitz (2000), Traugott (2007), etc.

149 Par exemple dans « Most people were surprised that it wasn't the creepy guy in the raincoat who went online trolling for young girls. If only. In real life it was the teacher, the doctor, the grandpa, the restaurant manager. » L. Wiehl (*Face of betrayal*, 2008) cité par Brinton (2017, ex.



classés comme connecteurs (*pourant, parce que, donc, en fait, en effet, etc.*)

En adoptant une perspective « œcuménique », j’entends ici par *pragmatique* le fait que la communication linguistique met en jeu des ressources complémentaires à celles du code linguistique, à savoir le recours au contexte et le recours à des stratégies d’interprétation<sup>150</sup>. Sous cet angle, il est difficile de comprendre en quoi des marqueurs comme *donc* ou *en fait* seraient des marqueurs moins pragmatiques que *bon* ou *zut*. *Donc* et *en fait* sont peut-être moins expressifs mais pas moins pragmatiques. On a besoin dans tous les cas de convoquer des éléments saillants du contexte pour construire une interprétation et d’appliquer des stratégies qui limitent cette dernière pour qu’elle soit pertinente<sup>151</sup>. Il n’est pas évident dans ce cadre de construire une échelle de « pragmatisme », sinon en faisant appel à d’autres aspects, en particulier au *type* des objets, ou domaines, mis en jeu dans l’interprétation. C’est-à-dire la portée de *donc, en fait, bon* ou *zut*, etc.

(ii) Dans sa liste, Brinton attribue aux MD un sens essentiellement procédural et ajoute qu’ils sont non compositionnels (2017, p. 9). Concernant la non-compositionnalité, il faut préciser à quel niveau la notion est utilisée. On distinguera un niveau *interne* et un niveau *externe*<sup>152</sup>.

Généralement, c’est au niveau *interne* que les MD multi-mots sont dits non compositionnels, dans le sens où leur signification comme MD ne semble pas complètement dérivable de la signification de chacun de leurs composants. En synchronie, on peut penser à des associations comme *ah bon, quand même, mais enfin, écoute donc*, etc. La gestion de la (non-)compositionnalité paraît plus complexe en diachronie, lorsqu’il s’agit par exemple d’expliquer l’évolution d’expressions comme *enfin, en effet, du coup, à la rigueur, en même temps*, etc.

On peut tenter d’étendre la question de la (non-)compositionnalité au niveau *externe*. On cible alors le mode de combinaison du MD avec le reste de l’énoncé. Ceci peut faire penser à la différence *restrictif* vs *non restrictif* proposée par Heine (2013). Il est généralement avancé que les MD ne tire pas leur fonction de la structure phrastique de leur énoncé-hôte ni n’interviennent dans son contenu propositionnel<sup>153</sup> (c’est-à-dire qu’ils ne le modifient pas). Ils sont alors dits non restrictifs<sup>154</sup>. Par exemple, dans

24d, p. 118).

150 Par exemple les maximes de Grice (1975, p. 45-47) ou de Horn (1985, 1989, p. 193 sq.), ou le principe de « pertinence optimale » de Sperber et Wilson (1995 [1986], p. 158).

151 Je pense précisément à ce que l’on range traditionnellement dans la maxime de pertinence (ou de relation) de Grice (1975, p. 46).

152 Je remercie Maj-Britt Mosegaard Hansen d’avoir attiré mon attention sur ce passage devenu flou sur ce point dans la première version du document, suite à une réduction un peu maladroite du paragraphe.

153 J’adopte ici l’approche « classique » des actes de langage, où l’on distingue comme chez Stenius (1967) le *radical* (propositionnel) et la modalité (*mood*), ou comme chez Searle (1969) le *contenu propositionnel* et la *force illocutoire*. Dans les cas les plus simples, un contenu propositionnel *p* qui décrit un état de choses possible se voit doté d’une force illocutoire. Cette force peut être par exemple l’assertion ou la question, qui vont spécifier que, pour le locuteur, cet état de choses existe (assertion) ou, dans une question, que le locuteur requiert une information à son propos, par exemple sur son existence (*est-ce que p ?*) ou ses causes (*pourquoi p ?*).

154 « [...] non-restrictive meaning is not anchored in the structure of a sentence or its constituents but rather concerns reasoning processes and inferential mechanisms that are grounded in the situation of discourse ». (Heine 2013, p. 1215).

(54a), l'interjection *zut* ne se combine pas à ce qui dans le reste de l'énoncé renvoie à la proposition *j'ai oublié ma clé*. Le test souvent utilisé est celui du clivage, sachant que seuls sont clivables les éléments appartenant au contenu propositionnel. Ceci explique pourquoi (54b) ne fonctionne pas.

- (54) a. Zut, j'ai oublié ma clé.  
b. ? C'est zut que j'ai oublié ma clé.

En fait, à y regarder d'un peu plus près, l'absence de contribution au contenu propositionnel, le caractère non restrictif des MD, en plus d'être parfois discutable, ne semble pas constituer pas un trait distinctif des MD. Elle semble valoir pour beaucoup de déclencheurs de présupposition (pour un développement, voir plus loin section 3.1.5, p. 114). Ajoutons à cela que, si les MD ne se composent pas avec le contenu propositionnel de leur énoncé-hôte, ils peuvent néanmoins se composer avec autre chose, par exemple qualifier un événement extérieur ou un état mental du locuteur. On pourrait dire que les MD sont non compositionnels (non restrictifs) au niveau du contenu propositionnel de leur énoncé-hôte, mais qu'ils se combinent, comme des fonctions aux arguments qu'ils relient ou caractérisent. Je reviens sur ce point dans la section 3.1.4 (p. 111).

(iii) Le terme *procédural* est régulièrement avancé pour désigner la partie du sens qui correspond à des instructions. Dire que les MD ont un sens essentiellement procédural a comme conséquence de réduire leur fonctionnement référentiel, de le rendre abstrait et de les tirer du côté des mots-outils. Par exemple, dans un énoncé comme

- (55) Alors, tu viens ?

on dira que *alors* donne l'instruction à l'interlocuteur de chercher dans le contexte une situation telle que le fait qu'il vienne soit une continuation d'une séquence d'actions ou d'étapes. En réalité, il n'est pas si simple de considérer que les MD n'ont de sens *que* procédural. Ce terme et le fonctionnement sémantique associé sont présentés et discutés dans la section 3.1.3.

### **3.1.1.2 Les propriétés pertinentes pour les MD**

J'indique ici brièvement les propriétés que j'ai prises en compte pour identifier la classe des MD dans une acception plus large que celle qui correspond aux marqueurs pragmatiques de Brinton. Après avoir comparé différentes approches et après avoir annoté des données authentiques, je me suis concentrée sur trois des propriétés fréquemment mentionnées pour identifier les MD et opérer des distinctions à l'intérieur de la classe. Ces propriétés recourent en partie les considérations précédentes sur le procédural et la non-compositionnalité. Il s'agit de l'indexicalité, de la portée et du niveau de signification.

*L'indexicalité.* Intuitivement, un MD demande qu'on identifie un référent plausible pour qu'une interprétation puisse être construite. Cela a pour conséquence qu'on peut envisager des analyses dites procédurales qui mettent en jeu des instructions pour l'identification de ce référent. Le caractère indexical et procédural des MD est discuté dans les sections 3.1.2. et 3.1.3.

*La portée.* En elle-même, l'indexicalité ne permet pas de distinguer un pronom comme *il* ou *la* d'un MD comme *donc* ou *bon*. Pour aller plus loin, il faut étudier le type de référent que les MD demandent de reconstruire (la portée ou domaine). C'est l'objet de la section 3.1.4.

Le *niveau de signification*. Les MD peuvent contribuer à la signification à différents niveaux. Généralement, ils sont présentés comme étant hors du contenu propositionnel, ce qui les rapproche des présuppositions. Dans une phrase, les tests habituellement utilisés pour faire le tri entre ce qui est intégré ou non à son contenu propositionnel sont la négation, l'interrogation et le clivage. En pratique, ils produisent des résultats variables pour des items que l'on reconnaît pourtant comme des MD (typiquement la différence de comportement de *parce que* et *puisque* face à ces tests). Le problème du niveau de signification est analysé dans la section 3.1.5.

### 3.1.2 Indexicalité et dépendance contextuelle

Les MD comportent dans leurs conditions d'emploi une indétermination sémantique qui ne peut être précisée que grâce à un recours au contexte. On peut ainsi les dire *dépendants du contexte*. Plus précisément, on peut les qualifier d'éléments *indexicaux*, comme le fait par exemple Diewald :

As to the connective function pragmatic markers [MD pour moi] in general are acknowledged to share, it is referred to in the present paper as indexical or relational function. It is a pointing relation between an *origo* (a starting point), and a target, i.e. the entity pointed to. As such – as an indexical relation – it is always a *two-place relation* no matter whether the target is expressed in the linguistic string or not. (Diewald 2013, p. 24, voir aussi Detges & Waltereit 2009, p. 44)

Une partie de la signification des MD est donc à un niveau très général une instruction de pointage vers une cible, linguistique ou non. C'est là la partie procédurale de leur signification et là que s'articulent indexicalité et dépendance contextuelle. L'indexicalité et la dépendance contextuelle sont liées mais ne sont pas équivalentes. Une expression indexicale est une expression qui invite conventionnellement à chercher un référent pour enrichir l'interprétation et, comme dans tout processus d'interprétation, on fait appel au contexte, qu'il soit linguistique, situationnel ou cognitif. Mais l'instruction d'« aller voir ailleurs » est-elle suffisante ? Les expressions indexicales, MD ou non, donnent aussi des indications, parfois assez abstraites, sur le type de référent à rechercher, c'est là une partie plus descriptive, conceptuelle, de leur signification. C'est ce qui fait que des pronoms comme *elle* ou *ça* vont demander tous les deux la recherche d'un référent dans le contexte mais avec des contraintes de sélection différentes (féminin, animé, etc.) Il en est de même pour les MD qui, comme indexicaux, donnent l'instruction de relier deux éléments (*two-place relation*), mais qui comportent aussi dans leur signification des contraintes sur la sélection des éléments à relier et sur la valeur de la relation. Dans le cas d'items polyfonctionnels dans la langue (*la preuve, du coup, disons, bon*, etc.), les contraintes sont souvent dues au sens lexical d'origine (littéral) attaché au MD qu'il n'a pas complètement perdu.

Ces remarques soulèvent deux types de problèmes : d'une part, il faut disposer d'une définition suffisamment générale de l'indexicalité (qui ne réduit pas à la *deixis*), d'autre part, il faut comprendre le rapport entre les informations sémantiques apportées par le code linguistique et le traitement des indexicaux.

La définition des indexicaux ci-dessous :

Indexicals are linguistic expressions whose meaning remains stable while their reference shifts from utterance to utterance. (Cappelen & Lepore 2002, p. 271)

est suffisamment large pour prendre en compte tous les types de contextes et de référents vers lesquels les indexicaux peuvent pointer. La différence qui y est opérée entre la stabilité du sens et la variation de la référence n'est pas sans rappeler celle entre *sens* (*Sinn*) et *dénotation/référence* (*Bedeutung*) chez Frege (1892). Pour simplifier, dans les célèbres exemples commentés, il faut distinguer deux cas : celui des expressions (prototypiquement les noms) et les propositions. La dénotation d'une expression renvoie globalement à une portion de la réalité. Le sens, comme « mode de dénotation de la dénotation », correspond aux propriétés conventionnellement associées à l'expression et qui permettent la sélection de la dénotation pouvant être nommée par l'expression en question. La dénotation d'une proposition complète ne renvoie pas à une portion de la réalité mais à la valeur de vérité, vraie ou fausse, de cette proposition dans le monde. C'est en cela que l'approche de Frege est dite véri-conditionnelle.

Que peut-on retenir pour les MD de ces réflexions générales sur l'indexicalité et la dénotation ?

– D'abord que les MD fonctionnent comme des expressions indexicales, et qu'à ce titre elles présentent des propriétés sémantiques relativement stables (non dépendantes du contexte et conventionnelles) et un référent variable (dépendant du contexte) mais sélectionné grâce aux propriétés sémantiques.

– Ensuite que ce à quoi les MD renvoient ne correspond pas bien à la dénotation stricte de Frege. Les MD ne sont pas l'équivalent d'expressions, au même titre qu'un nom. Ils ne renvoient ni à des faits ni à des objets. A priori, ils ne sont pas non plus l'équivalent de propositions complètes qui auraient comme dénotation une valeur de vérité. Cela peut expliquer pourquoi, pour les MD, les propriétés sémantiques (leur sens) et ce à quoi ce sens permet d'accéder (leur dénotation) ne sont pas faciles à circonscrire ni à formuler. Les MD semblent agir à un niveau plus abstrait, qu'un cadre plus pragmatique permet de mieux expliquer.

### 3.1.3 Composante procédurale et fonctions de base

La composante procédurale, qui est la seule propriété sémantique dans la liste de Brinton (2017, p. 9 et ici Annexe 4, p. 197), est reprise dans la plupart des définitions des MD. Cette propriété est présentée comme centrale pour en caractériser le fonctionnement sémantico-pragmatique et pour les rapprocher d'autres expressions.

Le procédural peut se comprendre par opposition au *conceptuel*. Cette distinction est empruntée à la théorie de la pertinence (Sperber & Wilson 1995 [1986])<sup>155</sup>. Elle est

<sup>155</sup> La théorie de la pertinence a déjà été évoquée dans la section 2.2.1.1 (p. 75) sur l'argumentation à propos de l'auditoire. Elle se présente comme une théorie *cognitive* de la communication qui repose sur une analyse des processus d'interprétation. Elle prend son point de départ dans les travaux de Grice (1975) et vise à unifier les différentes contraintes en une contrainte unique,

également associée à la distinction antérieure entre *connaissance procédurale/connaissance déclarative* en psychologie et en intelligence artificielle<sup>156</sup>. Son application aux MD est due à Blakemore (1987 avec quelques aménagements en 2002, p. 4 et chap. 3)<sup>157</sup>. Globalement, la distinction intervient à partir du moment où l'on considère que, dans la langue, il existe des expressions qui encodent non pas des concepts ou des représentations conceptuelles, mais des procédures (ou instructions) pour enrichir la forme logique de base. Il est important de souligner que ces instructions sont imposées au niveau du code linguistique, par le MD, et non pas construites à partir d'inférences utilisant uniquement le contexte.

This means that, in principle, linguistic form could encode not only the constituents of the conceptual representations that enter into inferential computations but also information which constrains the computations in which these computations are involved. In other words, it is possible for linguistic form to encode either *conceptual* information or *procedural* information. (Blakemore 2002, p. 78)

Quand Blakemore parle de marqueurs, elle cible spécifiquement les connecteurs (*discourse connectives*). Pour elle, ils servent de guides d'interprétation contraignant la construction des contextes et des effets cognitifs recherchés par le locuteur. En revanche, sans indice supplémentaire (voir par ex. la section 2.1 sur les juxtapositions, p. 50), la relation de cohérence entre deux phrases doit être inférée. Parfois, plusieurs interprétations sont en concurrence, comme le montrent les différents cas de (56) :

- (56) Il pleut. Je vais me promener.
- a. Il pleut. [Il découle du fait qu'il pleut que] Je vais me promener.
  - b. Il pleut. [Malgré le fait qu'il pleut] Je vais me promener.
  - c. Il pleut. [La preuve du fait qu'il pleut vient du fait que] Je vais me promener.
  - d. Il pleut. [il est également vrai que] Je vais me promener.

Si le locuteur utilise un connecteur, comme en (57), il choisit/impose une interprétation parmi celles possibles.

- (57) a. Il pleut. Donc je vais me promener.  
 b. Il pleut. Mais je vais me promener.  
 c. Il pleut. En effet, je vais me promener.  
 d. Il pleut. De plus, je vais me promener.

Ainsi, le locuteur contraint son interlocuteur à envisager les deux phrases sous un angle particulier. En calibrant l'effet cognitif visé, le locuteur réduit l'effort cognitif de l'interlocuteur et l'aide à construire une interprétation. Les connecteurs n'encodent pas ici un concept mais une procédure d'inférence qui vise à faciliter le travail d'interprétation et à renforcer la pertinence de l'énoncé<sup>158</sup>:

appelée *principe de pertinence optimale*. Pour simplifier, la pertinence d'un énoncé est dite optimale quand elle est la plus économique en termes de compromis effort-effet compte tenu des conditions concrètes de l'échange, incluant d'éventuelles limites des participants à cet échange. Cette théorie a donné lieu à de nombreux développements et de multiples applications. Pour des synthèses récentes, voir par exemple Clark (2013), Padilla Cruz (2016), Wilson (2017). Voir aussi le recensement des publications dans le domaine : <https://personal.ua.es/francisco.yus/rt.html>

156 Voir Anderson (1976), Winograd (1975) et Hofstadter (1989 [1979]) que cite Hansen (1998a, p. 78).

157 Pour des présentations ciblées de l'opposition *procédural/conceptuel* dans le domaine linguistique, voir Blakemore (2002, p. 89-98), Carston (2016), Clark (2013, p. 308-327) et Wilson (2017, p. 97-99).

158 Voir aussi : « These items indicate something about the context in which the utterances of which they are part are to be processed. They guide the hearer towards intended contextual effects,

Since the degree of relevance increases with the number of cognitive effects derived and decreases with the amount of processing effort required for their derivation, the use of an expression which encodes a procedure for identifying the intended cognitive effects would be consistent with the speaker's aim of achieving relevance for a minimum cost in processing. (Blakemore 2002, p. 79)

En plus du caractère procédural (instructionnel) des MD, Blakemore différencie trois catégories correspondant aux trois effets cognitifs reconnus par la théorie de la pertinence. La spécificité procédurale des MD est de faciliter l'obtention d'un de ces trois effets dans un espace de croyances : *inférer* une proposition, *renforcer* une proposition ou *affaiblir* (voire contredire ou éliminer) une proposition (1992 et 2002, p. 95). Par exemple, si l'on compare (58a) et (58b), l'intuition est que les marqueurs soulignés guident l'interprétation de l'interlocuteur.

- (58) a. Ben peut ouvrir le coffre fort de Tom. Après tout, il connaît la combinaison.  
 b. Ben peut ouvrir le coffre fort de Tom. Donc il connaît la combinaison. (d'après Blakemore 2002, p. 95)

*Après tout* l'incite à rechercher une proposition qui puisse être renforcée par « il connaît la combinaison »<sup>159</sup>. *Donc* l'incite à rechercher une proposition dont « il connaît la combinaison » puisse être dérivée.

Les analyses sont plus complexes dans le détail et il faudrait aller plus loin sur le type d'inférence en jeu et la sélection des contextes<sup>160</sup>. Cela soulève le problème de la distinction entre des MD aux significations apparentées comme par exemple *but-however-nevertheless* pour l'anglais ou *cependant-toutefois-mais-pourtant-néanmoins-par contre-en revanche* pour le français<sup>161</sup>.

Le recours au procédural m'inspire deux remarques. La première concerne l'extension de son usage et la seconde la répartition du conceptuel et du procédural dans la signification des MD.

– *Remarque 1.* La signification procédurale ne se réduit pas aux connecteurs. Le spectre s'est sensiblement élargi. D'autres études défendent l'idée que les pronoms, les particules énonciatives, les interjections, les expressifs en général, l'intonation et certains indicateurs de modalités fonctionnent également de cette manière, en ce sens qu'ils indiquent une procédure pour appréhender le sens d'un énoncé plutôt qu'ils ne le décrivent ou n'en décrivent un élément (Wilson & Sperber 1993, Sasamoto & Wilson 2016). La signification procédurale apparaît aujourd'hui plus hétérogène qu'elle ne l'était dans les premières définitions, au point que certains se demandent si tout ce qui relève du code n'est pas finalement procédural (Carston 2016, p. 12). En effet, à un certain de degré de généralité, le sens procédural ne peut-il pas désigner le

---

and hence reduce the overall effort required to process the discourse. » (Bezuidenhout 2003, p. 103)

159 « *after all* is analysed as encoding the information that the intended inferential route is one which results in the strengthening of an existing assumption. » (Blakemore 2002, p. 96). Pour l'idée que *après tout* en français fait référence à un implicite, voir Roulet (1987a).

160 Je renvoie à la discussion dans Wilson & Sperber (1993, 2012, p. 149-168), aux remarques de Blakemore (2002, p. 4) sur sa propre évolution, et à l'analyse de Clark (2013, p. 321-322).

161 Voir Blakemore (2002, p. 95) sur la différence entre *but*, *however* et *nevertheless*, Iten (2005) sur la différence entre *but* et *although* et Lenepveu (2007) sur la différence entre *cependant*, *toutefois* et *néanmoins*.

fonctionnement symbolique de base des formes linguistiques indexicales, qui demandent la recherche de leur référent ?

– *Remarque 2.* Dans la littérature, on attribue souvent un sens essentiellement procédural aux MD en en faisant une de leurs particularités. Autrement dit, ce qui les caractériserait serait d'être principalement des donneurs d'instructions. Une position qui ne leur ôte pas toute capacité « conceptuelle » pourrait être intéressante pour expliquer les différences entre les MD. On pourrait ainsi considérer que le sens des MD et des expressions indexicales en général résulte d'une collaboration entre la démarche de recherche d'un référent (une vision restreinte du procédural) et une dimension descriptive (pour chercher le référent, il faut utiliser la manière dont l'expression indexicale le décrit, le présente)<sup>162</sup>.

Dans ce cas-là, la partie procédurale du sens d'un MD serait le seul fait de donner l'instruction de rechercher un référent dans le contexte. Les contraintes sémantiques de sélection de ce référent qui sont associées à la forme linguistique du MD constitueraient plutôt la composante descriptive (conceptuelle) de son sens. Les exemples (59) et (60) illustrent ce mode de fonctionnement général – qui n'est donc pas une spécificité des MD. (59) présente une expression honorifique et (60) un MD. Dans les commentaires, les propriétés (a) correspondent à la partie procédurale du sens et les propriétés (b) à la partie conceptuelle.

Dans :

(59) Sa Majesté est trop généreuse.

L'expression honorifique *Sa Majesté* (a) invite conventionnellement à rechercher une entité qui (b) est une personne occupant une position hiérarchique particulière par rapport au locuteur. (a) est la composante instructionnelle de l'expression indexicale, (b) est un élément stable relevant d'une implicature conventionnelle (Potts 2005) ou de la signification dite *use-conditional* ou *secondaire* (Gutzmann 2015, Gutzmann & Turgay 2019).

Dans l'exemple suivant, qui reprend l'exemple (55), on peut préciser ce qui relève de l'instruction et ce qui relève de contraintes liées au sens de l'item lexical :

(60) Alors tu viens ?

Le MD *alors* invite à (a) rechercher une situation telle que (b) le fait que l'interlocuteur vienne constitue une continuation d'une séquence d'actions ou d'étapes. Le fait que *alors* invite à rechercher une situation fait partie de ses portées possibles.

Pour conclure, la notion de procédural est utile pour pour calibrer le fonctionnement des MD, mais elle peut s'entendre d'au moins deux manières : une manière restreinte qui consiste à se focaliser sur l'instruction générale de recherche d'un référent, et une manière plus large qui intègre aussi la nature des contraintes de sélection du référent (par exemple, le sens consécutif de *donc*, le sens temporel de *après* ou les valeurs émotionnelles ou attentionnelles de *ah*, etc.) Concernant cette deuxième interprétation, il faut noter que même pour les marqueurs réputés faibles, sémantiquement presque vides ou *blanchis*<sup>163</sup>, il reste souvent un minimum de contenu,

162 Cette idée n'est pas nouvelle, elle est déjà très présente chez Frege (1892).

163 Le blanchiment sémantique (*semantic bleaching*) est mis en avant depuis longtemps comme une propriété du changement linguistique et des processus de grammaticalisation et de

de teinte, dont l'origine est parfois plus concrète que ce qu'on pense (par exemple la composante de récence pour *du coup*, de complétude pour *bon* ou de spatio-temporalité pour *là*, etc.)

### 3.1.4 La portée

Derrière l'intuition que les MD ne font pas tous la même chose, même lorsqu'ils appartiennent à une même grande classe (les causaux/explicatifs, les concessifs, les additifs, etc.), on retrouve le plus souvent les deux aspects mentionnés précédemment et reformulables ainsi :

- Que « fait » conventionnellement le MD, quel est son « potentiel dynamique » ?
- De quoi peut-on ou doit-on partir dans le contexte pour appliquer les instructions qu'il donne ?

Je m'intéresse ici à un aspect particulier du deuxième point, à savoir ce qu'un MD code comme *type* d'objet sémantique du contexte admissible pour jouer le rôle de référent de l'indexical.

On sait, depuis au moins Redeker (1990) et Sweetser (1990), que certains MD (les connecteurs) peuvent concerner des objets sémantiques variés. Sweetser distingue ainsi différents plans, qu'elle appelle *domaines*, en particulier dans son analyse des connecteurs *and*, *or* et *but* et des conditionnelles (1990, p. 11, 86-112 et 113-123). Il s'agit du domaine des événements ou états de choses (*content domain*), du domaine des états épistémiques (*epistemic domain*) et du domaine des actes de langage (*speech act domain*)<sup>164</sup>. Cette tripartition semble la plus appropriée pour des connecteurs associés à une relation de discours de dépendance, c'est-à-dire une relation qui présente la survenue d'un objet (état de choses, état de croyances, acte de langage) comme provoqué par ou provoquant la survenue d'un autre objet. Les études mettant en évidence des différences de portée ont en effet privilégié les relations de discours causales, consécutives et conditionnelles et les MD associés<sup>165</sup>. L'exemple (61) illustre les trois dimensions pour *parce que*.

- (61) a. La route est mouillée parce qu'il a plu. [relation entre états de choses]  
 ⇒ L'événement la route est mouillée est causé par l'événement qu'il a plu.
- b. Il a plu parce que la route est mouillée. [relation entre états de croyances]  
 ⇒ Le locuteur pense qu'il a plu parce qu'il voit que la route est mouillée.
- c. T'as mis tes baskets ? Parce que la route est mouillée [relation à un acte de langage]  
 ⇒ Le locuteur demande si l'interlocuteur a mis ses baskets parce qu'il voit/sait/pense que la route est mouillée.

La relation de dépendance est assez évidente pour les états de choses et les états de croyances. Dans (61a), *parce que* donne l'instruction de considérer l'évènement à sa

---

pragmaticalisation (voir entre autres Sweetser 1988, Narrog & Heine 2012).

164 Plus récemment, la répartition en domaines a été significativement enrichie à l'aide de la notion d'*espace mental* initialement introduite par Fauconnier (1985). Voir aussi Sanders et al. (2009, 2012).

165 Voir par exemple Debaisieux (2004) et Torck (1996) pour *parce que* en français, ou Jayez & Rossari (2001) pour une application des trois domaines de Sweetser aux connecteurs de conséquence.



droite comme étant la cause de l'événement à sa gauche. Dans (61b), *parce que* donne toujours l'instruction de relier les deux propositions mais sa portée est différente, il indique que le locuteur « abduit » la connaissance de l'événement à sa gauche d'après sa perception de l'événement à sa droite.

Mais l'idée d'une dépendance est aussi très plausible pour les actes de langage. Ainsi, dans (61c), l'acte de question est présenté comme légitimé par le constat du locuteur que la route est mouillée et une paraphrase adéquate serait : *t'as mis tes baskets ? Je te demande ça parce que je vois/sais/pense que la route est mouillée.*

Le cas des états de croyances et des actes de langage rejoint donc le thème de l'argumentation, puisqu'un locuteur présente un objet comme garantissant la vérité ou la pertinence d'un autre objet. La situation est moins claire pour d'autres catégories de MD dont les fonctions ne sont pas aussi transparentes.

Tout en reconnaissant, au moins implicitement, le caractère instructionnel des MD, plusieurs approches insistent sur la variété des fonctions qu'ils remplissent, fonctions souvent qualifiées de pragmatiques.

Schiffirin analysait déjà les MD selon les plans du discours qu'ils mettent en jeu (1987, p. 24-29 et p. 316). Ainsi distingue-t-elle ceux qui interviennent au niveau de l'état d'information (*information state*), du cadre de la participation (*participation framework*), de la structure liée aux idées (*ideational structure*), de la structure liée à l'action (*action structure*) et de la structure liée à l'échange (*exchange structure*).

Plus tard, Dostie avance aussi l'idée que les marqueurs sont polyfonctionnels et peuvent jouer un rôle à plusieurs niveaux : textuel, interactionnel et métadiscursif, ce qui la conduit à différencier les connecteurs textuels, les marqueurs illocutoires et les marqueurs d'interaction (Dostie 2004, p. 45-48, voir ci-dessous Figure 20, p. 120).

Dans la synthèse de la littérature qu'elle propose, Fischer (2006, p. 1-20) fait également ressortir l'idée de fonctions diverses qui ne sont pas réductibles à des connexions entre les trois domaines évoqués par Sweetser.

Enfin, dans un article récent, Crible et Degand (2019) utilisent aussi un spectre large de domaines et de fonctions en se basant sur la classification de Redeker (1990). Chez ce dernier, les MD sont des expressions signalant la relation d'un énoncé à un contexte immédiat. Cette relation peut être de trois sortes : *idéationnelle* (*ideational*), *rhétorique* ou *séquentielle*. Les deux dernières relations sont présentées comme des relations d'ordre pragmatique (Redeker 1990, p. 369). Crible et Degand ajoutent un quatrième domaine dit *interpersonnel* (2019, p. 11) et une quinzaine de relations de discours (ou fonctions) définies de manière transversale aux quatre domaines ou selon des propriétés spécifiques à chaque domaine (p. 12-13).

Bien que n'étant pas toujours facilement articulables, ces différents travaux se répondent tout de même sur l'idée que les MD ont une portée assez diversifiée qui dépasse largement le niveau des états de choses.

Dans certains de mes travaux (Dagnat & Jayez 2020e [37], Dagnat 2020f [38], Dagnat 2021a [39]), j'ai été amenée à étudier des MD qui font référence à des objets/domaines encore plus variés. Cette possibilité intéresse surtout les MD que

j'appelle plus loin *particules énonciatives* (PEN), mais elle existe aussi pour des connecteurs et d'autres types de particules<sup>166</sup>. Ces MD peuvent selon les cas faire référence à :

- un événement extérieur,
- un événement cognitif réel ou visé<sup>167</sup> interne au locuteur (variation affective, attentionnelle et/ou épistémique du locuteur),
- un événement cognitif réel ou visé externe au locuteur (variation affective, attentionnelle et/ou épistémique d'un interlocuteur),
- la production du discours par le locuteur,
- un état cognitif antérieur à l'énonciation.

J'illustre ces possibilités par les exemples de (62) où les MD sont soulignés, leur catégorie entre parenthèses<sup>168</sup> et la catégorie du référent entre crochets.

(62) *Contexte : le locuteur veut rentrer chez lui avec un ami, mais la portée est fermée et c'est sa femme qui a la clef.*

- a. Bon (PEN) [événement extérieur], du coup (connecteur) [événement extérieur] on peut pas rentrer
- b. Et voilà (PEN) [événement extérieur]. Pourtant (connecteur) [événement extérieur], je lui avais dit de ne pas fermer à clef.
- c. Zut (PEN)[événement extérieur], on peut pas rentrer. Ah (PEN) [événement cognitif interne], le voisin a un double.
- d. On peut pas rentrer. Je vais voir le voisin. Ah ! (PEN) [événement cognitif interne] ma femme lui avait repris la clef.
- e. A – Zut (PEN) [événement extérieur], et j'ai oublié la clef !  
B – Attends (PEN) [événement cognitif externe visé de type variation attentionnelle], ton voisin avait la clef, non ?
- f. La porte est bien fermée (particule autre que PEN) [état cognitif antérieur à l'énonciation]. C'est exactement ce que ma femme m'avait dit.

Ces exemples appellent plusieurs remarques.

– *Remarque 1.* Ils mettent parfois en jeu des MD connecteurs (*du coup, pourtant*). D'une manière générale, certains connecteurs peuvent faire référence à des domaines plus riches que ceux de Sweetser.

– *Remarque 2.* Il peut y avoir un risque de confusion sur l'identité de la portée, notamment lorsqu'elle consiste en événements cognitifs internes (variation affective, attentionnelle et/ou épistémique). Ainsi, en (62d), le locuteur porte son attention sur le fait que sa femme a repris la clef au voisin et qu'il ne peut donc pas aller chez ce voisin pour récupérer la clef. On pourrait penser que c'est le contenu propositionnel de l'affirmation *ma femme lui avait repris la clef* qui constitue la référence de *Ah*, puisque c'est ce contenu qui est asserté. On pourrait aussi se demander si ce n'est pas l'acte de langage, l'assertion véhiculée par *ma femme lui avait repris la clef*, qui constitue la référence de *Ah*.

Il y a deux raisons de penser que c'est simplement la prise de conscience de la situation qui est ciblée (un événement cognitif), qu'elle soit formulée linguistiquement ou non. Premièrement, si la situation n'est pas formulée linguistiquement (*On peut pas rentrer. Je*

166 Voir section 3.2.1 pour des critères distinctifs entre ces catégories de MD.

167 J'entends par *visé* tout événement ou état dont un locuteur cherche à obtenir la réalisation.

168 La distinction PEN/connecteur est précisée dans la section 3.2.2.1 (p. 121).

*vais voir le voisin. Ah !*) mais reste au stade de la prise de conscience, cela ne change pas la portée de *Ah*. Deuxièmement, si la situation est formulée linguistiquement (*On peut pas rentrer. Je vais voir le voisin. Ah ! ma femme lui avait repris la clef*), il faut tenir compte du fait que *Ah* exprime une réaction émotionnelle positive (satisfaction, admiration, surprise agréable, etc.) ou négative (surprise désagréable, dégoût, colère, etc.) ou/et une augmentation de l'attention sur un élément jugé nouveau ou pertinent. Il serait difficile d'imaginer ici que le locuteur réagisse à son propre acte de langage et/ou au contenu propositionnel associé<sup>169</sup>. On rejoint tout à fait les propos de Ducrot à propos des interjections lorsqu'il en parle en termes de « réaction déclenchée par la représentation d'une situation » (Ducrot 1984, p. 187, c'est moi qui souligne).

– Une *troisième remarque* concerne les croyances. Un mot comme *bien* peut servir à confirmer une proposition. On fait alors référence à un espace de croyances dont la proposition faisait partie. (62e) pourrait alors être paraphrasé par : *conformément à ce qui avait été dit/pensé, la porte est fermée*. Cet espace de croyances n'est pas forcément celui du locuteur puisque ce dernier peut confirmer une croyance attribuée à quelqu'un d'autre. *Bien* se comporte dans des exemples du type (62e) comme une particule dite *modale*, pour utiliser la terminologie des études germaniques, c'est-à-dire une particule qui joue sur les croyances et les attentes des participants<sup>170</sup>.

– *La quatrième et dernière remarque* concerne l'accès au référent. Comme avec tous les indexicaux, celui-ci peut être plus ou moins clairement identifiable selon les contextes. Les MD offrent de ce point de vue deux particularités. Certains MD renvoient souvent à des événements internes qu'il n'est pas possible d'objectiver (sentiment, impression, etc.), ce qui ne les empêche pas d'être interprétables dans la mesure où l'interlocuteur peut discerner, notamment avec l'aide du contexte et de l'intonation, le *type* de l'événement interne qui constitue le référent<sup>171</sup>. D'autre part, dans des échanges relativement longs, les MD peuvent pointer sur une série d'interventions d'un locuteur ou d'échanges entre locuteurs. Dans une théorie de l'*attachement* comme la SDRT (Asher & Lascarides 2003), on considérerait, au moins dans certains cas, que le segment hôte du MD est attaché à toute la séquence discursive qui contient les interventions ou échanges en question. Dans Dagnat (2020f [38], p. 7, 2022b [42], p. 191), j'ai avancé que, dans ce type de cas, il faut prévoir la possibilité de référents « globaux », comme par exemple l'attitude générale d'un locuteur (ce que, à propos de *mais*, j'ai appelé un *comportement*, voir section 3.2.2.2, p. 140).

### 3.1.5 L'intégration des MD

Dans cette section, je discute de la façon dont les MD s'intègrent ou non dans le contenu sémantique des phrases (leur *compositionnalité* chez Brinton). J'insiste en

169 Reste à voir si cela est plus facile dans des situations dialogales où le locuteur utilisant *Ah* réagirait aux propos de son interlocuteur.

170 Il existe de très nombreux travaux depuis les premières définitions de Weydt pour l'allemand (1969). Voir par exemple Hansen (1998a, p. 41-46), Métrich et al. (1998-2002, p. 24-36), ou encore Waltereit (2001) et Karagjosova (2003). On ne reconnaît au français que de rares exemples (*bien, déjà, donc, peut-être, seulement, quand même*, etc.). Je reprends cette question à la section 3.2.1.2 à propos de la possibilité ou non d'établir une liste des MD.

171 Il s'agit donc d'un cas d'*accommodation* au sens bien connu dans la littérature sur les présuppositions. Voir Lewis D. 1979. « Scorekeeping in a language game ». *Journal of Philosophical Logic* 8, p. 339-359.

particulier (i) sur les limites de la notion de procédural pour étudier la participation des MD au contenu propositionnel et (ii) sur l'utilité d'emprunter des notions intervenant pour l'analyse des présuppositions (projection, effet local, déplaçabilité).

(i) *Limites de la notion de procédural pour l'étude des MD*

Un item fonctionnant de manière procédurale a parfois été présenté comme systématiquement hors du contenu propositionnel de l'énoncé-hôte. Cette position radicale, qui était celle de Blakemore (1987), est reprise dans la plupart des définitions des MD, comme l'illustre la propriété *g* chez Brinton (ici Annexe 4, p. 197) ou des propriétés analogues chez Heine (2013). En réalité, les deux aspects ne sont pas complètement assimilables, et Blakemore (2002) nuance elle-même sa première définition (voir aussi Carston 2016, Hansen 1998a, p. 75, Sperber & Wilson 1995 [1986], p. 258-259).

Bien qu'ayant une dimension procédurale, certains MD peuvent en effet intervenir au niveau du contenu propositionnel. En français, c'est le cas d'un certain nombre de subordinants comme par exemple *parce que*<sup>172</sup>, *afin/pour que*, *quand* etc. qui sont compatibles avec des structures clivées, négatives ou interrogatives (voir Abeillé & Godard 2021, en particulier chap. XIV-1, p. 1549, 1585, 1599, 1616-1617, 1635)<sup>173</sup>.

Une opposition bien connue est celle de *parce que* et *puisque* (Groupe λ 1975, Nazarenko 2000, p. 61-71, Zufferey 2014, etc.). Les manipulations de l'exemple (63) tendent à montrer que *parce que* est intégré au contenu propositionnel, et les manipulations en (64) montrent que ce n'est pas le cas de *puisque*.

- (63) a. Paul a raté son train parce qu'il était en retard à la gare.  
 b. C'est faux, il n'a pas raté son train parce qu'il était en retard, mais parce qu'il y avait une erreur dans l'affichage des horaires.  
 c. C'est parce qu'il était en retard à la gare que Paul a raté son train.
- (64) a. Paul a raté son train puisque il était en retard à la gare.  
 b. ? C'est faux, il n'a pas raté son train puisque il était en retard, mais puisque il y avait une erreur dans l'affichage des horaires.  
 c. ? C'est puisque il était en retard à la gare que Paul a raté son train.

Par ailleurs, lorsqu'un MD se trouve dans la portée syntaxique d'un verbe lié à un état épistémique<sup>174</sup>, on peut se demander si ce qui est exprimé à travers l'emploi d'un MD doit être attribué au locuteur ou à l'agent du verbe, autrement dit sous quelle *perspective* on interprète la relation associée au MD. Par exemple, dans (65) :

- (65) Marie se demande s'il a quand même eu son train.

est-ce que la relation de concession marquée par *quand même* fait partie des croyances de Marie ? En d'autres termes, est-ce que Marie pense qu'il y avait des raisons pour que Paul n'ait pas son train ?

172 Le référent de *parce que* ne peut pas être intégré au contenu propositionnel s'il porte sur la croyance ou l'acte de langage, ce qui est souvent manifesté par des propriétés prosodiques (pause et nature du contour terminal de la première phrase), voir la différence entre *Est-ce que Paul a eu un accident parce qu'il y avait du verglas ?* Et *Est-ce que Paul a eu un accident ? ... parce qu'il y avait du verglas.*

173 Il s'agit des tests habituellement utilisés pour tester la participation d'un élément au contenu propositionnel.

174 Par exemple : *penser, croire, se demander, estimer.*

La participation ou non au contenu propositionnel ne doit donc pas être un critère catégorique, car, de mon point de vue, il ne permet pas de distinguer ce qui fonctionne comme MD ou pas<sup>175</sup>. En revanche, ce critère peut être pertinent pour distinguer différents *types* de MD. On verra par exemple que des MD ou des emplois de MD plus expressifs ne codant pas une relation de discours (par. ex. *bon, ah, chic, merde*, certains emplois de *quand même, alors*) ne participent en général pas au contenu propositionnel et ne peuvent pas être dans la portée d'un verbe épistémique.

(ii) *L'utilité des notions intervenant pour l'analyse des présuppositions*

Plutôt que de m'appuyer sur la distinction conceptuel-procédural, j'ai exploré la contribution des MD à partir des notions de *projection*, d'*effet local obligatoire* et de *déplaçabilité* habituellement associées à l'étude des présuppositions. Dans Dagnat & Jayez (2020e [37], p. 8-11 et p. 16-20) et Dagnat (2020f [38]), j'ai exploré le statut présuppositionnel de certains MD. Dans la mesure où les MD sont des indexicaux, ils pourraient être considérés comme présupposant l'existence d'un référent ayant certaines propriétés, à l'instar des pronoms personnels.

Les tests couramment utilisés pour identifier un mot ou une expression qui déclenche une présupposition (un *déclencheur*) sont deux types :

- La résistance de la présupposition à certains opérateurs : la négation, l'interrogation et l'intégration du déclencheur dans une expression de possibilité, par exemple une conditionnelle en *si* (tests dits de *projection*<sup>176</sup>). Voir les exemples (66b,c,d)
- La possibilité de suspendre la présupposition en la transformant en hypothèse. Voir l'exemple (66e).

Sur la base de l'exemple (66a), on remarque que le pronom *elle*, élément indexical prototypique, présuppose l'existence d'un référent spécifique et que cette présupposition résiste aux trois opérateurs mentionnés (négation en (66b), question en (66c) et l'intégration dans une hypothétique (66d)). On dira que la présupposition *se projette*. La construction proposée en (66e) montre que l'on peut suspendre la présupposition que ce référent spécifique existe, puisqu'on le rend hypothétique.

- (66) a. Elle a acheté une voiture.  
[présuppose l'existence d'une personne spécifique de sexe féminin]
- b. Elle n'a pas acheté de voiture.
- c. Est-ce qu'elle a acheté une voiture ?
- d. Si elle a acheté une voiture, ...  
[b, c et d continuent à présupposer l'existence d'une personne spécifique de sexe féminin]
- e. Si le département a une directrice, elle dispose d'une voiture.  
[ne présuppose pas qu'il existe une personne féminine spécifique]

175 Ma position n'est pas forcément partagée dans toutes les analyses. Je remercie Maj-Britt Mosegaard Hansen pour sa demande d'éclaircissement sur ce point. En effet, un certain nombre d'études retiennent la (non-)participation au contenu propositionnel comme un des critères discriminants entre MD et non MD. Ce dernier choix conduit alors à distinguer deux *parce que*, l'un qui est dans le contenu propositionnel et pas MD (ex. 63a) et l'autre qui est hors du contenu propositionnel et donc MD. Dans un ex. comme *Paul est chez lui, parce que sa voiture est dans l'allée*, il est difficile de cliver la causale, car elle porte sur un état de croyance et pas sur un état de choses.

176 Voir Langendoen & Savin (1971). On dira qu'une présupposition *se projette* si elle résiste aux opérateurs, en d'autres termes si elle échappe à leur portée.

Le caractère présuppositionnel est aussi couramment admis pour des éléments d'autres catégories, en particulier certains verbes avec les exemples bien connus du type de (67a-b), ou des adverbes comme *encore*, *de nouveau* ou *aussi*.

Prenons l'exemple de *encore* dans son sens itératif. Dans (67c), *Paul a encore raté son examen* présuppose que Paul a déjà raté un examen et cette présupposition résiste à l'opérateur de négation (67d), elle se projette. C'est-à-dire, comme on vient de le rappeler, qu'ils échappent à la portée des opérateurs qui suspendent ou inversent la valeur de vérité du contenu véridictionnel du contenu propositionnel de leur énoncé-hôte.

De façon semblable, dans (67e-f) le MD *pourtant* se projette.

- (67) a. Paul a arrêté de fumer  
[Présuppose que Paul a fumé]
- b. Paul n'a pas arrêté de fumer  
[Continue à présupposer que Paul a fumé : projection]
- c. Paul a encore raté son examen  
[Présuppose que Paul a déjà raté son examen]
- d. Il n'est pas vrai que Paul a encore raté son examen  
[Continue à présupposer que Paul a déjà raté son examen]
- e. Paul a pourtant tout fait pour réussir son examen  
[Présuppose qu'il y a un référent incompatible avec ce qui est exprimé par *Paul a tout fait pour réussir son examen*]
- f. Est-ce que Paul a pourtant tout fait pour réussir son examen ?  
[Conserve la même présupposition]

Un autre indice du caractère présuppositionnel de certains MD est qu'on peut suspendre la présupposition en la rendant hypothétique. Reprenons en (68a) l'exemple (67a), *Paul a arrêté de fumer*. *Arrêter* y déclenche la présupposition que l'action a déjà eu lieu, mais on peut construire (68b) un contexte qui suspend cette présupposition. Les exemples (68c,d) sont une transposition de ce phénomène au domaine des MD, ici *donc*.

- (68) a. Si Paul a déjà fumé, alors il a arrêté  
[la conditionnelle suspend la présupposition]
- b. Paul qui est inscrit en Math1 ne peut donc pas s'inscrire en Math2  
[présuppose que l'on ne peut pas être inscrit à la fois en Math1 et en Math2]
- c. Si on ne peut pas suivre Math1 et Math2 la même année, alors, Paul, qui est inscrit en Math1, ne peut donc pas s'inscrire en Math2  
[la conditionnelle suspend la présupposition]

Enfin, comme les présuppositions, les MD que j'ai étudiés possèdent la propriété *d'effet local obligatoire*, décrite par Tonhauser et al. (2013). Cette propriété renvoie au fait qu'un contenu présupposé est présent dans un espace de croyance autre que celui du locuteur. Cet espace est ouvert par des verbes comme *penser*, *dire*, *croire*, *estimer*, *imaginer*, etc.

Les exemples montrent le parallèle entre le déclencheur classique *arrêter* qui nous sert de fil directeur (69a-b) et un MD, ici *pourtant* (b). Dans les deux cas, la présupposition est comme transférée à l'espace de croyances de l'énonciateur dont le locuteur rapporte les propos ou les pensées.

- (69) a. Paul a arrêté de fumer.  
[Implique que le locuteur pense que Paul a fumé]

- b. Marie pense que Paul a arrêté de fumer  
[Implique que Marie pense que Paul a fumé]
- (70) a. Paul a pourtant tout fait pour réussir son examen.  
[Implique que le locuteur pense qu'il y a un référent incompatible avec ce qui est exprimé par *Paul a tout fait pour réussir son examen*]
- b. Marie pense que Paul a pourtant tout fait pour réussir son examen  
[Implique que Marie pense qu'il y a un référent incompatible avec ce qui est exprimé par *Paul a tout fait pour réussir son examen*]

J'ajoute une dernière remarque concernant la non-correspondance entre les points (i) et (ii). L'appartenance au contenu propositionnel ne coïncide pas avec un statut systématiquement non-présuppositionnel. Par exemple, un MD peut participer au contenu propositionnel (peut être clivé, voir ex. (71b) et (72b)) mais être quand même présuppositionnel (ex (72a)). Cela semble concerner en particulier les MD temporels.

- (71) a. Est-ce que Paul est en retard *parce qu'il a raté son train* ?  
[ne présuppose pas qu'il a raté son train]
- b. C'est *parce qu'il a raté son train*, que...  
[fait partie du contenu propositionnel]
- (72) a. Est-ce que Paul a attrapé le covid *quand il a pris le train* ?  
[présuppose qu'il a pris le train]
- b. C'est *quand il a pris le train*, que...  
[fait partie du contenu propositionnel]

Les exemples que j'ai commentés concernent des MD connecteurs (causal, consécutif, concessif et temporel) pour lesquels les observations ne semblent pas généralisables. La situation semble en revanche plus tranchée avec les MD que j'appelle ci-dessous PEN (voir section 3.2.1, p. 119).

### 3.1.6. En résumé

Je retiendrai des sections 3.1.1 à 3.1.5 les points suivants :

1. Les MD correspondent à des instructions d'interprétation dont le schéma général est d'identifier, au moins partiellement, un référent dont les propriétés sont en parties déterminées par le sens du MD. Cela correspond à l'intuition fondamentale de Frege sur le sens comme auxiliaire de détermination du référent.
2. Les types de référent possibles comprennent des catégories bien décrites et des catégories moins souvent mentionnées. Les types de référent envisageables dépendent du MD considéré, du contexte et des stratégies d'interprétation.
3. La contribution du MD au message varie. On peut distinguer au moins trois cas : certains MD peuvent contribuer au contenu propositionnel, sauf lorsqu'ils portent sur une croyance ou un acte de langage (*parce que*), d'autres sont présuppositionnels et contribuent au contenu propositionnel (*quand*) et d'autres encore sont présuppositionnels mais ne contribuent pas au contenu propositionnel (*donc*).

Une des difficultés que présentent les MD est la formulation explicite de leur sens. Même lorsqu'on a une intuition apparemment claire, elle est prise en défaut lorsqu'on

compare les MD d'une même classe intuitive. Par exemple, quelles différences doit-on faire entre *donc*, *du coup*, *de ce fait alors* et *par conséquent* qui semblent tous associés à une idée de conséquence ou de cause ? Une remarque analogue concerne la variété des emplois d'un même MD, par exemple *alors*, *mais*, *bon* ou *tiens*. Il y a à cela au moins deux raisons possibles. D'une part, la variété des types de référent, donc la difficulté de décrire la portée des MD. D'autre part, le caractère « indicible » (la propriété d'*ineffability* chez Potts (2007, p. 66) du sens de ces items. Est-ce que cette deuxième difficulté provient de leur caractère procédural tel que je l'ai présenté dans la section 3.1.3 ? Sur la base des sections précédentes, il me semble que la résistance des MD à l'analyse ne vient pas forcément d'une absence de contenu conceptuel mais plutôt du caractère abstrait de ce contenu. Dans la mesure où le sens des MD mobilise des relations de discours, des attitudes et du typage, il concentre des propriétés complexes, qui n'ont pas de manifestation directe, par exemple sous forme d'images ou de familles de patrons syntaxiques. De plus, par rapport à des unités lexicales qui désignent des objets ou des événements concrets, les MD sont soumis à des évolutions diachroniques qui font intervenir massivement l'interprétation des locuteurs, comme on le verra dans la section 3.2.2 (p. 134).

## 3.2 Travaux sur les connecteurs et les particules énonciatives

La section précédente était un zoom sur des propriétés que j'ai trouvées complexes parmi l'ensemble des propriétés très fréquemment attribuées à la classe générale des MD et une discussion sur des propriétés moins souvent avancées. La présente section est l'occasion de faire le point sur mes travaux dans le domaine à différentes étapes de ma réflexion. Je soulève d'abord l'utilité de distinguer deux types de fonctionnement qui ont déjà émergé dans certains commentaires. Il s'agit du fonctionnement comme connecteur et du fonctionnement comme particule énonciative (PEN). Je pointe aussi la difficulté de dresser une liste des formes concernées, en particulier pour les PEN (section 3.2.1). Je présente ensuite des travaux portant sur des marqueurs de concession, en intégrant une dimension diachronique et une dimension argumentative (section 3.2.2.). Les deux autres sections me permettent d'insister sur les propriétés prosodiques pour la distinction des types de marqueurs et leurs différentes fonctions (section 3.2.3), et sur la combinaison de marqueurs (section 3.2.4).

### 3.2.1 Vers une cartographie des items et des fonctions

En cherchant à faire une liste des termes employés pour désigner les MD (Tableau 8, p. 103) et en cherchant ensuite à lister les items pouvant fonctionner comme MD, j'ai été confrontée à des sous-classes dont la distinction résulte de l'insistance sur telle ou telle propriété. Dans la littérature, les typologies et les listes ne sont pas toujours les mêmes et les critères de sélection des items candidats pas toujours très clairs. Cette difficulté à identifier les formes et les fonctionnements des MD, pour différentes raisons, a des conséquences sur la constitution et l'analyse de données authentiques. Dans le



cadre de projets ou de la préparation de publications<sup>177</sup>, je me suis demandé par exemple quelles requêtes formuler pour interroger les corpus disponibles déjà annotés, comment annoter ces items si l'annotation n'est pas faite et s'il est possible de les taguer automatiquement.

Dans un premier temps, j'ai repris la distinction classique entre connecteurs et particules énonciatives (PEN)<sup>178</sup>. Je me suis appuyée sur la typologie de Dostie (2004) en modifiant quelques étiquettes pour des raisons de lisibilité, car *marqueur discursif* risquait de porter à confusion. Ci-dessous, la Figure 20 reprend sa schématisation en indiquant en gras les étiquettes que je retiens et, le cas échéant, entre crochets les étiquettes d'origine. Les fonctions sont en italiques.

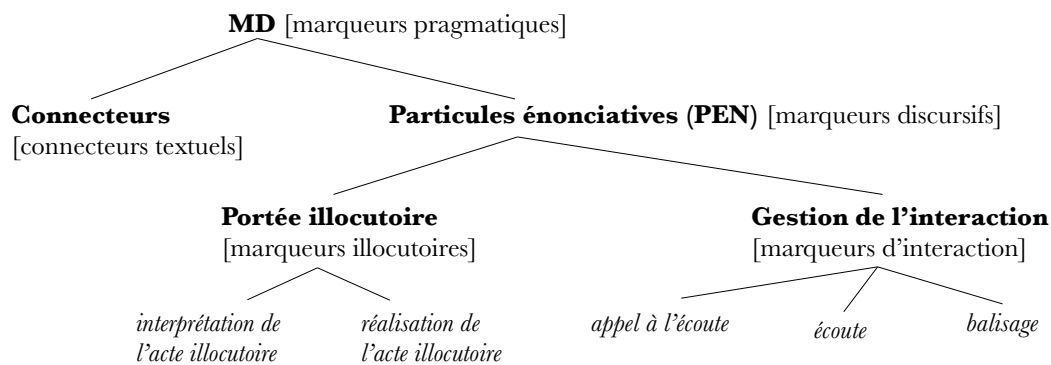


Figure 20. Sous-catégories et fonctions des MD d'après la classification de Dostie (2004, p. 46)

La catégorie des connecteurs n'est pas développée car ce n'était pas l'objet de l'étude. L'auteure renvoie principalement à la définition que propose Fraser des *discourse markers* (1999), qui correspond assez bien également à ce que disent Detges et Waltereit des *discourse markers* lorsqu'ils les distinguent des *modal particles* (2016, p. 638-639). Les connecteurs désignent alors les items lexicaux simples ou complexes qui portent sur des segments textuels – explicites – qu'ils mettent relation. La catégorie des PEN (*marqueurs discursifs* chez Dostie) est en revanche plus détaillée et les différentes sous-catégories permettent d'intégrer un grand nombre de fonctions et d'éléments. Ces sous-catégories sont illustrées plus loin par les exemples (77) à (82).

Les distinctions entre connecteurs et PEN ou à l'intérieur de la classe des PEN sont très utiles et productives mais doivent être envisagées avec une certaine souplesse, notamment pour les raisons suivantes : certains marqueurs sont polyfonctionnels, certains usages demeurent sous-spécifiés et les usages et les fonctions évoluent au fil du temps. Des auteurs remarquent que la distinction entre connecteur et particule n'est pas toujours aisée en pratique (Chanet 2003, p. 3-4), d'autres plaident explicitement pour une « catégorisation flexible » (Diewald 2013, p. 41-42) et l'idée de fonctionnements prototypiques et donc de propriétés « clignotantes » (il y aurait dans ce cas de meilleurs

177 Voir notamment ma participation au projet SYFRAP (resp. C. Cerisara & C. Gardent) pour l'annotation des MD, à la réflexion sur *pour* dans le cadre du French Discourse Treebank (Colinet et al. 2014a [21]), la codirection direction du stage de A. Hafed Dahou (2021), ainsi que, entre autres, Bartkova et al. (2017a [27]), les travaux autour de la thèse de L. Lee (Lee 2021 et Lee et al. 2018 [29]) et Dagnat (2022a [42]).

178 J'emprunte la formule *particule énonciative* à Fernandez-Vest (1994) pour insister sur la lexicalisation des items concernés, même lorsqu'ils sont complexes, et sur leur fonctionnement énonciatif.

représentants de la classe que d'autres) (Schoonjans 2013, p. 141-145). Le processus de pragmatization, y compris la subjectivisation et l'intersubjectivisation qu'il sous-entend, peut jouer un rôle dans ce passage et ce floutage des fonctions entre connecteurs et PEN (Traugott 1982, 1995 et 2010).

Les propriétés générales des MD ont été évoquées et quelques-unes détaillées dans la section précédente. Ci-dessous, j'insiste sur ce qui distingue le fonctionnement comme connecteur et le fonctionnement comme PEN (3.2.1.1). Ceci permet d'envisager des listes distinctes mais ouvertes et non étanches, ce qui pose des problèmes pour l'annotation et l'interrogation de corpus (3.2.1.2).

### 3.2.1.1. *Connecteurs et particules énonciatives*

En tant que MD, les connecteurs et les PEN ont en commun, entre autres, de « recruter » dans les mêmes catégories, d'être invariables, d'être des indexicaux, d'avoir un fonctionnement procédural au niveau du discours et de ne pas assurer de fonction syntaxique par rapport à un élément de l'énoncé-hôte<sup>179</sup>. Néanmoins, ils se distinguent sur plusieurs points, rappelés dans Dagnat (2020f [38], p. 5-8 et 2021a [39], p. 2024-2025). J'insiste ici sur : (i) le type de relation, (ii) l'autonomie, (iii) le rapport au contenu propositionnel et (iv) le comportement par rapport aux propriétés de la présupposition (la projection, l'effet local et la déplaçabilité).

D'autres propriétés pourraient être discutées. Du point de vue phonétique, les PEN ont tendance à être « écrasées » ou « affaiblies ». Du point de vue syntaxique, elles se combinent et peuvent parfois construire des compléments. Sur ce dernier point, voir Dagnat (2021a [39], p. 2019-2023) et la notice à venir dans l'*Encyclopédie grammaticale*<sup>145</sup>.

#### (i) *Le type de relation en jeu et son explicitation*

A ce niveau, la description des connecteurs paraît moins problématique, moins « indicible », que celle des PEN, dans la mesure où ils sont associés à l'identification d'une relation de discours essentiellement de type logico-temporel. La sémantique des relations a été largement abordée à la section 2.1.1.2 (p. 55). Il s'agit du sens de la connexion entre les deux arguments du connecteur, qui sont prototypiquement des propositions explicites, comme en (73) où *alors* marque la relation **Conséquence** (ou **Résultat**) entre les deux propositions, ou comme en (74) où *d'autant que* marque une relation **Explication** entre les deux propositions. Mais il est tout à fait envisageable que le domaine vers lequel il pointe ne soit pas celui des états de choses ou qu'une des deux propositions connectées soit implicite. Leur qualité de connecteurs tient avant tout au fait qu'ils marquent une relation de discours entre deux éléments formulables comme des propositions.

(73) la bourse ferme ce soir et ne rouvrira que mardi matin. Quatre jours de fermeture, alors ça ça n'incite pas beaucoup aux initiatives évidemment. (ESTER 20030417\_1200\_1300\_franceinfo\_dga)

(74) Les enfants peuvent clairement s'identifier aux Power Rangers, d'autant que ce sont des humains qui leur ressemblent. (*Le Monde*, 1995 dans Combettes & Dagnat 2021c [41] ex. 45, p. 65)

179 La question est plus complexe pour les marqueurs de corrélation, car on peut attribuer à la partie adverbiale une portée sur un élément pivot de l'énoncé-hôte (par exemple un adjectif dans les concessives en « *si... que* ». Voir section 3.2.2.1.

Les PEN, si elles « connectent » bien quelque chose également, ne le font pas à la manière des connecteurs. Elles n'encodent pas une relation logico-temporelle entre deux propositions ou deux situations, mais elles servent à connecter le discours dans lequel elles surgissent directement à la situation d'énonciation, par exemple pour intensifier la surprise du locuteur pour le deuxième *alors* dans (75), ou la réaction à une croyance initiale du locuteur lui-même pour *mais* dans (76).

(75) Alors il se tournait vers la photographe en plein reportage sauvage, affublée de ses voiles grecs au troisième rang, et me faisait en riant un geste signifiant : « Oh, toi, alors ! Tu m'as encore eu... » (A. Duperey, *Les photos d'Annie*, 2018, p. 84)

(76) Ah mais c'est qu'il est cool en fait ce jeu d'merde... (blog, <https://www.senscritique.com>)

Quelques marqueurs très fréquents en français ont les deux emplois, connecteur et PEN, par exemple *alors*, *donc* et *mais*. En pratique, il n'est pas toujours simple de trancher. Les fonctions des PEN sont en général formulées en termes d'attitudes épistémiques ou affectives du locuteur vis à vis d'une proposition, d'un événement intérieur ou extérieur ou en termes de gestion de l'interaction. Les sous-catégories de PEN dans la Figure 20 proposée ci-dessus montrent ces deux orientations en les précisant. Je les rappelle et les illustre :

– Les PEN ayant une *portée illocutoire*, sur l'acte de langage, peuvent servir par exemple à orienter l'interprétation d'une assertion (une suggestion dans (77)) ou manifester l'attitude du locuteur (émotionnelle dans (78) et épistémique en (79)).

– En ce qui concerne les PEN permettant la *gestion de l'interaction*, elles sont centrées sur le locuteur et montrent sa disponibilité dans l'échange (ex. 81), ou sur l'interlocuteur qu'elles cherchent précisément à « connecter » (ex. 80). Le fonctionnement dit de balisage correspond aux « ponctuels de la langue » (Vincent & Sankoff 1992, Vincent 1993) que (82) illustre, la haute fréquence étant souvent une donnée associée.

(77) Tu peux apporter de la bière, tiens.

(78) Tiens, vous ici ?

(79) A : Pourquoi il l'aimait à ce point ?

B : Parce que c'était Chet Baker, tiens ! Un immense musicien ! (Anna Gavaldà, *La consolante*, 2008)

(80) tu sais quoi moi j'envie et c'est sincère ce que je vais dire d'accord j'envie les gens qui n'ont pas euh qui n'ont pas de style ok (*Fary is the New Black*, spectacle d'humour, 2018, Netflix)

(81) L2 – moi je suis né à Tiaret

L1 – d'accord

L2 – elle elle est née à Mostaganem (CEFC, Alg-jac-06)

(82) et puis je commence à chanter des trucs un peu hyper cul-cul quoi et genre euh j'écris le texte et je le regarde je le lis je dis putain mais c'est trop cul-cul quoi [...] mais Mon Dieu la meuf c'est une psychopathe quoi genre c'est pas possible et en fait c'était voilà tellement d'amour que s'en s'en on aurait dit une psychopathe qui chantait quoi (Izia Higelin, *France Info*, 11-07-2012, dans Dagnat & Jayez 2020e [37], ex.33 p. 18)

**(ii) L'autonomie**

Beaucoup de PEN peuvent former un énoncé à elles-seules, ce qui est très difficile voire impossible pour un connecteur. Dans (83), *tiens* ou *ah* expriment la surprise du locuteur à propos d'un élément du contexte (soit le tour de parole précédent, soit la situation elle-même) sans avoir besoin d'autre chose. Un connecteur seul demeure incomplet, il sous-entend deux éléments à connecter, qu'il faut reconstruire s'il en manque un. En (84), la personne qui laisse un *pourtant* en suspens est invitée à poursuivre et à développer le second argument. Ce qu'elle fait d'ailleurs par la suite. Mais cela n'est pas toujours le cas. Si l'argument introduit par le connecteur n'est pas explicite ou saillant dans la situation, il demande à être reconstruit, en fonction des indices de l'autre argument et du sens de la relation en jeu. La PEN en revanche est capable d'introduire par sa seule énonciation l'information qui concerne l'élément qui la déclenche. Cette information peut parfois être explicitée par une proposition (83 B'), mais qui n'apporte alors rien de plus.

(83) A – Paul a eu son permis de conduire. / (*Paul passe au volant d'une voiture*)

B – Tiens ! / Ah !

B' – Tiens ! / Ah ! je suis surpris.

(84) – Mon opinion vous importe-t-elle ?

– Oui.

– Eh bien... Je... Je pense que je l'ai toujours redouté. Il se montrait affable, il m'offrait des cadeaux, il m'emmenait en promenade. Pourtant...

– Pourtant ? fis-je me penchant légèrement et la fixant intensément.

– Je le sentais plein d'un mépris inhumain. J'avais l'impression qu'il se moquait de moi comme de nous tous. (M. del Castillo, *La Nuit du décret*, 1981, 144)

**(iii) Le rapport au contenu propositionnel**

Comme cela a été rappelé dans les propriétés générales sur l'intégration des MD au contenu sémantique (section 3.1.5 (i), p. 114), le contenu propositionnel renvoie au contenu sur lequel il est possible d'enchaîner et dont on peut, hors contexte, suspendre ou inverser la valeur de vérité par une question, une négation ou une structure hypothétique<sup>180</sup>. Un élément contribuant au contenu propositionnel peut en outre être clivé, c'est-à-dire inséré dans une structure en *c'est... que*. Les PEN ne font jamais partie du contenu propositionnel, c'est une de leur propriété sémantique, alors que certains connecteurs peuvent y être intégrés, comme la proposition introduite par *parce que* en (85a)<sup>181</sup>. Cela vaudrait aussi pour les connecteurs *encore*, *après (que)*, *avant (que)*, *ensuite*, etc.

(85) a. Louise est rentrée chez elle plus tôt parce qu'elle était fiévreuse.

b. Est-ce que Louise est rentrée chez elle parce qu'elle était fiévreuse ?

c. Louise n'est pas rentrée chez elle plus tôt parce qu'elle était fiévreuse.

d. C'est parce qu'elle était fiévreuse que Louise est rentrée chez elle plus tôt.

Les connecteurs ont donc un fonctionnement plus souple et leur contribution ou non au contenu propositionnel ne peut pas être un critère strict de leur définition. Pour simplifier, on dira que les connecteurs sont associés à une relation de discours entre deux éléments formulables de façon propositionnelle mais pas forcément explicite et

180 Cela rejoint des propriétés du *posé*, lorsqu'on l'oppose au *présupposé* chez Ducrot (1972, p. 22-23), *truth-conditional content* chez Chierchia & MacConell-Ginet (1990), ou le *what is said* ou le *groundfloor statement* chez Grice (1989 [1975], p. 22-40 et 1989, p. 362).

181 Voir aussi ce qui a été dit de *parce que* à la note 172, p. 115 et des exemples (63) et (64), p. 115.

qu'ils ne sont pas arrimés à l'énonciation. Les PEN, elles, sont nécessairement arrimées à l'énonciation.

(iv) *Comportement face aux tests pour les présuppositions*

Les éléments développés ici sont extraits de Dagnat (2020f [38], p. 8-10) qui porte précisément sur le traitement des PEN dans la perspective des présuppositions, en particulier du point de vue de la propriété de projection.

Le fait que les PEN soient nécessairement accrochées à l'énonciation (le *hic et nunc* et les acteurs de cette énonciation, notamment le locuteur) a comme conséquence qu'elles se projettent toujours. C'est-à-dire que les présuppositions qu'elles déclenchent résistent aux opérations de négation, de question et d'insertion dans une construction hypothétique. Ce qui a été noté pour les pronoms dans l'exemple (66) (voir aussi note 176, p. 116) et pour les verbes, les adverbes et les connecteurs dans l'exemple (67) (p. 117), vaut aussi pour les PEN. Prenons l'exemple de *tu m'étonnes* en (86) :

- (86) a. L1 – Paul est venu à la réunion.  
 b. L2 – Tu m'étonnes !  
 c. L2 – Je suis très étonné.  
 d. L2 – Tu es sûr ? Tu ne m'étonnes pas. (d'après Dagnat 2020f [38], ex. (15) p. 8)

La réponse de L2 en (86b) ne peut subir aucune modification. *Tu m'étonnes* fonctionne ici comme une PEN qui pointe vers l'assertion de L1, c'est son caractère indexical, et manifeste que la présence de Paul constitue, pour L2, un événement totalement attendu. C'est le contenu sémantique conventionnellement reconnu à *tu m'étonnes* dans cet emploi. Ce contenu est clairement figé, puisqu'il signifie le contraire de ce que pourrait constituer le contenu littéral de *tu m'étonnes* (= « je suis surpris ») exprimé par L2 dans la réponse (86c)<sup>182</sup>. La manipulation effectuée dans la réponse de L2 en (86d)<sup>183</sup> montre que la présupposition déclenchée par la PEN *tu m'étonnes* en (86a) (= « je ne suis pas surpris ») ne peut pas être affectée par la négation. Si c'était le cas, *tu ne m'étonnes pas* en (86d) devrait signifier la négation de « je ne suis pas surpris » (autrement dit, « il n'est pas vrai que je ne suis pas surpris »), mais ce n'est pas ce qui passe. En (86d), L2 dit simplement qu'il n'est pas étonné.

On remarque aussi que les PEN n'ont jamais d'effet local. On pourrait appeler cela leur *indexicalité restreinte*. Pour rappel, un déclencheur de présupposition provoque un effet local lorsque, quand on l'enchâsse dans un opérateur de croyance, typiquement des verbes comme *penser, estimer, croire*, la présupposition déclenchée est perçue comme « embarquée » dans la portée de cet opérateur. Autrement dit, elle sera attribuée au sujet de l'opérateur (du verbe). Les manipulations de (87a) montrent que dans (87b) et (87c) les PEN sont arrimées aux croyances du locuteur avec différentes valeurs (surprise, acceptation, évidence, etc.). Dans (87e) et (87f), on voit que ces valeurs ne sont pas transférables aux croyances de l'énonciateur (Marie) dont le locuteur rapporte les propos, comme est transférable le contenu propositionnel « la boulangerie est fermée ».

182 Une des explications possibles de la valeur antiphrastique de la PEN par rapport au sens littéral pourrait être un mécanisme d'ironie. Dans ce cas-là, il me semble que la prosodie et les mimiques seraient de bons indicateurs à l'oral.

183 Le contexte de doute exprimé par *Tu es sûr ?* permet de positionner L2 en (87d) comme n'adhérant pas spontanément à l'assertion de L1 en (87a) et comme pouvant donc être surpris par cette assertion.

Ces valeurs collent à l'énonciation en cours, celle du locuteur. Dans (87e), il n'est même pas sûr que les énoncés produits soient tous acceptables, ou alors avec une prosodie ou une typographie (tirets, parenthèses) marquant un commentaire parenthétique.

- (87) a. La boulangerie est fermée.  
 b. Mais/ah/zut/bon/tiens/sans blague (!) la boulangerie est fermée.  
 c. La boulangerie est fermée, quoi.  
 d. Marie croit que la boulangerie est fermée.  
 e. % Marie croit que mais/ah/zut/bon/tiens/sans blague (!) la boulangerie est fermée.  
 f. Marie croit que la boulangerie est fermée, quoi.

Enfin, j'ai rapproché cette absence d'effet local de la notion de *non-déplaçabilité* reprise par Potts (2007) à propos des expressifs<sup>184</sup>:

Expressives cannot (outside of direct quotation) be used to report on past events, attitudes, or emotions, nor can they express mere possibilities, conjectures, or suppositions. They always tell us something about the utterance situation itself. This is the nondisplaceability property. (Potts 2007, p. 169)

À l'origine, la notion de déplaçabilité a été définie d'un point de vue anthropologique sur la capacité référentielle du langage. En cherchant à caractériser l'origine et les spécificités du langage humain, Hockett (1960, p. 91) a dégagé treize caractéristiques de la communication animale et huit systèmes de communication. La déplaçabilité ressort comme une des spécificités du langage humain. Il s'agit de la capacité du langage à renvoyer à un référent absent de la situation spatiale et ou temporelle de la communication en cours<sup>185</sup>. La non-déplaçabilité renvoie donc à un processus référentiel « moins évolué », cantonné à l'immédiateté et présentant des similarités avec la communication non-verbale basée sur l'ostension et l'accès immédiat aux référents. C'est précisément ce point qui est repris plus tard chez Potts (2007, p. 169-173) et chez Cruse (1986, p. 270-277, 2011, p. 200-2002) à propos du mode sémantique expressif ou des expressifs, et chez Gutzmann (2013, p. 37-40) à propos des expressions dépendantes du contexte (*use-conditional expressions*).

Les MD pris dans leur ensemble ne réagissent pas tous de la même manière et la non-déplaçabilité apparaît comme une propriété « en dur » des PEN. Les PEN sont souvent présentes en filigrane dans les études sur les expressifs à travers quelques exemples d'interjections et de jurons, mais les remarques sur la déplaçabilité s'appliquent en fait à toutes les catégories de PEN de la Figure 20 (ici p. 120) et pas seulement à celles qui véhiculent une variation émotionnelle. L'exemple (88) servira d'illustration.

- (88) a. Paul a raté son examen d'histoire hier.

184 Les expressions appelées *expressifs*, ou fonctionnant selon le mode expressif, sont présentées non seulement comme s'appliquant de façon concomitante à leur énonciation (pas dans le passé ni dans le futur de cette énonciation : elles ne peuvent pas être rapportées ni anticipées), mais également comme liées à la situation d'énonciation elle-même. Les expressifs désignent en fait plusieurs phénomènes marquant la subjectivité.

185 « Man is apparently almost unique in being able to talk about things that are remote in space or time (or both) from where the talking goes on. » (Hockett 1960, p. 90). Cette capacité correspond au mode symbolique de Bühler, chez qui les signes symboliques sont qualifiés de « détachables » par A. Eschbach (Bühler 2011 [1934], p. 513) et mode propositionnel chez Cruse : « Another characteristic distinguishing expressive meaning from propositional meaning is that it is valid only for the utterer, at the time and place of utterance. » (1986, p. 272)

- b. Bon/Tiens/Zut Paul a raté son examen d'histoire hier.
- c. # Bon/Tiens/Zut Néron a incendié Rome !
- d. Ce satané Néron a incendié Rome en 64 après J.C.

En (88a), le locuteur rapporte un événement passé, mais les PEN de (88b) ne portent pas directement sur cet événement. On s'en rend compte en utilisant un événement temporellement lointain, comme en (88c). La phrase (88c) est étrange. Pour améliorer son interprétation, il faut supposer que le locuteur apprend que Néron a incendié Rome, que sa réaction ne porte donc pas directement sur l'événement mais sur le fait qu'il en prend conscience (événement interne).

L'exemple (88d) me sert à montrer que les PEN sont plus restreintes que les expressifs, illustrés ici l'adjectif *satané*.

Bien que l'usage d'un adjectif expressif comme *fichu*, *foutu*, *satané* ne puisse être attribué qu'au locuteur et donc, dans cette mesure, verrouillé à l'énonciation, rien ne l'oblige à porter sur une entité immédiatement appréhendable ni même récente, à la différence des PEN. Il y a donc une différence avec les PEN dont la portée est restreinte à des entités (individus ou événements) contemporains de la situation d'énonciation.

En résumé, j'ai cherché à approfondir et à compléter des propriétés généralement avancées pour caractériser les MD et pour différencier le fonctionnement comme connecteur et le fonctionnement comme PEN. J'ai questionné le caractère indexical et procédural des MD en général, j'ai insisté sur le rapprochement avec les propriétés des présuppositions et des expressifs pour mettre en évidence que les PEN avaient un fonctionnement plus « strict », plus restreint que les connecteurs. En tant que MD, elles sont indexicales et fonctionnement de manière procédurale. En tant que PEN, elles ne marquent pas une relation de discours, mais servent à présenter leur référent sous un certain angle et cette perspective est souvent difficile à formuler, elles se projettent obligatoirement, n'ont jamais d'effet local et ne sont jamais déplaçables, c'est-à-dire qu'elles renvoient nécessairement à un référent concomitant de leur énonciation.

### 3.2.1.2 (Im)possibilité d'une liste

Cartographier les MD repose sur l'identification de critères (propriétés et tests) qui permettent d'en pointer les fonctionnements généraux et spécifiques. C'est aussi, dans la mesure du possible, en dresser une liste. Les deux approches se complètent, puisqu'en général on extrait des propriétés de l'usage d'items que l'on estime être de bons candidats MD et on interroge l'intégration d'autres items moins fréquents ou très polyfonctionnels sur la base des propriétés que l'on retient. C'est toute l'ambiguïté de la définition d'une catégorie.

Dans un premier temps, je me suis demandé s'il était possible, sur la base des propriétés mentionnées, de dresser la liste de tous les MD du français, en distinguant systématiquement connecteurs et PEN (i). Ceci amène à s'interroger sur une catégorie reconnue par la grammaire, l'interjection, et sur celle des particules modales, bien étudiées pour l'allemand mais dont l'existence en français est discutée (ii).

#### (i) Quels sont les éléments listés comme MD ?

Dans la mesure où les MD, connecteurs et PEN, sont présentées comme des catégories fonctionnelles, on n'en trouve pas de recensement dans les grammaires traditionnelles. Les travaux qui les mentionnent portent en général sur un MD ou quelques MD, et il

est rare de rencontrer une vision d'ensemble qui cherche à pointer l'ensemble des formes. Je rappelle ci-dessous quelques approches qui ont produit des listes sur lesquelles je me suis appuyée, en particulier pour les connecteurs (a et b). J'insiste sur ma démarche pour produire une liste des PEN les plus fréquentes en français (c) et sur l'intégration de toutes ces données dans le dictionnaire du logiciel Unitex/GramLab<sup>188</sup> dans le cadre d'un suivi de stage (d).

### (a) MD en général

Chanet propose une liste commune de 85 formes (2003, p. 15-16) sur la base d'un relevé dans le corpus CORPAIX et le Corpus de Référence du Français Parlé (CRFP). Même si les analyses quantitatives sont limitées, cela permet d'avoir une vision de leur présence et de leur distribution dans les données.

Lors de mes premières analyses sur le français québécois, j'avais listé les éléments fonctionnant comme MD dans les deux corpus utilisés, *fcapop* et *tremblay* (voir section 1.3.1.1, p. 37). J'avais alors abouti à la liste suivante :

à part de (t) ça, ah, (ça) fait que, après, aussi, ben, bien, bon, donc, certain, coudonc, déjà, disons, donc, écoute, en effet, en fait, en tout cas, entéka, ensuite, euh, hein, je pense, je sais pas, tu sais, je veux dire, tu sais pour dire, là, mais, ouais, par contre, par exemple, peut-être, pis, puis, pis tout, quand même, quoi, dis/dites donc, voyons, voyons donc, écoute/écoutez (donc), ha, aïe, aye, ayoye, eh, hé, euh, hein, hon, hm, humhum, mmm, mmmh, oh, bah, ouf(ff). (Dagnat 2006 [6], p. 259)

À l'époque, je n'avais donc pas différencié connecteurs et PEN et j'avais intégré les formes orthographiques normées et des néographies phonétisantes conventionnelles. De plus, la liste n'était pas bien organisée et mettait sur le même plan des phénomènes qui auraient pu être distingués plus finement. Ce premier contact avec le « monde » des MD m'a été utile pour comprendre l'importance de tels éléments dans la communication orale et pour sa représentation et la difficulté de circonscrire leur sens en contexte.

### (b) Connecteurs

On trouve assez facilement des listes de connecteurs (souvent dits *logiques*) dans des manuels et des sites méthodologiques, mais ces listes sont parcellaires et souvent intuitives. Il existe une liste des connecteurs du français, LEXCONN, qui résulte du travail de Roze (2009) et de l'équipe autour du French Discourse Treebank (Roze et al. 2012). La liste proposée dans Roze (2009, p. 67-77) est produite sur la base de ressources diverses. On dénombre plus de 320 formes différentes. L'arrière-plan théorique est celui de la SDRT, ce qui explique pourquoi les connecteurs sont présentés en fonction de leur catégorie syntaxique (adverbe, conjonction, préposition, etc.), du type (relation discursivement subordonnante ou coordonnante) et du sens (Result, Contrast, Condition, etc.) de la relation de discours qu'ilsinstancient. L'intérêt de cette liste est sa richesse. D'une part, les connecteurs polyfonctionnels sont mis en évidence (ex. *alors* et *pendant que*), d'autre part, les formes composées sont prises en compte (ex. *quand (bien) même, du coup, en d'autres termes, la preuve*) alors qu'elles restent problématiques dans les corpus annotés, qui les décomposent. Le Tableau 9 illustre quelques cas.

Connecteur	Catégorie	Type	Relation
------------	-----------	------	----------



alors	adv	coord	Consequence
alors	adv	coord	Narration
alors	adv	coord	Result
pendant que	conj. sub.	coord	Contrast
pendant que	conj. sub.	sub	Background
quand bien même	conj. sub.	sub	Concession

Tableau 9. Description des connecteurs *alors*, *pendant que* et *quand bien même* (Roze 2009, p. 67)**(c) PEN**

Il est plus difficile de trouver des listes globales de PEN ou équivalent. Le cadre lexicographique franco-allemand de Métrich et al. (1988-2002) permet un bon état des lieux de ces « invariables difficiles », mais dans une perspective contrastive dont les choix sont orientés de l'allemand vers le français. La rareté et la difficulté des listes de PEN peuvent s'expliquer pour plusieurs raisons. D'abord parce que leur catégorisation est loin d'être uniforme et que l'intégration de tel ou tel item dans la famille PEN dépend des propriétés retenues par les approches. Ensuite, parce que les PEN sont plus sujettes à la variation que les connecteurs. Cette grande variabilité vient du fait qu'elles sont plus fréquentes à l'oral, qui est un mode de communication plus immédiat (Koch & Oesterreicher 2001). Elle vient aussi du fait que leur indexicalité est centrée sur le locuteur les énonçant et sur le contexte spatio-temporel et cognitif concomitant de leur énonciation. Ceci les rend fondamentalement plus expressives et plus spontanées que les connecteurs et donc très sensibles à toutes les dimensions variationnelles (diachronie, diatopie, diastratie et diaphasie).

Dans le cadre de la *Grande Grammaire du Français* (Dagnat 2021a [39], p. 2018-2019), il m'a été demandé de produire une liste des particules les plus fréquentes en français standard en pointant quelques variations régionales. Une autre contrainte a été d'intégrer les interjections au chapitre tout en leur conservant le statut de catégorie grammaticale à part entière. Ce qui n'est pas sans soulever un certain nombre de questions. Dans l'*Encyclopédie Grammaticale du Français*, en revanche, j'ai choisi de traiter les interjections comme une sous-partie des PEN. Dans la Figure 20 (ici p. 120), elles correspondent en effet à une des sous-catégories de Dostie, celle qui sont liées à la réalisation d'un acte illocutoire. Dans les deux cas, *Grande Grammaire* ou *Encyclopédie*, la liste de départ s'est considérablement agrandie et a été organisée. Le Tableau 10 ci-dessous reprend l'essentiel des tableaux produits pour la *Grande Grammaire*, en fusionnant celui sur les interjections et celui sur les « autres particules », en modifiant quelques paramètres et en ajoutant quelques expressions. La première colonne donne la catégorie prototypique des autres emplois de l'item dans la langue, la deuxième colonne liste les items fonctionnant comme PEN, en distinguant des formes simples et des formes complexes sans en discuter la (non-)compositionnalité.

Catégories des autres emplois dans la langue	PEN
Adjectif	<i>formes simples</i> : super, chic, chiche, chouette, bon <i>formes complexes</i> : tout beau, tout doux
Adverbe	<i>formes simples</i> : alors, assez, ben, bien, comment, donc, effectivement, enfin, hélas, là, non, oui, (m)ouais, yes

	<i>formes complexes</i> : jamais de la vie, mais enfin, non mais, non mais enfin, non mais quand même, quand même, tant pis, tant mieux
Conjonction	Mais [donc est classé dans les adverbes]
Interjection	amen, allô, aïe, ah, aha, ahou, bah, beh, bernique, beuh, beurk, bof, bravo, chut, crac, eh, euh, eurêka, fi, ha, haha, hé, hein, hem, hep, heu, hip hip hip hourra, ho, holà, hop, hourra, hum, na, oh, ohé, ouah, ouf, ouh, ouille, oups, ouste, paf, patatras, peuh, pouh, pouah, pouët (pouët), prout, pst, sniff (snif), ta ta ta, tagada, taratata, tchin (tchin), tintin, tralala, waouh, youp là, youpi, zou, zut
Nom	<i>formes simples</i> : adieu, attention, banco, bingo, bon- jour, bonsoir, bonté, bordel, bougre, camembert, catastrophe, ciel, chapeau, chiotte, Christ, con, courage, crotte, dame, diable, dommage, grâce, flûte, halte, malheur, merci, merde, misère, Jésus, pardon, pétard, pitié, punaise, purée, putain, quoi, respect, salut, santé, seigneur, stop, tonnerre <i>formes complexes</i> : nom de Dieu, ça alors, ça par exemple
Pronom	quoi
Syntagme nominal	<i>formes simples</i> : de l'air, du balai, du calme, la vache, ma foi, ma parole, mon Dieu, mon salaud, ta/vos gueules <i>formes complexes</i> : Bon Dieu, bonne journée, bonne soirée, bonté divine, bon sang, chapeau bas, Dieu du ciel, doux Jésus, grand(s) dieu(x), juste ciel, point barre
Syntagme prépositionnel	à bientôt, à la bonne heure, à la fin, à la revoyure, à la tienne/nôtre/vôtre, à plus, à tantôt, au revoir, au plaisir, d'accord, en fait, par exemple
Verbe, construction verbale	<i>impératif</i> : admettons, allons, allez, dis, dites, disons, écoute, écoutez, ferme-la, fermez-la, la ferme, figure-toi, figurez-vous, regarde, regardez, remarque, remarquez, tiens, tenez, va, voilà, voyez, voyons <sup>186</sup> <i>indicatif</i> : tu comprends, vous comprenez, comprends-tu, comprenez-vous, tu penses, penses-tu, pensez-vous, vous pensez, tu parles, tu m'étonnes, tu rigoles, tu sais, vous savez, tu vois, vous voyez <i>subjonctif</i> : soit <i>infinitif</i> : foutre [voir aussi comme nom]
Fragment, phrase	c'est ça, je veux dire, je dis ça je dis rien, je veux mon neveu, qu'est-ce que tu veux, que veux-tu, qu'est-ce que vous voulez, que voulez-vous, n'est-ce pas, si tu veux, si vous voulez et la marmotte, et ta mère, et ta sœur, et puis quoi encore, et mon cul c'est du poulet, etc.

Tableau 10. Principales expressions fonctionnant comme PEN en français

Bien que cette démarche ait été stimulante, je ne suis pas encore complètement satisfaite du résultat. Il faudrait ajouter des éléments sur la variation, les emprunts (*shit, basta, olé, yes*, etc.) et des altérations diverses visant à atténuer les particules relevant des jurons et blasphèmes (*diantre, fichtre, fouchtra, tabarnouche, vinzou*, etc.) (voir Dagnat 2021a [39], p. 2017 et 2022-2023). L'histoire et l'usage ont figé certaines formes complexes et il n'est pas toujours facile de trancher, notamment entre les syntagmes prépositionnels et les adverbes (*d'accord, en fait, par exemple*, etc.). La ligne « verbe et construction verbale » n'est pas uniforme : j'y mentionne des impératifs et des indicatifs qui pourraient tout aussi bien être déclarés comme « phrases » pour la grande majorité dans la mesure où la construction est saturée. La ligne « fragments et phrases » mériterait aussi d'être détaillée. S'y trouvent des segments prédicatifs

186 La plupart de ces constructions à l'impératif peuvent être combinées avec *don(c)* ou *voir* : *dis donc, dis voir, tiens donc, tiens voir*, etc.

figés divers dont le sens, très culturel pour certains, est devenu assez commun. Ils pourraient d'ailleurs être qualifiés de phrasèmes, c'est-à-dire d'énoncés multilexématique non libres (Mel'čuk 2011). Morphologiquement, leur statut de PEN peut être discuté en fonction de leur degré de figement, observable en particulier dans leur réalisation phonétique et du caractère compositionnel ou non de leur association.

**(d)** Liste commune et analyse de corpus

Posséder une liste des MD la moins incomplète possible présente plusieurs intérêts. Pour mes recherches, j'y vois en particulier la possibilité de construire des requêtes pour analyser finement leurs fonctions dans des corpus de grande taille et, dans l'idéal, la possibilité de leur annotation dans des données brutes. Posséder une liste mêlant connecteurs et PEN présente aussi l'intérêt de prendre en compte la polyfonctionnalité de certains items dans l'usage sans avoir à choisir telle ou telle fonction a priori. Enfin, une liste commune permet d'interroger des données et d'analyser quantitativement et qualitativement les combinaisons contiguës et non contiguës de MD (connecteurs entre eux, PEN entre elles et connecteurs et PEN). C'est précisément ce que j'ai commencé à faire dans Dagnat & Jayez (2021b [40]) et Dagnat (2022a [42]) avec les associés de *mais*, et ce qui constitue le sujet central du projet CODIM (voir section 4.1.1, p. 159).

Dans le cadre du suivi de stage de A. Hafd Dahou et du projet PROSOMDI & CO<sup>187</sup>, nous avons utilisé l'interface Unitex/GramLab<sup>188</sup> en enrichissant son dictionnaire au moyen de la liste des connecteurs de Roze (2009) et de la liste des PEN du Tableau 10. L'objectif était d'extraire les combinaisons de MD les plus fréquentes en français dans un ensemble de données écrites et orales assez important<sup>189</sup>. La démarche était exploratoire et les résultats limités mais cela a permis de débroussailler le domaine et de pointer quelques problèmes. Je renvoie à Hafd Dahou (2021) pour l'exposé complet. Cette collaboration a été constructive pour moi dans la mesure où elle m'a familiarisée avec l'interface d'Unitex/GramLab et avec la construction d'automates obligeant à formuler des contraintes sur les items recherchés. Elle a aussi permis de mettre en place un puits de données assez diversifiées et d'aboutir à une liste conséquente de MD qui dépasse largement les 85 formes identifiées par Chanet et que je compte utiliser pour la suite de mes travaux.

**(ii)** *Que faire des interjections et des particules modales ?*

La littérature sur les MD est touffue et propose des catégories et des termes à géométrie variable. Néanmoins, deux catégories auxquelles il est fréquemment fait référence ont un relief particulier car elles ont droit à un statut grammatical : il s'agit des interjections et des particules modales. Comment sont-elles définies ? Doit-on les considérer comme des sous-catégories de MD, des sous-catégories de PEN ? Quelles sont les formes concernées ?

**(a)** Les interjections

187 PROSOMDI & CO : PROSodie des Marqueurs Discursifs et COrpus. Projet que j'ai coordonné et qui a été financé par l'ATILF, le Pôle scientifique CLCS de l'Université de Lorraine et Lorraine Université d'Excellence (OLKI) (2021).

188 Unitex/GramLab, suite logicielle pour l'analyse de corpus : <https://unitexgramlab.org/fr>

189 Corpus utilisés : ESLO1, *Le Monde*, Frantext, CORPAIX, DÉCLICS, Fra80.

J'ai considéré les interjections comme des PEN car, dans la classification que j'ai adoptée (voir Figure 20, ici p. 120), leur fonction correspond à la fonction de réalisation d'un acte illocutoire telle que définie par Dostie :

Ils [les marqueurs] ont la possibilité d'accomplir un acte illocutoire, le plus souvent expressif ou directif, parfois assertif. Il s'agit de mots-phrases ou de mots associés à un SN/une proposition traduisant l'état psychologique de l'énonciateur. (Dostie 2004 : p. 47)

On aurait pu penser qu'établie en classe grammaticale et reconnue dans les dictionnaires depuis longtemps, l'interjection serait relativement bien définie et circonscrite, comme les autres catégories grammaticales. En réalité, son traitement n'est pas uniforme et ses contours pas si nets. Elle apparaît à la fois comme une classe de mots et comme un usage (voir Améka 1992, p. 102, 2006, Hansen 1998a, p. 38). Elle correspond chez certains à des critères très restreints et renvoie alors aux lexèmes invariables ne fonctionnant dans la langue *que* comme PEN et pouvant former un énoncé à eux seuls. La restriction peut aussi porter sur l'interprétation, l'interjection désigne alors uniquement les lexèmes ayant les propriétés des PEN et exprimant l'état émotionnel du locuteur. Chez d'autres encore, la classe est nettement plus accueillante et recouvre l'ensemble des PEN désignées ci-dessus dans le Tableau 10 et l'ensemble de leurs fonctions. L'interjection renvoie alors à tous les lexèmes pouvant fonctionner ainsi, qu'ils aient ou non d'autres emplois dans la langue<sup>190</sup>.

Dans *La Grande Grammaire* (Dagnat 2021a [39]), c'est d'abord la définition syntaxique restrictive qui a été retenue, c'est-à-dire que les interjections y sont des lexèmes invariables présentant les propriétés des PEN, ne relevant d'aucune autre classe syntaxique et pouvant former un énoncé à elles seules. En revanche, une interjection peut servir à autre chose qu'à l'expression des émotions du locuteur (*allô, tchin-tchin, hein, heu,...*). Pour les raisons évoquées plus haut au sujet des PEN en général (indexicalité restreinte, oralité, spontanéité et expressivité), les interjections, en tant que PEN, sont très sujettes à la variation – elles en sont même un excellent traceur –, ce qui explique l'impossibilité d'en fournir une liste exhaustive. De plus, piochant facilement dans la vulgarité, les blasphèmes et les jurons, elles développent aussi des formes modifiées adoucies, sortes d'euphémismes tout aussi expressifs mais socialement plus acceptables (*diantre* <diable, *parbleu* <par Dieu, *tabarnouche* <tabernacle, etc.).

#### (b) Les particules modales

La question des particules modales est plus délicate, car il s'agit d'une catégorie grammaticale bien étudiée pour les langues germaniques, en particulier l'allemand, mais dont la pertinence est discutée pour le français (Abraham 1991, Diewald 2013, p. 27, Hansen 1998a, p. 41-46, 1998b, p. 135, Métrich et al. 1998-2002, p. 24-36, Schoonjans 2013, p. 155-157, Waltereit 2001, 2006).

La liste des particules modales est relativement restreinte et concerne des mots brefs. Pour l'allemand, on compte une quarantaine de mots (voir la comparaison de Diewald 2013, p. 27 et de Schoonjans 2013, p. 140), par exemple *aber, auch, doch, eben, halt, ja, nur*. Parmi la dizaine de propriétés attribuées à cette classe, certaines correspondent à celles que j'ai rappelées ou mises en évidence pour les PEN. D'autres

190 Le fait de séparer les formes n'existant que comme interjection (*ah, oh, youpi, etc.*) et celles empruntées à d'autres catégories grammaticales (*chic, la vache, merde, bon, etc.*) rejoint la distinction entre *interjections primaires* et *interjections secondaires* (Ameka 1992, p. 105 et 111).

diffèrent, en particulier du point de vue syntaxique et du point de vue de la portée. Les particules modales sont syntaxiquement et prosodiquement/graphiquement intégrées à leur proposition-hôte. Elles sont aussi présentées comme obligatoirement positionnées dans le champ médian (*middle field*) de la phrase et, à ce titre, fonctionnent comme des adverbes<sup>191</sup>. Enfin, elles sont attachées à l'acte de langage de la phrase dans laquelle elles se trouvent et portent sur l'ensemble de la phrase.

Leur signification est généralement abordée dans une perspective large et interactionnelle<sup>192</sup> de la modalité en termes d'attitudes épistémiques ou évaluatives du locuteur envers son énoncé. Métrich et al. (1998-2002) isolent six fonctions communicatives dont « l'indication ou la modification de la valeur illocutoire des énoncés » (p. 31) et « l'expression de l'attitude du locuteur vis-à-vis du contenu de son énoncé ou de l'interlocuteur » (p. 33). Ces fonctions rappellent les marqueurs à portée illocutoire de Dostie, en particulier les PEN donnant des instructions sur l'interprétation de l'acte illocutoire (voir ici Figure 20, p. 120 et l'exemple (77), p. 122).

Les études<sup>193</sup> qui font une place au français proposent une liste d'items bien plus limitée que pour l'allemand. On trouve principalement : *bien, déjà, quand même, donc, peut-être, seulement, tout de même*, illustrés pour certains par les exemples suivants :

(89) Les diamantaires ne semblent pas préoccupés par la progression des rebelles zairois [titre] / Ils sont quand même curieux, ces diamantaires : à première vue, l'offensive des troupes rebelles [...] ne les émeut guère. « Nous suivons la situation de très près », se borne à déclarer un porte-parole de la Central Selling Organisation. (*Le Monde* 15.3.1997 dans Waltereit 2001, ex. (13a) p. 1400)

(90) Si j'avais seulement pu le consulter avant de rencontrer Reslaut ! (C. Mauriac, *La Marquise sortit à cinq heures*, 1961, dans Schoonjans 2013, ex. 20 p. 155)

(91) Regarde donc où tu vas ! (dans Hansen 1998a, ex. 6 p. 41)

(92) C'est bien la première fois que ça m'arrive ! (Hansen 1998b, dans Detges & Waltereit 2016, ex. 12 p. 647)

Globalement, dans (89), *quand même* sert à marquer une situation comme étant inattendue (« les diamantaires ne s'émeuvent guère ») par rapport à une situation présentée comme « normale » (« ils devraient être préoccupés »). *Quand même* introduit aussi une part d'expressivité qui traduit à la fois une évaluation (le locuteur trouve cela « curieux », étrange) et peut-être une surprise du locuteur sur le caractère inattendu de la situation<sup>194</sup>.

191 Le champ médian correspond en allemand à la place entre une forme conjuguée et une forme verbale non conjuguée ou à la place entre le mot subordonnant et le groupe verbal dans les subordonnées. L'idée est que les particules modales ont une place fixe en allemand, celle dite de la position médiane. Mais cette propriété est discutée par Schoonjans (2013, p. 136 et p. 155) qui montre, d'une part, que cela est applicable au français et, d'autre part, que ce n'est pas toujours vrai.

192 « La signalisation de ces attitudes se justifie dans le cadre d'une conception fondamentalement interactionnelle du langage. Ainsi, l'échelle des modalités est redéfinie en termes communicatifs, comme indiquant si, et dans quelle mesure, le locuteur s'attend à ce que la factualité de ses propos soit remise en cause par l'allocutaire, et, du même coup, si ces propos ont besoin d'une justification évidentielle. » (Hansen 1998b, p. 133)

193 Pour des analyses focalisées sur certaines particules, voir Schoonjans (2013, p. 155-158), Waltereit (2001, p. 1400-1405) et (Hansen 1998b).

194 On peut s'interroger sur la présence de *même* qui introduit une scalarité dont on pourrait dériver une valeur intensive à l'évaluation. Ceci reviendrait à dire que le locuteur trouve les diamantaires *très* curieux (= « même avec un seuil de bizarrerie très élevé, les diamantaires

Dans (90) et dans (91), *seulement* et *donc* servent à renforcer l'acte de langage associé à leur proposition hôte, un souhait dans le premier cas, un ordre dans le second.

Dans (92), *bien* fonctionne de manière polyphonique. Il sert à renforcer une assertion (« c'est la première fois que ça m'arrive ») présentée comme réfutant le proposition complémentaire négative (« ce n'est pas la première fois que ça m'arrive »). Cette proposition complémentaire peut être explicite ou implicite, être le fait du locuteur ou de son interlocuteur.

Ainsi formulées, les fonctions de *quand même*, *seulement*, *donc* et *bien* entrent effectivement dans le champ de certaines des fonctions avancées par Dostie que j'ai reprises. Mais les items sont plus contraints syntaxiquement, moins mobiles que des PEN comme *tiens*, *bon*, *ah*, *disons* et liés par constructions (*si seulement*, *c'est bien*, verbe à l'impératif + *donc*). Du point de vue syntaxique, les particules dites modales semblent effectivement se comporter comme des adverbes.

Qu'en est-il de la projection, de l'effet local et de la déplaçabilité ? Je rappelle que les PEN ont été décrites comme se projetant toujours, sans effet local et non-déplaçables. Pour des raisons syntaxiques, les exemples (89) à (92) sont compliqués à tester car ils présentent des contraintes constructionnelles importantes, ce pourquoi je pars d'exemples plus simples et n'illustre que quelques tests. L'exemple et ses manipulations dans (93) qui contient la PEN *disons* sert de contraste avec l'exemple (94) qui présente une valeur modale de *quand même*.

- (93) a. Paul se lève à sept heures  
 b. Paul se lève disons à sept heures  
 ⇒ Paul se lève je dirais en gros à 7 heures  
 c. il n'est pas vrai que Paul se lève disons à sept heures  
 ⇒ il n'est pas vrai que Paul se lève je dirais en gros à 7 heures

- (94) a. Paul est intelligent.  
 b. Paul est quand même intelligent.  
 ⇒ Quoiqu'on dise, ou de quelque manière qu'on fixe le seuil d'intelligence, Paul reste intelligent<sup>195</sup>.  
 c. Il n'est pas vrai que Paul est quand même intelligent.  
 ⇒ Il n'est pas vrai que quoiqu'on dise, ou de quelque manière qu'on fixe le seuil d'intelligence, Paul reste intelligent.

– *Projection*. J'utilise le test de la négation<sup>196</sup>. On dira qu'il y a projection quand la proposition évoquée par *disons* ou *quand même* ne peut pas entrer dans la portée de la négation. Dans (93b), la proposition évoquée est « je dirais en gros » et elle n'est pas atteinte par la négation en (93b). Dans (94b), sans contexte plus précis, la proposition évoquée est « quoiqu'on dise, ou de quelque manière qu'on fixe le seuil d'intelligence ». Si en (94c) on veut garder la valeur modale de *quand même*, il faut le déplacer et l'intégrer au niveau de l'acte de langage de la phrase (l'assertion *il n'est pas vrai que*), c'est sa place de particule modale. Cela donnerait : *il n'est quand même pas vrai que Paul est intelligent*. Il n'est pas possible d'étendre la portée de *quand même* sans le déplacer (voir la contrainte

l'atteignent ou le dépassent »).

195 Il est possible de percevoir un effet intensif dans la mesure où l'intelligence semble résister à tous les contre-arguments auxquels le locuteur a accès.

196 Je laisse de côté les tests avec la question et l'insertion dans une structure hypothétique, car trop de paramètres interviennent et je n'arrive pas à analyser l'acceptabilité des résultats.

du champ médian). Dans (94c), *quand même* avec une valeur modale ne peut pas être atteint par la négation, autrement dit se projette.

– *Effet local*. Le test de l'effet local consiste à enchâsser l'élément à tester dans une proposition du type *X pense/estime/croit que*. Ce qui donne :

(93) d. Marie pense/estime/croit que Paul se lève disons à sept heures.

(94) d. Marie pense/estime/croit que Paul est quand même intelligent.

Dans (93d), *disons* reste une approximation du locuteur et n'a donc pas d'effet local. Dans (94d), la valeur modale de *quand même* est préférentiellement attribuée à Marie, il a donc un effet local.

– *Déplaçabilité*. Il s'agit de tester ici si *disons* et *quand même* peuvent s'ancrer sur des périodes temporellement distantes (voir ex. 88, p. 125).

(93e) A l'époque, Paul se levait disons à sept heures.

(94e) A l'époque, Paul était quand même intelligent.

Dans les deux cas, le contenu associé au marqueur concerne le moment de l'énonciation.

J'en conclus provisoirement que, pour *quand même* en tout cas, la situation est assez hybride. Sémantiquement, le fonctionnement est proche de celui des PEN, mais les contraintes de composition sont plus strictes : sémantiquement, *quand même* porte sur l'acte de langage de la phrase dans laquelle il est inséré. S'il figure dans une subordonnée complétive, il ne peut pas porter sur l'acte de langage de la phrase matrice. Il resterait à analyser en détail si ce comportement peut être étendu à l'ensemble des particules présentées comme modales.

## 3.2.2 Quelques aspects de la concession

Cette section présente des recherches sur des marqueurs discursifs ciblés. Pour des raisons de cohérence, je me concentre sur les travaux traitant du marquage de la concession. Ils me permettent d'illustrer deux directions d'analyse : une direction diachronique qui emprunte le cadre des Grammaires de construction et une direction argumentative et interactionnelle.

### 3.2.2.1 De la comparaison à la concession

La concession n'est pas toujours très bien définie. Pour dire les choses très simplement, elle repose sur la perception d'un contraste, d'une opposition, entre une proposition et la conclusion implicite d'une autre proposition. Que l'on regarde la théorie de la SDRT, où la concession est présentée comme un cas particulier de la relation **Contraste** (Asher & Lascarides 2003, p. 465), le système de Renkema (2009, p. 78) ou de Nazarenko (2000, p. 31-34), pour qui la relation **Concession** est présentée comme un cas particulier de la cause, ou la description proposée par Morel (1998), se dessine un schéma commun : une proposition assertée, dite concédée dans les approches argumentatives, et la négation d'une conséquence normale, attendue, de cette proposition. Le français possède des moyens assez variés pour exprimer cette relation et plusieurs connecteurs peuvent la marquer. Avec Bernard Combettes, nous nous sommes intéressés plus particulièrement aux constructions corrélatives de type « adverbe intensif + adjectif +

que + proposition », comme dans les exemples (95), (96) et (97), repris à Combettes & Dagnat (2016b [26], p. 2 et 4).

(95) Mon mari si gentil qu'il soit a du mal à suivre mes changements d'humeurs (internet, forum, avril 2015, ex. 2, p. 2)

(96) Tout chrétiens qu'ils fussent officiellement devenus, ils s'efforçaient de vivre comme ils avaient toujours vécu. (ex. 6, p. 4)

(97) Cette jeune fille était, quelque fallacieuse qu'elle fût, indispensable à ma vie. (ex. 7, p. 4)

La perspective adoptée est celle d'une analyse diachronique qui emprunte certains éléments au cadre des CxG. Ceci nous a amenés à chercher l'origine de la construction et des marqueurs associés du côté de la comparaison et à interroger le passage de la comparaison à la concession. Le cadre des CxG a permis d'articuler différents niveaux de la description linguistique et de penser les constructions apparentées de manière hiérarchique. Je reprends ci-dessous les principales idées et conclusions de la démarche.

### (i) CxG et diachronie

Depuis déjà quelques années, la linguistique diachronique appréhende certains phénomènes dans le cadre des CxG. La possibilité d'une représentation multidimensionnelle est un des atouts souvent mentionnés de ce cadre. Nous n'avons pas décrit toutes les dimensions, mais nous avons cherché à mettre en évidence le fonctionnement d'une construction, c'est-à-dire ici l'association d'une forme syntaxique/discursive et d'une signification. Un autre point pertinent pour l'étude du changement linguistique est la hiérarchisation entre les constructions d'une même famille sur la base d'une recherche d'héritage de propriétés. Des réflexions récentes (par ex. Hoffmann 2013, p. 235, Traugott 2007, p. 525 et 2008, p. 236, Trousdale 2012 et 2015) aboutissent à un classement des constructions par degrés d'abstraction enchâssés. Dans son étude sur les modificateurs de degré en anglais (*a sort of, a lot of, a shred of*, etc.), Traugott propose de distinguer au moins quatre niveaux :

- *macro-constructions*: meaning-form pairings that are defined by structure and function, e.g., Partitive, or Degree Modifier Constructions, etc.,
- *meso-constructions*: sets of similarly-behaving specific constructions,
- *micro-constructions*: individual construction-types,
- *constructs*: the empirically attested tokens, which are the locus of change. (Traugott 2008, p. 236)

Nous avons proposé de transposer cette vision à la construction concessive étudiée pour aboutir à la distribution suivante :

- Les macro-constructions désignent les constructions générales du type « X intensif + item + subordonnée », avec une interprétation comparative (*Pierre est aussi/plus grand que Paul (est)*), concessive (*Si grand que soit Paul, il passe sous la porte*) ou consécutive (*Paul est si grand qu'il touche le plafond*).
- Les méso-constructions désignent l'ensemble des constructions particulières ayant des fonctionnements analogues, ici « ADV intensif + adjectif + subordonnée ».
- Les micro-constructions désignent des constructions de type méso-constructions mais dont certains composants sont spécifiés. Par exemple : « *si* + adjectif + subordonnée en *que* au subjonctif », ou « *tant* + adjectif + subordonnée en *que* au subjonctif », etc.



- Les *constructs*, qu'on peut traduire par *énoncés*, sont des instanciations de micro-constructions et correspondent aux exemples attestés.

Une telle hiérarchisation est utile en synchronie pour établir une typologie de constructions à partir d'usages. Elle l'est du point de vue diachronique pour mettre en évidence des pivots interprétatifs dans l'évolution d'une construction (par exemple comparaison/concession), la disparition, l'apparition voire même la réapparition d'éléments dans la langue.

(ii) *À partir de la construction concessive en si*

Notre démarche a été ascendante, c'est-à-dire que nous sommes partis d'une micro-construction assez fréquente aujourd'hui avec une interprétation concessive, celle qui met en jeu l'adverbe *si* (ex. 95). D'une part, nous avons cherché à mettre en évidence quelques étapes de son parcours diachronique, ce qui nous a amenés à décrire aussi la construction apparentée en *tant* qui n'existe plus aujourd'hui. D'autre part, nous nous sommes interrogés sur d'autres micro-constructions qui, bien que présentant quelques particularités, pourraient aussi rentrer dans le schéma « ADV intensif + adjectif + subordonnée » (notamment celles qui mettent en jeu *tout* et *quelque*). Je rappelle ici les points essentiels concernant les constructions en *tant* et *si*.

**Tant.** Les premières constructions en *tant* à valeur concessive semblent remonter au XII<sup>ème</sup> siècle<sup>197</sup> et perdurent jusqu'au XVII<sup>ème</sup> siècle.

- (98) Mes il estoit tex chevaliers qu'il ne dotoit nul home, tant fust forz ne fiers. (C. de Troyes, *Chevalier à la charrette*, 1180) / *Mais il était chevalier à ne redouter aucun homme, si fort ou farouche fût-il.*

On peut faire plusieurs remarques sur ce type de construction :

- La portée est sur l'adjectif et c'est la construction attributive dans son ensemble qui a le statut d'apposition. Le subjonctif marque le lien de subordination de type concessif<sup>198</sup>.
- Les contextes d'emplois sont très souvent négatifs.
- Il y a une coréférence entre les deux propositions, ce qui explique la non-expression du sujet du verbe *être*.

Ces propriétés semblent s'effiloche assez tôt. On peut considérer que certains exemples ambivalents sont les témoins de cette période de flottement où la portée n'est plus systématiquement un adjectif mais un locatif, où l'adverbe perd son sens intensif et où la coréférence n'est plus une obligation<sup>199</sup>. Un autre élément est à prendre en compte, l'apparition, dès le moyen français et surtout à partir du XVI<sup>ème</sup> siècle, de la corrélation avec *que* :

- (99) Nuls, tant hauls ne nobles que il fust, n'estoit espargniés. (Froissard 1400) / *Nul, si haut ou si noble que il fust, n'était épargné.*

Pour expliquer cette évolution et cette présence de *que* avec *tant*, nous avons avancé l'hypothèse d'une influence du système des constructions en « *si* + adjectif + *que* +

197 En fonction des attestations relevées dans le corpus Frantext intégral.

198 Un indicatif ici conduirait à une interprétation de type causal.

199 Un exemple comme « Bataille ne aventure ne savront que il n'aillent querre, tant soit an felenesse terre (*Perceval*, fin XII<sup>ème</sup> siècle) peut s'interpréter de deux manières en fonction du sujet que l'on reconnaît à *soit*. Ce sujet peut renvoyer à « bataille ne aventure » (*ils ne connaîtront ni bataille ni aventure qu'ils ne recherchent même si elle était en terre païenne*) ou à la situation plus générale (*ils ne connaîtront [...] même si c'était en terre païenne*).

proposition au subjonctif » et « *si* + adjectif + *comme* + proposition à l'indicatif », les premières exprimant la concession et les secondes la comparaison.

**Si.** L'histoire des constructions concessives en *si* présente des caractéristiques communes avec celle des constructions en *tant*, mais s'en distingue sur plusieurs points :

- Le contexte est moins systématiquement négatif dans les constructions en *si*.
- Les premiers exemples présentent des constructions avec un sens comparatif et descriptif où il est parfois difficile de repérer une opposition. De plus, *si* se construit au départ avec *comme* qui demande l'indicatif, ce mode ne permettant pas d'inférer une relation de type concessif. La comparaison entre les exemples (100) et (101) montre que le contexte joue un rôle important.

(100) Belinanz, si correciez com il estoit, monte a cheval... (*Le roman de Tristan en prose*, XIII<sup>ème</sup> siècle) / *Belinant, (aus)si irrité qu'il était, monte à cheval.*

(101) Quant il fut remonte, si las comme il estoit, si se fiert en la bataille (*La mort Artu*, 1230) / *Quand il fut remonté à cheval, si las qu'il fût, il se jette dans la bataille.*

Dans le premier cas, l'interprétation semble purement descriptive, il n'y a pas de lien d'implication contrariée entre le fait d'être en colère et le fait de monter à cheval. En revanche, dans (101), il y a bien une opposition entre une conséquence implicite d'être fatigué et de se jeter dans la bataille.

- Deux autres étapes achèvent le parcours de la construction que nous connaissons aujourd'hui, il s'agit du remplacement de *comme* par *que* et l'apparition du subjonctif que ce remplacement rend possible. De rares exemples montrent un stade ambivalent, où « *si* + adjectif + *que* » est suivi d'une proposition à l'indicatif mais peut recevoir une interprétation concessive<sup>200</sup>. Pour l'utilisation systématique du subjonctif, nous avons avancé l'idée d'un alignement des constructions en « *si* ... *que* » sur le paradigme des concessives concurrentes en *bien que* et *quoique*.

D'autres micro-constructions mériteraient une analyse et certains points nécessitent encore des approfondissements. On peut se demander en quoi les structures présentées relèvent des MD, dans la mesure où elles impliquent toute une construction et une dépendance de type subordination corrélatrice. Cela soulève la question d'éléments discontinus (« *si*... *que* »), qui ne sont généralement pas mentionnés dans les listes de MD. Cependant, ces éléments sont figés, sont associés à une relation de discours (la concession), et demandent pour que l'interprétation soit complète qu'on accède à une autre proposition (propriété d'indexicalité).<sup>201</sup>

Je retiens de cette approche diachronique le fait qu'elle apporte un cadre et des contextes d'emplois permettant d'expliquer et d'explicitier des parcours d'interprétation et des nuances sémantiques « en réseau ». Elle met en évidence à la fois l'influence du contexte sur la construction et les marqueurs associés et l'impact de la construction et des marqueurs associés sur le contexte. De plus, l'approche diachronique soulève la question même de l'identification des processus d'évolution du sens, désormais

200 « Si so et si enfes que je suis, suis-je plus sages que vos n'estes » (*Placides et Timeo*, fin 13<sup>e</sup>) / *Si sot et si puéril que je sois, je suis plus sage que vous n'êtes.*

201 Les autres items que nous avons étudiés du point de vue diachronique sont plus « lexicalisés » et sont généralement relevés parmi les connecteurs ou les MD en général : *dans la mesure où* et *en principe* (Combettes & Dagnat 2020a [33]) et *d'autant que* (Combettes & Dagnat 2021c [41]).

relativement bien conceptualisés, et qui occupent une place importante dans la réflexion sur les MD. Je les indique brièvement ci-dessous.

**(iii) Processus de XXXcalisation ?**

Les MD sont fréquemment présentés comme le résultat d'un processus d'évolution – on parle d'ascension (*rise*) (voir Heine et al. 2021a, 2021b) – qui donnerait à des formes d'origine nonMD certaines des propriétés évoquées ci-dessus. Les étapes de cette évolution varient en fonction des items concernés mais certaines sont récurrentes et correspondent à des « chemins » que l'on a coutume de nommer *grammaticalisation*, *pragmaticalisation* et dans certains cas *lexicalisation*. Ces trois processus reçoivent des définitions plus ou moins strictes et sont plus ou moins liés. D'où la formule XXXcalisation que nous avons employée dans Combettes & Dagnat (2021c [41], p. 67) à propos de l'évolution du marqueur *d'autant que*. Les réflexions générales sur ces processus ne sont pas récentes. Marchello-Nizia (2006, chap. 1) rappelle l'histoire des études utilisant le terme *grammaticalisation* et son articulation avec d'autres processus comme la lexicalisation, la subjectivisation, la réanalyse et l'analogie. L'idée d'une pragmaticalisation semble plus récente et correspondre à une étape supplémentaire dans l'évolution de certains items déjà grammaticalisés. Je ne cherche pas à ordonner le champ conceptuel, assez variable et déjà très documenté, mais le domaine des MD m'a nécessairement exposée à des définitions de ces processus et à leurs enjeux. Le cadre de cette synthèse d'HDR est l'occasion de clarifier certains points afin de mieux les exploiter dans mes travaux futurs.

Il existe plusieurs versions de la grammaticalisation. Elle est d'abord vue comme un processus que l'on décrit par son résultat : l'acquisition par un item lexical de fonctions grammaticales ou l'accès d'un item déjà grammatical à de nouvelles fonctions grammaticales<sup>202</sup>. Ce processus permet donc l'introduction de nouveaux éléments dans la grammaire d'une langue. Typiquement, pour le français, on peut citer le cas du futur périphrastique (*aller* + infinitif), l'auxiliarisation du verbe *avoir* ou l'utilisation comme préposition du nom *côté*. Les paramètres utilisés pour dire s'il y a ou non grammaticalisation renvoient généralement à ceux avancés par Lehmann (2015 [1982], p. 129-189).

La pragmaticalisation est une idée plus récente. Elle est présentée comme le processus par lequel, dans un contexte donné, la valeur propositionnelle d'un item lexical (ou d'un syntagme) change. Cet item ou ce syntagme assure alors une fonction au niveau du discours, de l'organisation textuelle, de l'interaction ou de la métacommunication. Ce processus de changement est dès le départ associé aux MD (voir initialement Erman & Kostinas 1993 sur *ba'* et *you know* en anglais, et, pour le français, Dostie 2004 sur les MD d'origine verbale comme *tiens*, *écoute*, *dis*). En effet, les études sur l'émergence des MD montrent que la mise en place de leur fonctionnement discursif ne se laisse pas décrire totalement en suivant le schéma prototypique de la grammaticalisation et qu'un nouveau schéma d'évolution est nécessaire. Certains critères de la grammaticalisation ne sont pas respectés. Ce qui « coince » concerne notamment le fait que l'acquisition de fonctions discursives et pragmatiques correspond généralement à l'optionnalité sémantique, à l'élargissement de la portée et à

202 D'après Hancil & König : « Most linguists working in this domain would define the term roughly as follows : Grammaticalization is a process through which lexical elements develop into grammatical markers and grammatical markers become more grammatical. » (2014, p. 2)

l'autonomisation syntaxique de l'item d'origine, au contraire de ce qui se passe dans le processus de grammaticalisation.

La nécessité de créer un nouvel espace de réflexion semble correspondre en fait au besoin d'une conception large de la notion de grammaire, qui intègre la structuration du discours, c'est-à-dire une prise en compte de phénomènes textuels et pragmatiques. L'intérêt de distinguer ou d'articuler grammaticalisation et pragmaticalisation et d'élargir la grammaire au discours fait toujours débat<sup>203</sup>. A un niveau général, j'en retiens que se pose frontalement la question de l'empan et des dimensions d'une grammaire et, pour les MD, de leur légitimité ou non à être décrits comme des éléments de cette grammaire. Au niveau de propriétés plus spécifiques, les différentes études, qu'elles fassent converger ou diverger grammaticalisation et pragmaticalisation, permettent de mettre en évidence la dynamique de formation des MD. Je pense en particulier aux phénomènes de subjectivisation<sup>204</sup> et d'intersubjectivisation qui introduisent des dimensions cognitive, expressive, interactionnelle et argumentative dans le sens (voir Frank-Job 2006, Traugott & Dasher 2002, Traugott 1982, 1995, 2010).

– Le processus de lexicalisation est aussi régulièrement convoqué et son rapport avec la grammaticalisation fait également débat (voir Brinton & Traugott 2005, p. 20-32, Lightfoot 2011). De Mulder (2021, p. 1820) isole deux définitions : une qui insiste sur un processus de conventionnalisation d'une expression linguistique et de son sens, qui finissent par former une unité lexicale à part entière, une qui insiste sur la nature complexe des expressions dites lexicalisées et leur caractère non-compositionnel (voir aussi Hancil & König 2014<sup>205</sup>). Dans mon cas, cette perspective s'avère particulièrement pertinente pour l'étude des combinaisons de marqueurs. C'est une question que je me suis posée à propos de *mais enfin* (Dagnat 2022a [42]). Il me semble que l'agencement est encore compositionnel (combinaisons de valeurs de *mais* et de valeurs de *enfin*) et que la lexicalisation n'est pas encore avérée. Mais des réductions phonétiques et graphiques comme [mãfẽ] et *m'enfin*, conventionnelles et spécifiques à certains contextes et certaines interprétations, pourraient laisser penser qu'une lexicalisation est en cours.

D'autres processus ou d'autres manières de penser ces processus de changement ont été avancés plus récemment. Il s'agit de la constructionnalisation (Traugott & Trousdale 2013, Traugott 2014), de la routinisation (Detges et Waltereit 2016) et de la cooptation (Heine et al. 2021a et 2021b). Ils me paraissent très productifs pour l'analyse des MD, mais je n'ai pas encore eu le temps d'y confronter plus précisément mes recherches. Ils s'attachent pour certains à des théories plus larges, comme la *Thetical Grammar*<sup>206</sup> ou les CxG. Il me semble que dans les trois cas les auteurs cherchent à

203 Voir par exemple Aijmer & Simon-Vandenberghe 2011, Badiou-Monferran & Buchi 2012, Brinton 2010, Degand & Evers-Vermeul 2015, Detges & Waltereit 2016, De Mulder 2021, Diewald 2012, Heine 2013, Traugott & Trousdale 2013, section 3.2, p. 96-112.

204 « Subjectification is the semasiological process whereby [speakers/writers] come over time to develop meanings for [lexemes] that encode or externalize their perspectives and attitudes as constrained by the communicative world of the speech event, rather than by the so-called 'real world' characteristics of the event or situation referred to. » (Traugott & Dasher 2002, p. 30)

205 « [lexicalization] : the change from complex expressions to lexical items as a result of a loss of compositionality. » (Hancil & König 2014, p. 2)

206 Dans *Thetical Grammar*, *thetical* est une réduction de *parenthetical*. Une telle grammaire a pour objectif de répertorier les expressions et les constructions qui répondent à un certain nombre de

conceptualiser le changement linguistique de manière large et à dépasser les oppositions classiques entre grammaticalisation, pragmaticalisation et lexicalisation. Parmi les dimensions qui m'intéressent tout particulièrement, je retiens la place de la fréquence dans le processus de routinisation, le changement de portée dans le processus de cooptation, et, pour la constructionnalisation, le rôle du contexte et la question de la (non-)compositionnalité.

### 3.2.2.2 *Concession et interaction*

La concession a un rapport étroit avec l'argumentation, puisqu'on peut considérer qu'elle implique la prise en compte d'autres points de vue que celui du locuteur. Elle sert d'ailleurs souvent à anticiper et désamorcer de potentiels contre-arguments à la thèse défendue et par là-même la renforcer (voir section 2.2.1, p. 75). Dans des situations dialogales, elle peut également servir à remettre en cause des croyances ou des attitudes qu'on attribue à l'interlocuteur. Dans (Dagnat 2022b [43]), je me suis intéressée à des emplois de *mais* où un locuteur met en cause la légitimité de la contribution de son interlocuteur à l'échange. Il s'agit d'un entretien médical entre un médecin et son patient<sup>207</sup>. Le médecin sollicite la réaction du patient à propos d'un compte rendu d'analyse qu'il lui a envoyé et lui explique certains aspects de ce compte rendu. Je reprends les principales conclusions de l'analyse d'un extrait, repris sous (102), pour montrer la variété des référents possibles de *mais*. Les tours de parole sont numérotés entre crochets pour faciliter le commentaire. Les # indiquent les chevauchements dans la transcription.

- (102)  
 MED : #1 donc il y a pas de problèmes # [1]  
 PAT : #2 mais j'ai eu des C.R.P à dix # [2]  
 MED: oui mais dix c'est rien [3]  
 PAT: ah bon {ton de stupéfaction} + mais pourquoi alors les laboratoires sanguins {ton interrogatif} [4]  
 MED: mais parce que il faut qu'ils aient des normes [5]  
 PAT: mais oui mais pourquoi ils adoptent des normes qui seraient pas adaptées alors {ton interrogatif} [6]  
 MED: oui mais {rires} là je m'excuse + vous êtes en train de euh de [7]  
 PAT: je cherche à comprendre moi [8]  
 MED: oui mais moi #1 je vous explique # [9]  
 PAT: #2 oui #  
 MED: j'ai l'expérience [10]  
 PAT: non mais j'en doute pas c'est pour ça que je suis venu vous voir [11]  
 MED: j'ai l'expérience de de de gérer ce type de problèmes + et je sais très bien que une C.R.P à dix je vais pas bouger [12] (DÉCLICS CO6, repris à Dagnat 2022b [43], extrait IV, p. 196)

Le médecin vient d'indiquer que la valeur 2 pour la mesure de CRP (C Reactive Protein) ne pose pas problème [1]. En [2], le patient remet en cause la légitimité de la

---

propriétés, notamment le fait d'être syntaxiquement et prosodiquement autonomes et de ne pas participer au contenu propositionnel de leur énoncé-hôte (*non-restrictive meaning*) (voir Kaltenböck et al. 2011, p. 7, Heine 2013, p. 1213-1224).

207 Cette étude a été faite dans le cadre de ma participation au projet interdisciplinaire DÉCLICS (Blasco 2021), qui proposait un accès à des entretiens médicaux divers. Les données ne sont pas en libre accès pour des raisons de confidentialité. La structure de l'entretien est disponible sous forme de graphe à [http://mathilde.dagnat.free.fr/index\\_fichiers/Dagnat-explication-argumentation-ref1-Graphe.jpg](http://mathilde.dagnat.free.fr/index_fichiers/Dagnat-explication-argumentation-ref1-Graphe.jpg)

conclusion *il n'y a pas de problème* en signalant qu'il a eu des valeurs plus élevées, qui pourraient être problématiques. Il introduit donc une nouvelle prémisse (CRP à 10) dans le raisonnement tout en acceptant la première prémisse (CRP à 2). En [3], le médecin accepte à son tour la deuxième prémisse (*oui*), mais en refuse la pertinence (*mais*) par rapport à la conclusion de [1]. Que la CRP soit à 2 ou à 10 cela ne pose pas de problème. En [4], la réaction du patient est double. D'une part, la particule complexe *ah bon* et son intonation montante marquent l'étonnement, d'autre part, le *mais* présente la question en *pourquoi* comme argumentativement opposée à quelque chose. Il est peu probable que le patient mette en doute la parole du médecin à laquelle il fait référence avec le *alors*. Il est plus probable que le *mais* signale une situation qu'il juge anormale (si la valeur de CRP 2 n'est pas problématique, le laboratoire ne devrait pas la souligner dans son compte rendu). Le patient pointe ainsi un certain illogisme du compte rendu. En [5], le médecin réplique en expliquant (*parce que*) l'attitude du laboratoire par la présence de normes. Cette explication est elle-même introduite par un *mais* qui peut souligner soit qu'elle est très accessible et donc n'aurait pas dû donner lieu à une question (le *mais* porte sur la légitimité de l'acte de langage de question du patient<sup>208</sup>), soit que la question est inappropriée dans la relation patient-médecin. Les deux aspects ne sont d'ailleurs pas incompatibles. En [6], le patient concède la nécessité d'avoir des normes, mais remet en cause l'adéquation de ces normes. La réponse du médecin en [7] présente plusieurs traits. Le *mais* marque la distance par rapport à [6]. Le *je m'excuse* indique que l'acte sur lequel il porte peut être interprété comme une critique, une réserve, un désaccord, etc. La phrase incomplète *vous êtes en train de* vise quelque chose d'indéterminé mais dont on peut faire l'hypothèse qu'il s'agit d'un sous-entendu (« les laboratoires ne sont pas sérieux ») ou d'un comportement (« vous créez des difficultés dans notre échange », « vous posez des questions inutiles »). Les échanges qui suivent font pencher pour la seconde hypothèse, dans la mesure où l'enjeu se déplace sur les positions des participants à l'entretien (l'ignorant qui *cherche à comprendre* [8] et le sachant qui *explique* [9]). Le *mais* de [9] conteste l'inférence que le patient serait livré à lui-même pour comprendre la situation. En [11], on peut interpréter le *mais* du patient comme excluant un sous-entendu porté par [10], à savoir que le médecin ne serait pas tout à fait compétent.

De telles interactions sont intéressantes pour deux raisons principales : d'une part, elles illustrent la flexibilité indexicale des MD, en particulier *mais*, d'autre part, elles montrent l'interconnexion entre les agencements argumentatifs et les places sociales dans un environnement institutionnel où elles sont très pertinentes<sup>209</sup>. Si l'on revient sur la présentation des structures argumentatives (voir section 2.2.2.1, p. 83), on voit comment les relations argumentatives articulant du contenu sont aussi opérationnelles au niveau métadiscursif (en particulier pertinence des actes de langage, traitement des sous-entendus, gestion des places dans l'interaction).

208 Cela relève des *conditions préparatoires* dans l'analyse que Searle (1972 [1969], p. 108-109) donne des actes de langage.

209 Voir Heritage J. & Clayman S. 2010. *Talk in Action : Interactions, Identities, and Institutions*. Chichester : Wiley-Blackwell, Part 3.

### 3.2.3 Particules énonciatives et prosodie

Cette section est une synthèse de travaux sur les propriétés prosodiques de certaines PEN et sur leur pertinence pour la désambiguïsation en cas de polyfonctionnalité. Les questions de fond qui ont animé ces recherches étaient de savoir s’il était possible pour des items polyfonctionnels<sup>210</sup> d’utiliser certaines propriétés prosodiques comme indicateurs d’usages PEN ou non-PEN (section 3.2.3.1), et s’il était possible d’utiliser certaines propriétés prosodiques pour discriminer les différentes fonctions d’une PEN (par ex. *bon* marquant une conclusion et *bon* marquant une ouverture) (section 3.2.3.2).

Les travaux concernés sont essentiellement des travaux en collaboration. Ils ont été effectués dans le cadre du projet PARDI!<sup>211</sup>, que j’ai initié et coordonné et qui s’est ensuite développé dans le cadre de l’opération *ProsodCorpus*<sup>212</sup> et autour du co-encadrement de thèse de L. Lee (2021). Ci-dessous, j’utilise l’étiquette PEN par souci de cohérence, mais dans les travaux cités c’est DP (*discourse particle*) qui apparaît.

#### 3.2.3.1 Particule vs non-particule ? Approche d’une classe

L’objectif du projet PARDI! était de lancer une analyse multidimensionnelle (syntaxe, sémantique, prosodie) des PEN en français. Dans les faits, nous nous sommes principalement centrés sur la prosodie et sur les PEN les plus fréquentes dans les corpus : d’abord *tu vois, vous voyez, tu sais, vous savez, tu comprends, vous comprenez, écoute, écoutez*, puis *quoi* et *voilà* et enfin *alors, bon* et *ben*. Pour rendre compte des analyses, de la méthode et des résultats, je m’appuie sur la co-direction du mémoire d’A. Bastien (2015), de Dagnat et al. (2015b [24]) et Bartkova et al. (2016a [25]).

Une première étape a consisté à récupérer des données quantitativement et qualitativement suffisantes pour l’analyse prosodique et à les prétraiter pour pouvoir les exploiter (i). Une deuxième et une troisième étapes ont consisté à annoter manuellement les emplois PEN et non-PEN des items étudiés, et à tester ensuite s’il était possible, sur la base des propriétés prosodiques, de détecter automatiquement cette différence d’usage (ii).

##### (i) Préparation et prétraitement des données

Les données ont été extraites de corpus oraux déjà transcrits<sup>213</sup>. L’étude prosodique a été rendue possible grâce à plusieurs étapes de prétraitement effectuées par D. Jouvét.

Tout d’abord, il a fallu segmenter et annoter automatiquement les fichiers récupérés au moyen de plusieurs outils, notamment *CMUSphinx*<sup>214</sup> (Mesbahi L. et al. 2011) pour la traduction graphème-phonème et l’alignement forcé, ensuite les programmes *Prosotran* (Bartkova et al. 2012) et *Prosotree* (Bartkova & Jouvét 2013).

210 Par exemple *bon* adjectif vs *bon* PEN. D’autres items, par exemple *ben*, ont essentiellement un fonctionnement comme PEN dans les corpus cités. En français québécois, *ben* a aussi les valeurs adverbiales intensives de *bien*.

211 Projet PARDI! (PARTicules DIscursives : sens et corrélats prosodiques), financement ATILF et Université de Lorraine (2014-2016).

212 Opération *ProsodCorpus* du CPER Langues, Connaissances et Humanités Numériques, piloté par D. Jouvét (2015-2017).

213 Corpus utilisés : ESTER, ETAPE et TCOF, mais occasionnellement et partiellement pour TCOF, à cause de la médiocre qualité sonore. La « matière première » consistait en un fichier son (.wav), une transcription orthographique « classique » avec Transcriber (.trs) indiquant les locuteurs, les tours de parole et les événements particuliers perturbant le signal.

214 *CMUSphinx* : <https://cmusphinx.github.io>

*Prosotran* code les propriétés prosodiques relatives de chaque voyelle (niveau, longueur, intensité) et permet de calculer et de noter des valeurs de pente entre la voyelle-cible et la voyelle des syllabes à gauche et à droite de la voyelle-cible.

*Prosotree* permet une segmentation hiérarchisée en groupes prosodiques sur la base du repérage de la durée et de l'accentuation des voyelles et de repères lexicaux (mots).

Ensuite, on a extrait automatiquement les expressions étudiées (ici *tu sais*, *alors*, etc.) avec un contexte de 5 mots avant et 5 mots après. L'objectif pratique était de pouvoir visualiser avec *Praat* la courbe mélodique des extraits ainsi que les différents niveaux d'annotation. La Figure 21 ci-dessous en est une illustration.

La courbe mélodique (en haut) est exprimée en semi-tons pour lisser les variations individuelles. Les quatre tiers, en partant du haut, correspondent aux phonèmes en SAMPA (T1), à la transcription orthographique (T2), aux groupes prosodiques obtenus avec *Prosotree* (T3) et aux propriétés prosodiques de chaque voyelle obtenues avec *Prosotran* (T4).<sup>215</sup>

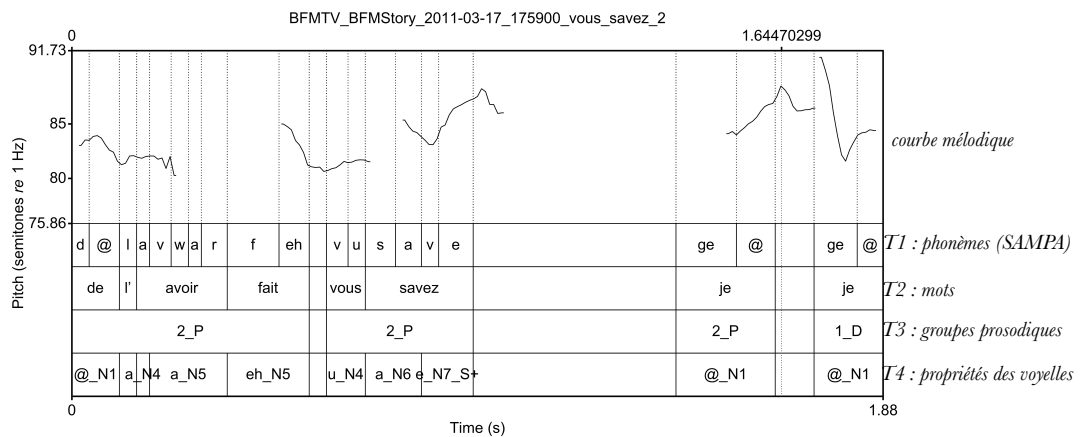


Figure 21. Segmentation et annotation prosodique automatique (ex. de *vous savez*)

### (ii) Annotation et détection des emplois *PEN* et *nonPEN*

Nous avons procédé à une annotation manuelle pour séparer systématiquement les emplois *PEN* des emplois *non-PEN*. Techniquement, ceci a été rendu possible grâce à l'interface *CorpusDisplay*<sup>216</sup>, qui permet d'assigner un label *PEN* ou *non-PEN* à chacun des extraits en ayant la possibilité de jouer le signal et d'afficher son analyse dans *Praat* selon le modèle de la Figure 21 ci-dessus. Grâce à ces multiples traitements, il était possible pour chaque occurrence de *tu vois*, *vous voyez*, etc. d'obtenir sous forme de tableur des valeurs pour les propriétés suivantes :

<sup>215</sup> La notation proposée par *Prosotree* (2\_P, 1\_D) rend compte du niveau de cohésion entre deux groupes prosodiques, plus le chiffre est élevé, moins la frontière entre les deux groupes est marquée. La notation proposée par *Prosotran* comporte plusieurs éléments : la durée des voyelles (D+ à D+++), le niveau tonal (N1 à N10) et la slope (S+ à ++++) entre l'expression étudiée et dernière syllabe de son contexte gauche et entre l'expression étudiée et la première syllabe de son contexte droit, en prenant en compte le seuil de glissando pour intégrer un paramètre perceptif. L'énergie de chaque voyelle est aussi récupérable, mais nous ne l'avons pas utilisée.

<sup>216</sup> L'interface a été réalisée et adaptée par D. Jouvét. Elle permet de visualiser la courbe prosodique et les 4 tiers d'annotation tout en écoutant le fichier son correspondant. Ensuite, nous pouvions comparer les annotations faites par les étudiantes (A. Bastien, L. Lee en M2 puis en doctorat) et par K. Bartkova et moi-même, et nous tranchions après discussion et réécoute.



<b>Pause</b>	– absence de pause – présence de pause avant et après la PEN et durée de la pause (seuils : $\leq 150$ ms, $> 150$ ms et $\leq 250$ ms, $\geq 250$ ms)
<b>Niveau tonal</b>	(exprimé sur une échelle de 10 relative à l'étendue vocale du locuteur) – de chaque syllabe de la PEN – de la dernière syllabe du contexte gauche – de la première syllabe du contexte droit
<b>Pente</b>	(valeur positive = pente montante ; valeur négative = pente descendante) – inclinaison entre le contexte gauche et la PEN – inclinaison entre la PEN et le contexte droit – inclinaison entre la première et la dernière syllabe de la PEN
<b>Slope</b>	(calculée en prenant en compte le niveau tonal de la dernière syllabe du contexte gauche, de la dernière syllabe de la PEN et le seuil de glissando) Cette valeur est exprimée en quatre niveaux : S+, S++, S+++ et S++++, selon que la valeur est 1, 2, 3 ou 4 plus grande que le seuil de glissando
<b>Durée</b>	– de la PEN (sur la base de la voyelle de sa dernière syllabe) – de la voyelle de la dernière syllabe du contexte gauche Cette valeur est exprimée en trois niveaux : D+, D++, D+++
<b>Position dans le GP</b>	– position initiale – position médiane – position finale

Tableau 11. Propriétés prosodiques pour l'analyse des PEN

Une fois effectués les prétraitements des données et l'annotation manuelle, un certain nombre de comparaisons ont été faites sur la base de cinq propriétés du Tableau 11 : la présence de pauses avant et après l'item étudié et la durée de cette pause, la position de l'item dans le groupe intonatif, son niveau tonal, sa valeur de slope en fonction des contextes gauche et droit.

Par exemple, pour *quoi*, dans les corpus ESTER et ETAPE, il est possible de faire les remarques suivantes (Dagnat et al. 2015 [24], p. 43-47) :

- *Quoi* apparaît majoritairement comme nonPEN (630 occ. nonPEN vs 399 PEN).
- *Quoi*-nonPEN (pronom) apparaît majoritairement sans pause (66,5%). *Quoi*-PEN apparaît majoritairement avec une pause longue après ( $> 250$  ms). Cela peut être corrélé à des propriétés syntaxiques et informationnelles : *quoi*-nonPEN est le plus souvent argument d'un verbe et *quoi*-PEN est majoritairement conclusif.
- *Quoi*-PEN apparaît le plus souvent comme mot isolé, c'est-à-dire non intégré à un groupe intonatif – ce qui correspond à son autonomie syntaxique et sémantique – contrairement à *quoi*-nonPEN, qui apparaît autant comme mot isolé que comme mot intégré en position médiane dans le groupe intonatif – ce qui correspond à une plus grande intégration sémantique et syntaxique.
- Sur une échelle de 10 représentant l'étendue vocale du locuteur, *quoi*-PEN est très souvent prononcé avec des valeurs de F0 basses et très rarement avec des valeurs hautes, contrairement à *quoi*-nonPEN. Ceci est sans doute à corrélérer avec la prédisposition conclusive de l'emploi PEN.
- Concernant la slope (pente en fonction du contexte et du seuil de glissando), *quoi*-PEN présente très souvent une slope descendante à partir de son contexte gauche, et à un niveau assez bas. Le comportement de *quoi*-nonPEN est sensiblement différent, puisque, dans ce cas-là, la slope à partir du contexte gauche est majoritairement montante.
- Enfin, concernant la durée et l'allongement potentiel de la voyelle de l'item, qu'il soit PEN ou nonPEN, dans environ 80 % des cas, aucun allongement n'est détecté. Parmi les 20 % restant, les différences entre PEN et nonPEN sont ténues et il paraît

difficile de tirer des conclusions à partir de ce critère. Mais, pour d'autres items, comme *voilà*, le critère de la durée s'avère plus pertinent.

Pour aller plus loin et tester la pertinence des cinq paramètres prosodiques retenus et leur intérêt dans le domaine du TAL, nous avons construit une expérience de détection automatique de la fonction PEN ou nonPEN. Pour cela, il a fallu utiliser un arbre de décision<sup>217</sup>. L'arbre a été entraîné sur 60 % des données des deux corpus (ESTER et ETAPE) et les 40 % restant ont servi pour l'évaluation. Il s'avère que pour *quoi* l'identification automatique a été correcte dans 71 % des cas, ce qui est un score perfectible mais relativement encourageant. Dans Bartkova et al. (2016b [25]), la démarche a été la même sur un corpus et une liste d'items augmentés, ce qui a permis d'améliorer la tâche de détection automatique.

Nous avons complété l'expérience en analysant plus spécifiquement les propriétés prosodiques des emplois comme PEN (2016b [25], p. 462). Le but était de mettre en évidence plusieurs profils prosodiques pour chaque item et l'hypothèse était que ces profils pouvaient correspondre à des différences de signification d'une même PEN. Ainsi, une première approche a consisté à utiliser la méthode des K-moyennes (*K-means*), toujours avec *Weka*, et à extraire les cinq clusters les plus homogènes pour *alors*, *ben*, *bon*, *quoi*, *voilà* et *vous savez*. Le partitionnement en K-moyennes est une méthode répandue de classification qui consiste à rechercher dans un ensemble de données la division optimale en un nombre (*k*) de groupes jugé pertinent par l'utilisateur. Pour l'analyse, les données sont des vecteurs composés des valeurs des propriétés présentées plus haut dans le Tableau 11 pour chaque occurrence de PEN. L'interface permet de jouer avec la taille de ces vecteurs et de choisir certaines propriétés seulement (par exemple les propriétés propres à la PEN ou les propriétés relatives à son contexte gauche et/ou droit). L'analyse pilote menée s'est concentrée sur la hauteur tonale de la PEN, la hauteur tonale du contexte droit et du contexte gauche, la durée de la voyelle finale de la PEN. Le Tableau 12 présente les cinq classes les plus homogènes si l'on prend en compte le niveau tonal des contextes gauche et droit et de la dernière syllabe de la PEN étudiée (valeur au centre et en rouge pour une impression couleur). L'échelle de hauteur est sur 10 en semi-tons.

	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5
<i>alors</i>	1-3-2	7-4-2	5-5-6	5-6-8	7-8-8
<i>ben</i>	2-2-7	3-2	5-4-1	4-6-6	7-6-7
<i>bon</i>	2-2-4	3-3-3	4-4-5	6-6-5	7-7-6
<i>quoi</i>	2-2-3	4-2-4	4-2-8	5-5-4	6-5
<i>voilà</i>	3-3-4	3-5-4	4-6-4	6-6-5	7-8-5
<i>vous savez</i>	3-4-3	5-5-3	7-7-1	7-7-6	1-8-7

Lecture du tableau (ex. *alors* cluster 1)

- 1-3-2 signifie qu'un des patrons prototypiques de *alors* est :
  - niveau tonal du contexte gauche : 1
  - niveau tonal de la dernière syllabe de *alors* ([lɔʁ]) : 3
  - niveau tonal du contexte droit : 2

Tableau 12. Classification et propriétés prosodiques des PEN *alors*, *ben*, *bon*, *quoi*, *voilà*, *vous savez*. (Bartkova et al. 2016b [25], p. 462)

Au-delà de l'intérêt descriptif pour chaque PEN, la question est de savoir si ces cinq sous-classes correspondent à des fonctions particulières, soit transversales (un même patron pour plusieurs PEN qui renverraient à une même fonction), soit intrinsèques (plusieurs patrons pour une même PEN qui renverraient à plusieurs fonctions de cette

217 L'arbre utilisé est le J48 de la boîte à outils de *Weka* (Eibe et al. 2016), interface sélectionnée pour le travail statistique dans le cadre de cette collaboration.

PEN). Même si certaines hypothèses ont pu être faites, elles sont restées limitées. Par exemple, pour *quoi*, on remarque un niveau souvent bas pour la PEN, et un plateau ou une pente descendante depuis le contexte gauche, ce qui pourrait être corrélé avec une fonction conclusive ou parenthétique. Pour *vous savez*, c'est l'inverse, le niveau tonal de la PEN est relativement haut et il y a un plateau haut ou une pente montante depuis le contexte gauche.

A ce stade, et en se limitant aux paramètres relevés, il n'est pas évident de faire correspondre une classe et une fonction. J'irais même plus loin, la démarche comporte certaines caractéristiques qui rendent les interprétations difficiles et les résultats discutables pour une généralisation : problèmes d'ordre matériel attendus relatifs par exemple à la qualité des données et à des erreurs d'alignement automatique ; définition du groupe intonatif, qui est un critère intéressant mais parfois flou ; prise en compte de la pause dans le contexte de la PEN ; contexte textuel trop limité et contexte situationnel trop pauvre pour offrir un cadre interprétatif suffisamment riche pour comprendre les fonctionnements fins des PEN, leurs points communs et leurs différences. Toutefois, cette première approche a permis d'explorer le domaine, d'affiner techniquement l'analyse automatique et de soulever des questions fondamentales qui seront exploitées dans des recherches futures.

### 3.2.3.2 De la classe aux fonctions pragmatiques

La recherche de liens entre les propriétés prosodiques et les significations des PEN a été approfondie et systématisée dans d'autres travaux (voir en particulier Bartkova et al. 2017a [27], Jouvét et al. 2017b [28], Lee et al. 2019b [31]) et dans le cadre du co-encadrement de la thèse de L. Lee (Lee 2021).

#### (i) Données

Le (pré)traitement des données était identique à celui présenté dans la section 3.2.3.1 (i), mais les données ont été augmentées et diversifiées pour que plusieurs genres soient représentés (contes, informations, interviews et conversations, interactions diverses). Ainsi, en plus de ESTER (EST) et TCOF (TCO), nous avons « pioché » dans les corpus suivants : French Oral Narrative (FRE) (contes lus), Corpus de Français Parlé Parisien (CFP) (entretiens), C-ORAL-ROM (parole spontanée plus ou moins formelle), Corpus de Référence du Français Parlé (CRF) (interactions variées), FLEURON (FLE) (interactions étudiants/personnel administratif), Corpus de Langue Parlée en Interaction (CLAP) (interactions variées), DECODA (DEC) (dialogues d'appels au service clientèle de la RATP), corpus de la thèse de M. Husianycia (HUS) (Husianycia 2011, p. 87-131) (réunions de travail), et extraits des corpus TUF<sup>218</sup> (TUF), VALIBEL (VAL) et OFROM (OFR)<sup>219</sup>. Ceci nous a permis d'atteindre 5,32 millions de mots.

Nous avons choisi de travailler sur *alors*, *bon* et *donc* (Jouvét et al. 2017b [28]), puis nous avons ajouté *quoi* (Bartkova et al. 2017a [27]) et *enfin* et *voilà* (Lee et al. 2019b [31]). Pour aller plus loin dans l'analyse sémantique, un sous-ensemble d'occurrences des mots étudiés a été sélectionné aléatoirement (50 à 100 occurrences pour chaque

218 TUF<sup>218</sup> : corpus recueilli dans le cadre d'un projet (2014) de la Tokyo University of Foreign Studies. Nous avons travaillé sur la partie française uniquement.

219 VALIBEL et OFROM introduisent des variations régionales (respectivement Belgique, Suisse) qui peuvent affecter la prononciation et la prosodie.

corpus), afin de rendre possible l'étape d'annotation manuelle pour différencier les emplois PEN et nonPEN et les différentes interprétations comme PEN.

**(ii) Analyses**

Ci-dessous, je synthétise quelques comparaisons en distinguant deux niveaux : (a) la répartition et la représentativité des données (fréquence des items dans les corpus, proportion d'emplois PEN vs nonPEN) et (b) la distinction de différentes fonctions/interprétations comme PEN (conclusion, suspension, hésitation, etc.)

**(a)** Le traitement automatique a rendu possibles le calcul précis de la taille des données et le décompte des occurrences des items étudiés. Ci-dessous, le Tableau 13 reprend les résultats pour *alors* (pour le tableau complet, voir Bartkova et al. 2017a [27], tableau 1 p. 13). Sur un total de 5,32 millions de mots, *alors* est le plus représenté des items étudiés (0,6%, contre 0,1% pour les autres). Si l'on se penche de plus près sur ses emplois, l'annotation manuelle a permis de mettre en évidence qu'il fonctionne comme PEN dans 69 % des cas. Mais il existe une forte disparité selon les corpus et les genres de discours<sup>220</sup> (par exemple, dans le corpus French Oral Narrative, *alors* ne fonctionne comme PEN que dans 23 % de ses usages contre 93 % dans FLEURON).

Corpus	Contes	Infos	Interviews, conversations, ...					Interactions, ...					total		
	FRE	EST	CFP	COR	CRF	TUF	VAL	CLA	FLE	TCO	OFR	DEC		HUS	
Nb. mots. (millions)	0.14	1.82	0.41	0.22	0.38	0.58	0.25	0.02	0.03	0.36	0.28	0.65	0.17	5.32	
<i>alors</i>	Fréq. occ.	0,6%	0,2%	0,4%	0,4%	0,4%	0,2%	0,3%	0,2%	0,5%	0,4%	0,5%	0,8%	0,4%	---
	Nb. annot.	100	178	88	87	92	87	82	38	74	77	91	66	81	1144
	Fréq. DP	23%	56%	78%	76%	68%	68%	71%	58%	93%	77%	84%	80%	88%	69%

*Rem.*  
DP = PEN

Tableau 13. *Alors* : taille des corpus, fréquence totale, occurrences annotées et fréquence d'emploi comme PEN

Afin de tester si les paramètres prosodiques étudiés étaient suffisants pour prédire la distinction PEN vs nonPEN, l'analyse a été complétée par des expériences de classification automatique, comme cela a été décrit dans la section 3.2.3.1(ii), et une approche de type réseau de neurones<sup>221</sup> sur *alors*, *bon* et *donc* uniquement pour des raisons quantitatives (Jouvet et al. 2017b, p. 40-41).

**(b) Différentes fonctions des emplois PEN**

L'annotation manuelle a aussi été l'occasion d'aller plus loin dans l'interprétation des emplois PEN, chaque item étant a priori considéré comme polyfonctionnel en tant que PEN. Une première revue de la littérature dans le domaine avait permis d'isoler au préalable certaines fonctions, mais, en pratique, l'application aux données n'a pas été simple et a nécessité plusieurs aller-retours, plusieurs discussions et passes d'annotation. De plus, le contexte linguistique des données étant très limité, certaines fonctions sont difficiles à identifier. De mon point de vue, c'est là que se jouent une partie de la validité des résultats obtenus, une partie de la prédictibilité des fonctions comme PEN sur base

<sup>220</sup> Cette perspective, que l'on pourrait qualifier de sociolinguistique, a finalement été peu exploitée et elle mériterait selon moi d'être davantage fouillée.

<sup>221</sup> Les expériences ont été menées au LORIA, au moyen de Keras : <https://keras.io>

uniquement prosodique et, à un niveau général, une partie de la définition sémantique et pragmatique des PEN. Néanmoins, même si cette étape du travail pourrait être améliorée, les résultats de l'annotation donnent un premier aperçu de la polyfonctionnalité évoquée.

*Alors, bon* et *donc* ont servi de laboratoire pour l'identification et l'annotation de ce qui a été nommé « fonction pragmatique » (Bartkova et al. 2017a [27], p. 14, Jouvét et al. 2017b [28], p. 37). Les labels renvoyant aux fonctions ont subi plusieurs réorganisations, car elles ont été formulées au fil de l'annotation et de la répétition des configurations. Par exemple, pour *alors*, quatre fonctions/labels ont d'abord été utilisé(e)s : suspension d'une activité en cours ou mise en attente (ex. 103), introduction ou réintroduction d'un thème (ex. 104), conclusion ou acceptation d'une idée (c'est dans ce cas-là un équivalent de *bon*) (ex. 105), et gestion de l'interaction (ex. 106)<sup>222</sup>.

- (103) L1 – non ce n'est pas celui-ci  
 L2 – il faut que j'aïlle sur quoi alors ?  
 L1 – sur le site de la RATP  
 L2 – alors je suis sur le site de la RATP  
 L1 – RATP  
 L2 – RATP ok  
 L1 – alors ensuite vous cliquez... (DEC 20091112\_RATP\_SCD\_0711)
- (104) les forces régulières les forces loyalistes vont mettre le paquet sur Boauké (pause)  
alors la question qui qui se pose à la mi journée c'est de savoir qui... (EST 20020919\_1400\_1500\_rfi)
- (105) en achetant tout simplement des produits vous savez étiquetés satisfait ou remboursé alors c'est une gestion mais ça marche il l'a prouvé il a rempli son frigo... (EST\_19991026\_0700\_0800\_inter)
- (106) L1 – ... et vous pensez l'avoir perdu où madame ?  
 L2 – alors euh j'ai deux endroits possibles alors je sais que je l'ai passé à le au métro... (DEC\_20091112\_RATP\_SCD\_0989)

Pour terminer sur cet aspect d'attribution de labels à des usages en corpus, je fais deux remarques. Premièrement, les labels correspondent à des fonctions qui sont à la fois sémantiques et pragmatiques et dont l'interprétation demeure relativement intuitive. Une description plus fine, peut-être plus « formelle », de ces classes d'usages serait un apport souhaitable. Deuxièmement, l'annotation met en évidence la capacité combinatoire des PEN avec d'autres MD, connecteurs ou PEN. Certaines PEN sont plus combinables que d'autres, c'est le cas de *bon* qui a été étiqueté comme « particule complexe » dans 35 % de ces occurrences<sup>223</sup>.

En plus du développement de collaborations à plusieurs niveaux, les travaux mentionnés dans cette section m'ont été utiles, car ils ont permis : l'exploration systématique de données authentiques variées, l'utilisation ou la mise en place d'outils pour leur annotation automatique et semi-automatique, ainsi qu'une réflexion sur les modèles statistiques utiles pour la reconnaissance de classes correspondant à des patrons prosodiques spécifiques. Même si la description sémantique et pragmatique des mots

<sup>222</sup> Dans un second temps, ce sont les labels *conclusion*, *hésitation*, *introduction*, *réintroduction* qui ont été utilisés.

<sup>223</sup> C'est un des enjeux du projet CODIM dans lequel une tâche de représentation de la signification des MD et de leur combinaison est prévue (voir section 4.1.1, p. 159).

étudiés est restée, pour moi, un peu rapide, elle a eu le mérite de soulever de réels problèmes sur la catégorisation en général et de montrer la difficulté de « dire » la signification des MD en particulier.

### 3.2.4 Combinaisons de marqueurs

La question de la combinaison des MD est intervenue assez tôt dans mes travaux, mais ne constitue la cible de l'analyse que récemment. Je donne quelques exemples des premières mentions à travers l'exemple de *coudonc* (3.2.4.1), mais me concentre sur les études récentes à travers l'exemple de *mais enfin* (3.2.4.2).

#### 3.2.4.1 Coudon(c) : *indice de réalisme langagier*

Dans mes travaux sur la variation linguistique (voir section 1.2.1.1(i), p. 28), j'ai utilisé les MD comme des traceurs parmi d'autres du phénomène qui m'intéressait alors : la représentation de l'oralité populaire québécoise (OPQ) (voir le relevé (i)(a) p. 126). Dans ce cadre, sans distinguer emplois connecteurs et emplois PEN, j'ai procédé à un relevé et à une étude des valeurs de quelques MD dans les corpus *fricapop* et *tremblay* pour voir si les formes retenues par Michel Tremblay étaient crédibles et efficaces (Dagnat 2006 [6], p. 257-298). Les MD étudiés étaient *coudon(c)*, *entéka*, *fait que*, *hostie*, *là* et *par exemple* dont certains emplois sont assez représentatifs du français québécois. Dans cette liste, la forme *coudon(c)* est en elle-même particulière, dans la mesure où elle renvoie à la combinaison et à la concaténation de *écoute(z)* et *donc*. *Coudon(c)* est une graphie conventionnelle depuis les années 1970 qui n'est pas propre à l'écrivain puisqu'on la trouve chez les linguistiques (voir Dostie 2004, p. 81-92 et 194-204, Laurendeau 1985, Léard 1991). Ceci va dans le sens d'une pragmatization et d'une lexicalisation bien avancées, comme le note Dostie (2004, p. 83). Il est admis que *Coudon(c)* vient de la combinaison de *écoute(z) donc*, mais l'usage n'en fait pas une simple variante complexe de *écoute(z)*. La forme concaténée s'est autonomisée et a des usages propres qui correspondent à un fonctionnement de type PEN. *Coudon(c)* assume une fonction expressive et ne réclame pas nécessairement un co-énonciateur, autrement dit, il « aurait perdu en partie la prise en charge naturelle de l'autre (l'adresse à 'tu') lui venant de sa forme impérative initiale » (Dostie 2004, p. 104). Les corpus utilisés présentent la forme lexicalisée et la différencient des formes impératives « classiques » *écoute/écoutez donc*. J'ai trouvé à peu près le même nombre d'occurrences : 5 dans le corpus *fricapop*, graphiées *coudon* par les transpositeurs (le [k] n'était pas prononcé par les locuteurs) et 6 dans le corpus *tremblay*, avec les graphies *coudonc*, *'coudonc* et *'Coudonc*. A l'époque, je n'ai pas cherché à analyser le degré de compositionnalité de la combinaison, je me suis contentée de montrer que les deux grands emplois de *coudon(c)* étaient bien représentés dans la parole authentique et dans la parole représentée. Je les illustre sous (i) et (ii).

(i) Premièrement, les emplois impliquant l'interaction.

Dans (107), *'coudonc* marque la réaction du locuteur au silence de son fils Claude. Dans (108), il sert à marquer l'énervement et à appeler une réaction de l'allocutaire. Enfin, en (109), *coudonc* intervient avant tout comme phatique et/ou comme marqueur de

hiérarchie discursive dans un propos de L2 entrecoupé d'insertions de l'enquêteur L1<sup>224</sup>.

(107) Alex I – J'ai conduit tellement longtemps pis sur des routes tellement épouvantables que j'ai encore tout ça dans les bras... même après mon bain...  
(*Silence*) Aie, 'coudonc, c'est pas avec toé que j'vas avoir une conversation qui a de l'allure à soir, hein ? Parliez-vous de moé quand chus rentré, là ? (*tremblay* LVM, p. 56)

(108) Nana – Le monde l'entendent, aussi, quand a' crie, mais y pensent que c'est un fantôme ! T'as treize ans, pis tu sais pas lire ? Y pensent que c'est le fantôme de la Dame de Couette-Couenne ! Y'a une chanson, là-dessus, dans le livre, pis toute ! En as-tu passé des bouttes, 'coudonc ?  
Le Narrateur – Ben non. (*tremblay* EF, p. 27)

(109) L2 – j'ai des défauts ## mais là ils te donnent ça à choisir ou l'adultère ## qu'est-c'est- ' tu prends ? Saint-Chrême ##  
L1 – le moins pire  
L2 – baf ## adultère ## <INS>oui</INS> le plus le-fun ## <INS> oui <RIRE/></INS> coudon ## tanq: </RIRE> hein ? tant qu'à te faire pincer pour de quoi prends l'adultère ## [...] (*ficapop* O, 7-9)

(ii) Deuxièmement, les emplois n'impliquant pas directement l'interaction, c'est-à-dire où *coudon(c)* fonctionne dans le discours d'un seul locuteur, sans faire directement référence à l'allocutaire. Dans (110), un équivalent pourrait être *d'ailleurs* et, dans (111), *alors* ou *ma foi* fonctionneraient également.

(110) Nana – Si y'avait parlé en français à l'autre, y'a ben dû y parler en anglais à lui ! Pis, en espagnol pour le roi d'Espagne. Pis en italien... 'coudonc, y'a-tu un roi, en Italie ?  
Le Narrateur – Y'en connaît, des langues ! (*tremblay* EF, p. 36)

(111) L1 – puis: vous trouvez que: à cette époque là: les jeunes en général # enfin les jeunes de votre âge là: ?  
L2 – hum ## c'est: c'est pas la vie d'aujourd'hui ## ça: ça: [...]  
L1 – qu'est-ce-qui: ?  
L2 – il-y-a un gros changement : un gros changement dans la vie # assez que moi : je peux pas rien dire à mon dernier moi là ## il : je l'ai laissé aller # celui-là ## j'ai dit « envoie ## si la vie est de même # coudon ## » mais les autres c'est pas : c'était marié puis tout ça ## ça fait qu'ils ont : resté[sic] plus tranquille[sic] (*ficapop* F, 28-33)

Mes conclusions portaient essentiellement sur le caractère représentatif de cette PEN et sur le travail phonographique du dramaturge. Je n'ai pas cherché à analyser quelles valeurs de *écoute(z)* et de *donc* se combinaient ni à quelle époque la combinaison s'est spécialisée. Parmi les autres MD étudiés, certains présentaient des emplois comme connecteur et comme PEN (*par exemple, entéka, là, fait que*). Leur étude m'a permis de prendre conscience de leur importance pour l'organisation et l'ancrage situationnel du discours, de la difficulté parfois de discriminer connecteur et PEN, et de l'intérêt qu'il y a à envisager la polyfonctionnalité et la pragmatization en lien avec la polysémie,

224 Même si le spectre des interprétations sémantiques paraît ici assez large, on pourrait lui allouer le sens d'un « ben » marquant l'évidence du propos ; interprétation que le contexte renforce puisque *coudon(c)* est suivi (repris) par un « tant qu'à + infinitif ».

plutôt que l'homonymie, c'est-à-dire à « sauver » un minimum de liens sémantiques entre les différentes acceptions.

### 3.2.4.2 Mais enfin : *construction par association*

Plus récemment (Dagnat 2021b [40]), j'ai étudié la cooccurrence *mais enfin* en français contemporain écrit et oral, comme dans (112), (113) et (114) en me concentrant sur le mode de combinaison des deux termes.

(112) Mais enfin, bon dieu, qui aurait pu croire un truc pareil ? (D. Pennac, *Monsieur Malaussène*, 1995)

(113) quand c'est large au départ c'est un peu plus long mais enfin ça va venir à mon avis ça va se refermer aux alentours de deux ans (TCOF, *Pédiatrie\_lam\_08*)

(114) Vous en doutez encore ? M'enfin ! Regardez autour de vous, voyez ? voyez ? Il est là, partout ! (L. Lang, *Les Indiens*, 2001)

L'approche diffère de celle sur le français québécois. Elle emprunte aux autres réflexions en ce sens qu'elle repose sur un listage systématique des fonctions sur la base d'exemples authentiques, mais elle s'en éloigne aussi pour deux raisons. Première raison, parce qu'il s'agit d'une expression complexe, combinaison de deux MD. Deuxième raison, parce que les arrière-plans théoriques et méthodologiques mêlent des aspects qualitatifs et des aspects quantitatifs en lien avec la question de la compositionnalité, les mesures d'association et la notion de collostruction. Je reprends les questions, les étapes et les conclusions principales de ce travail.

Chaque composant a déjà fait l'objet de plusieurs études<sup>225</sup>, en synchronie et en diachronie, et, pour les deux, il a été mis en évidence un processus de pragmatization, c'est-à-dire que des emplois pragmatiques, notamment expressifs et interactionnels, existent déjà pour *mais* et *enfin* séparément. On trouve peu d'études systématiques sur la combinaison des deux (voir Razgoulieva 2002 et quelques remarques dans les travaux ciblant *enfin*). Chez Razgoulieva, la perspective est ouvertement compositionnelle, c'est-à-dire que la combinaison ne crée pas de nouvelle signification, elle combine certaines des valeurs de *mais* (adversatif) et *enfin* (temporel et reformulatif), essentiellement dans des contextes monologiques écrits et inter-propositionnels.

(i) La première étape a comme objectif de lister ces différentes valeurs et à mettre en évidence les contraintes combinatoires dans des contextes écrits et oraux, monologiques et dialogaux<sup>226</sup>. Toutes les configurations ne sont pas possibles. Ainsi, pour *mais*, j'ai retenu des emplois inter-propositionnels (plutôt monologiques) où la conjonction lexicalise une relation de discours de type **Contraste**, **Parallèle** ou **Concession**, et des emplois dit pragmatiques (plutôt dialogaux) où l'opposition n'a pas lieu au niveau des états de choses, mais au niveau épistémique et au niveau des actes de langage (voir Dagnat 2022a [42], ex. (4) à (12)). Pour *enfin*, je me suis inspirée des emplois mentionnés chez Buchi & Städler (2010) qui proposent une revue assez systématique

225 Pour *mais*, voir en particulier : Bruxelles et al. (1976), Ducrot et al. (1980, p. 93 sq) mais aussi Sweetser (1990) et Winterstein (2012) sur *but*. Pour *enfin*, voir Bertrand & Chanet (2005), Buchi & Städler (2010), Cadiot et al. (1985), Hansen (2005), Nêmo (2000) et Petit (2010).

226 Pour l'écrit, principalement *Le Monde*, pour l'oral CORPAIX, CRFP, DÉCLICS, ESTER, Fra80 J'ai aussi utilisé ESLO2 pour l'oral et le corpus contemporain de Frantext (36 396 589 mots) pour l'écrit, mais pas pour les mesures d'association pour le moment).



des autres études. Ainsi, j'ai repris les valeurs suivantes : *enfin* temporel, aspectuel, épistémique, récapitulatif, énumératif, rectificatif, performatif, réprobateur (voir Dagnat 2022a [42], ex. (22) à (27)). J'ai regroupé les valeurs épistémique, récapitulative et énumérative qui ont en commun la perception de la fin d'une liste. J'ai également ajouté la valeur de modération pour des cas comme (115) où le locuteur accepte ce qui est dit avant ou ne le remet pas en cause directement. On peut lui adjoindre *tout de même* ou *quand même* et il a une valeur adversative.

(115) Il n'est pas Général, enfin il est Colonel.

Il est ressorti de cet examen que *mais* et *enfin* pris séparément présentent des emplois variés (connecteur et PEN) mais que demeure plus ou moins fortement, en fonction du domaine (à la Sweetser) sur lequel ils portent, l'idée d'une opposition pour *mais* et de l'aboutissement de quelque chose pour *enfin*.

(ii) La deuxième étape consiste à énumérer les possibilités et impossibilités combinatoires. Ci-dessous, le Tableau 14 synthétise les contraintes et précise les cas particuliers qui demandent un emploi contigu de *mais* et *enfin* (cas avec astérisque, cases grisées). Les incompatibilités ont fait l'objet de plusieurs remarques (voir Dagnat 2022a [42], p. 6-9).

ENFIN MAIS	temporel i	aspectuel ii	épistémique, récapitulatif, énumératif iii	rectificatif iv	modérateur v	performatif vi	réprobateur vii
a. correctif	-	+	-	-	-	-	+
b. parallèle	-	+	-	-	-	-	+
c. concessif direct	+	+	-	-	-	_*	+
d. concessif indirect	+	+	-	-	+	_*	+
e. événement interne	-	+	-	-	-	-	+
f. événement externe	-	+	-	-	-	-	+
g. actes de langage	-	-	-	-	-	-	+

Tableau 14. Possibilités combinatoires des valeurs de *mais* et *enfin*<sup>227</sup>

La question de fond qui m'intéresse dans cette étude est celle de la compositionnalité ou non de la combinaison où *mais* et *enfin* sont contigus, pour chercher, sinon à expliquer, au moins à mettre en évidence les prémices d'une lexicalisation et d'une pragmatization pouvant aller jusqu'au *m'enfin* de l'exemple (114). La combinaison contiguë est-elle une simple juxtaposition ou un figement (*frozen concessive collocation*, Hansen 2005) avec des valeurs propres ? Dans le premier cas, les valeurs de *mais enfin* contigus devraient correspondre strictement aux cases positives du tableau ci-dessus. La fréquence d'appariement serait alors un simple agencement compositionnel, une collocation, peut-être en voie de lexicalisation. Dans le second cas, *mais enfin* contigus

227 – signifie que les valeurs ne sont pas compatibles, + signifie qu'elles le sont, \* signifie que les valeurs sont compatibles uniquement pour *mais enfin* contigus.

devaient présenter une ou des valeurs qui ne s’y trouve(nt) pas. Les valeurs spécifiques iraient alors dans le sens d’une construction au sens fort, ou locution, c’est-à-dire un phrasème dont l’interprétation n’est pas une simple combinaison de l’interprétation de ses parties (Mel’čuk 2013). Concrètement, les possibilités de combinaison pour les emplois contigus et non contigus se correspondent, sauf dans le cas de *mais* concessif et *enfin* performatif, qui ne peuvent se combiner qu’en étant contigus, comme le montrent (116/116’) et (117/117’) :

(116) Mon entrecôte est trop cuite, mais enfin (je vais la manger).

(116’) # mon entrecôte est trop cuite, mais je vais enfin la manger.

(117) La situation est difficile, mais enfin (il faut faire avec).

(117’) ?La situation est difficile, mais il faut enfin faire avec.

Cette particularité ne remet pas en cause le caractère compositionnel de l’agencement. L’obligation de co-présence vient du fait que les deux items demandent à être placés à l’initiale de la proposition (qui d’ailleurs peut-être elliptique), et ce même quand ils sont seuls, le premier à cause de sa catégorie, le second car il laisse entendre que la proposition qui précède est acceptée à regret et teinte la proposition qui suit de résignation. À ce stade, il est difficile d’appréhender la combinaison *mais enfin* autrement que comme un agencement compositionnel. Tout au plus peut-on lui reconnaître une fréquence d’apparition qui pourrait être associée à un processus de lexicalisation<sup>228</sup>.

**(iii)** La troisième étape est une approche quantitative de la notion de construction. *Mais enfin* n’est donc pas une expression constructionnelle au sens fort du terme. Cependant, la notion de construction a évolué et n’exige plus systématiquement aujourd’hui le critère de la non-compositionnalité. En effet, le champ des grammaires dites de construction n’est pas uniforme (Hilpert 2014, p. 1-24, Hoffmann & Trousdale 2013). Les phénomènes décrits sont aujourd’hui très variés et l’analyse des marqueurs de discours, en particulier les marqueurs complexes, y trouve sa place. Certes, un des paramètres fondateurs et récurrents dans la définition d’une construction a d’abord été son caractère non compositionnel (Goldberg 1995, p. 4), mais, plus récemment, la fréquence de l’appariement de deux items est entrée en ligne de compte, même en cas de prédictibilité de l’expression complexe qu’ils composent (Goldberg 2006, p. 5). Le taux de répétition d’une construction dans ce sens quantitatif, parfois appelée *collostruction*, est même un des chemins reconnus de son acquisition comme bloc (Bybee 2010, p. 97 sq., Desagulier 2015, Gries 2013, 2019, Hilpert 2014, p. 66, Stefanowitsch & Gries 2003).

#### **(a)** Mesures d’association

Pour évaluer la fréquence de collocation de *mais* et *enfin* (contigus dans mon cas), qui peut être en lien avec le degré de figement, il est possible d’utiliser ce que l’on appelle des mesures d’association, dites aussi mesures d’attraction. Intuitivement, cela permet de mettre en évidence que la cooccurrence n’est pas le simple fait du hasard. Il faut disposer de données relativement propres et uniformes afin de pouvoir catégoriser les items et les trier automatiquement. J’ai sélectionné plusieurs corpus oraux et écrits, pour lesquels j’ai eu accès à l’intégralité des données<sup>226</sup>. Ils ont été étiquetés avec

228 Une analyse plus approfondie de la forme concaténée *m’enfin* (ou *menfin*) montrerait peut-être une spécialisation de certains emplois à des contextes ou des valeurs particuliers, qui conduirait sur la voie de la non-compositionnalité. L’étude reste à faire.

Unitex/GramLab<sup>229</sup>, qui permet de prendre en compte les multi-mots, à la différence d'autres parsers comme Treetagger<sup>230</sup> ou SpaCy<sup>231</sup>. L'objectif était de pouvoir tester la force d'attraction entre *mais* et *enfin*, avec *mais* comme pivot (*node*) et *enfin* comme associé (*collocate*), en le comparant aux autres candidats associés parmi les catégories conjonction, adverbe et interjection. Il n'est pas possible de reproduire ici la totalité des résultats, ni de détailler les 16 mesures effectuées (voir Brézina 2018, Desagulier 2017) mais le tableau complet est disponible en ligne<sup>232</sup>.

### (b) Fréquence et exclusivité

Comme le note Brézina (2018, p. 66-75), ces mesures sont plus ou moins sensibles à deux traits des associations : la fréquence et l'exclusivité.

Les calculs fondés sur la fréquence utilisent diverses fonctions appliquées à la fréquence *observée*, c'est-à-dire le nombre effectif de pivots, d'associés ou de combinaisons impliquant l'associé et/ou le pivot, et à la fréquence *attendue* dont le calcul dépend de la présence ou non du pivot et de l'associé.

L'exclusivité estime la tendance pour le pivot et l'associé à figurer davantage ensemble que séparément.

Je cible simplement trois mesures : les mesures MI3, G et DPF<sup>233</sup>. MI3 et G sont sensibles à la fréquence et à l'exclusivité, G ayant en outre l'avantage d'être sensible à la catégorie grammaticale du pivot (conjonction de coordination pour *mais*) et de l'associé (adverbe ou interjection pour *enfin*). DPF estime la probabilité de l'associé d'apparaître à droite du pivot, autrement dit la probabilité que *mais* soit suivi de *enfin*.

### (c) Résultats

#### – Mesure MI3

*Enfin* arrive en tête dans CORPAIX et Fra80, il arrive au 2<sup>ème</sup> rang dans CRFP, au 9<sup>ème</sup> dans DÉCLICS et au 32<sup>ème</sup> dans ESTER et dans *Le Monde*. Dans tous les cas, on voit que les valeurs de *enfin* sont dans les 25% supérieurs (la valeur de la mesure est supérieure au troisième quartile Q3).

#### – Mesure G

On obtient les rangs suivants pour *enfin* comme associé de *mais* : 1<sup>er</sup> rang dans CORPAIX et Fra80, 2<sup>ème</sup> dans CRFP, 14<sup>ème</sup> dans DÉCLICS, 33<sup>ème</sup> dans ESTER et 34<sup>ème</sup> dans *Le Monde*. Les valeurs de *enfin* sont supérieures à Q3 pour DÉCLICS et *Le Monde*, et quasiment égales à Q3 pour ESTER.

#### – Mesure DPF

Parmi les associés de *mais*, *enfin* apparaît aux rangs suivants : 1<sup>er</sup> dans Fra80, 2<sup>ème</sup> dans CORPAIX et CRFP, 6<sup>ème</sup> dans DÉCLICS, 15<sup>ème</sup> dans ESTER et 18<sup>ème</sup> dans *Le Monde*. Les valeurs de *enfin* sont toujours supérieures à Q3 sauf pour *Le Monde* où il est quasiment sur la médiane.

Les données des autres mesures vont dans le même sens. Il apparaît évident qu'il y a une attraction marquée entre *mais* et *enfin* dans les corpus oraux en interaction

229 <https://unitexgramlab.org/fr>. Pour cela, j'ai bénéficié du travail effectué dans le cadre du stage de A. H. Dahou (2021).

230 <https://cis.uni-muenchen.de/~schmid/tools/TreeTagger>

231 <https://spacy.io>

232 [http://mathilde.dagnat.free.fr/index\\_fichiers/DARGNAT-Langages-2020-Association\\_Measures.tsv](http://mathilde.dagnat.free.fr/index_fichiers/DARGNAT-Langages-2020-Association_Measures.tsv)

233 MI3 : *Mutual Information with a cubed numerator* ; G : *Lexical Gravity* (Daudaravičius & Marcinkevičienė 2004) ; et DPF : *Delta-P-Forward*. Dans le tableau en ligne : colonnes T-U-V pour MI3 ; colonnes AR-AS-AT pour DPF ; colonnes AX-AY-AZ pour G.

(CORPAIX, CRFP, DÉCLIC, Fra80). Cela est moins net pour ESTER et *Le Monde*, mais quand même sensible pour les mesures MI3 et G. Cette attraction est directionnelle : *mais* attire *enfin*, mais pas l'inverse, comme le montre la mesure symétrique de DPF dans l'ensemble des 16 mesures (mesure *Delta-P-Backward*<sup>234</sup>), sauf pour *Le Monde*, où il n'y a pas de prévisibilité particulière de *enfin* à la suite de *mais*.

Bien que cette analyse ne soit pas complètement aboutie, elle m'a permis d'explorer une voie quantitative que je n'avais pas empruntée jusque là et que je compte développer en l'appliquant à d'autres combinaisons de MD et de PEN (pour une première approche, voir Dagnat & Jayez (2021b [40])). Du point de vue théorique, la notion de collostruction revisite avec pertinence la notion de construction en lui donnant le relief de l'usage. Une des hypothèses concernant les combinaisons de MD et de PEN est que la force d'attraction ou d'association peut être le symptôme d'un processus de routinisation de type pragmatique associé à l'acquisition de valeurs sémantiques propres à l'expression complexe (à sa non-compositionnalité) et parfois à sa lexicalisation.

### Synthèse 3

La diversité des recherches sur les MD constitue à la fois une cause et un effet. Elle explique en partie la difficulté d'étudier les MD, car la variété des approches et la juxtaposition de points de vue très généraux et d'études très spécifiques rendent difficile d'appréhender la littérature dans son ensemble. Cette situation résulte probablement de la diversité intrinsèque des MD, qui ne forment pas une classe homogène. Dans cette partie, j'ai délibérément mis l'accent sur des notions ou des méthodes qui me semblaient suffisamment pertinentes pour « attraper » des propriétés partagées par des sous-ensembles importants de MD (connecteurs et PEN).

Je ne prétends pas avoir proposé une nouvelle taxinomie qui permettrait de catégoriser tous les MD d'une façon commode et transparente. Néanmoins, j'ai fait apparaître des propriétés transversales comme le caractère procédural, l'indexicalité, la nature de la portée et la relation au contenu propositionnel. Je retrouve ainsi d'une manière plus explicite un certain nombre de caractéristiques souvent mentionnées mais peu développées dans les différentes tentatives de définition.

Le caractère procédural recouvre la dépendance contextuelle et la difficulté de formuler la signification des MD, notamment des PEN (leur aspect « indicible »).

L'indexicalité renvoie à la fois au côté procédural (construction d'un référent) et au comportement présuppositionnel des MD au regard de la projection, de l'effet local et de la déplaçabilité.

La portée des MD est très riche, elle ne se cantonne pas aux états de choses. Elle reflète pour les connecteurs la flexibilité des relations de discours, et pour les PEN la subtilité de l'inscription du locuteur dans le discours (par ex. aspects épistémiques, interactionnels et émotionnels).

La relation au contenu propositionnel rejoint la question des présuppositions et plus largement celle de la hiérarchie/collaboration entre les niveaux de signification. On a

234 Dans le tableau en ligne, voir colonnes AU-AV-AW.

vu que certains connecteurs (causaux, temporels) sont intégrés au contenu propositionnel alors que les PEN en demeurent indépendantes. Toutefois, lorsqu'un MD porte sur un acte de langage, il porte aussi « indirectement » sur le contenu propositionnel. Ainsi, dans le cas de *quand même* analysé comme particule modale, la portée est l'acte de langage et le contenu propositionnel est essentiel à l'interprétation.

Par ailleurs, le changement de portée et la relation au contenu propositionnel sont des facteurs essentiels de l'évolution diachronique (voir section 3.2.2.1) ainsi que des postures dans l'interaction (voir section 3.2.2.2).

Outre le problème de la classification, j'ai présenté des travaux en collaboration sur le profil prosodique des particules. L'enjeu était d'utiliser des traits prosodiques dans la détection supervisée du statut (PEN/nonPEN) et de la valeur sémantique (PEN polyfonctionnelles) de différentes PEN.

Enfin, j'ai abordé la question des combinaisons de MD, pour deux raisons. Il s'agit d'un phénomène fréquent et il amène à étudier le caractère plus ou moins compositionnel des combinaisons.

# 4

## Perspectives en guise de conclusion

*La sagesse est un parapluie troué : personne n'est à l'abri de la mode.*

Christian Bobin, *Les ruines du ciel*. Paris : Gallimard, 2009, p. 44.



Les perspectives présentées ci-dessous vont dans deux directions qui peuvent se rejoindre sur certains points. La première est celle d'une continuation et d'un approfondissement de mes réflexions sur les marqueurs discursifs (section 4.1). La seconde concerne la valorisation et l'enrichissement de ressources pour l'étude de la variation linguistique (section 4.2).

## 4.1 Marqueurs discursifs : combinaison et expérimental

Dans les quelques années à venir, je souhaite continuer mon travail sur les combinaisons de marqueurs discursifs dans deux directions principales. La première vise en priorité l'inventaire et l'analyse de combinaisons de MD (4.1.1), la seconde l'interaction des MD de cause et de concession (4.1.2).

### 4.1.1 Compositionnalité et combinaisons de MD : le projet ANR CODIM

J'ai commencé à aborder cette question dans deux articles récents sur *mais enfin* (Dagnat 2022a [42]) et plus généralement sur des associés de *mais* (Dagnat & Jayez 2021b [40]). Elle fait l'objet du projet d'ANR que je coordonne, l'ANR [CODIM](#) (COmpositionality and DIscourse Markers), qui été accepté en juillet 2022 et qui commencera début 2023<sup>235</sup>.

Il s'agit d'un projet interdisciplinaire (sémantique et informatique) dont le point de départ est le constat qu'il existe en linguistique deux tendances majeures dans l'approche des expressions complexes. L'une insiste sur la compositionnalité, c'est-à-dire l'élaboration de la signification d'une structure à partir de la signification de ses parties. L'autre met l'accent sur le caractère partiellement conventionnel de l'association entre l'expression et son sens en insistant sur le caractère non prédictible du sens global à partir de celui de ses parties. Les principaux représentants de ces deux tendances sont, d'une part, la sémantique compositionnelle et, d'autre part, les travaux sur les expressions multi-mots, la phraséologie et les CxG<sup>236</sup>.

La plupart des études sur cette double perspective n'abordent pas les combinaisons de MD. Cependant, la cooccurrence de MD comme *alors donc*, *alors du coup*, *mais enfin*, *ah bon*, *non/oui mais*, *écoute donc*, *et/mais pourtant*, *tiens voir*, *dis donc*, *ah mais non du coup*<sup>237</sup> est un phénomène très fréquent<sup>238</sup> sur lequel il existe finalement assez peu d'études en synchronie. Même si les cooccurrences font l'objet de remarques, la majorité des études se concentrent sur des MD isolés. Cependant, la question de leur

235 Outre l'ATILF, qui est le laboratoire porteur à travers moi, les deux partenaires scientifiques sont le LORIA à Nancy (responsable Y. Toussaint) et le LLF (responsable J. Ginzburg). Lancement du projet en janvier 2023.

236 Pour un panorama des approches en sémantique compositionnelle, voir Werning, Hinzen & Machery (2012, parties III et IV). Pour les expressions multi-mots : Constant et al. (2017). Pour la phraséologie, voir Kauffer & Keromnes (2019) et Mel'cuk (2013). Pour l'approche constructionnelle, voir Desagulier (2015) et Gries (2019).

237 *Ah mais non du coup* : exemple emprunté à une publicité pour *Midas*, juillet 2022.

238 Par exemple, dans le travail d'annotation autour de la thèse de L. Lee (2021), nous avons observé que *bon* apparaissait dans des combinaisons dans 35 % de ses occurrences comme PEN.



(non-)compositionnalité se pose, comme le montrent les réflexions de Hansen sur *ah bon* et *eh bien* (1998a, p. 44, p. 233 et p. 289), Razgoulieva (2002, p. 167) sur *mais enfin* ou Waltereit sur *bon ben* et *enfin bref* (2007, p. 104).

Le projet CODIM aborde cette question sous deux angles principaux : (i) un aspect quantitatif et (ii) un aspect sémantique, y compris prosodique (durée, pause et contour).

(i) L'aspect quantitatif concerne la « force d'association » entre les éléments entrant dans une combinaison de MD. Il est très difficile de se fier uniquement à son intuition dans ce domaine, ce qu'ont bien montré des études préliminaires (Dagnat 2022a [42], Dagnat & Jayez 2021b [40], Hafed Dahou 2021). Il est prévu d'attaquer ce problème avec deux types de méthode : des mesures d'association, qui sont des indicateurs statistiques de la tendance de deux ou plusieurs éléments à se présenter ensemble dans des corpus, et de l'apprentissage automatique, qui construit des modèles de prédiction des cooccurrences (par exemple en estimant la probabilité qu'un MD déterminé soit précédé ou suivi d'un autre MD). Les ressources pour ces deux approches sont des données massives issues de corpus écrits et oraux.

(ii) L'aspect sémantique consiste à étudier et à représenter les propriétés sémantiques et prosodiques de chaque item de la combinaison et les propriétés sémantiques et prosodiques de la combinaison elle-même. Je liste quelques problèmes que le projet rencontrera.

– L'identification des MD dans des données « brutes » et la polysémie de certains MD qui nécessitent a minima que l'on s'arrête sur une liste (voir ce qui a été dit plus haut à la section 3.2.1.2, p. 126).

– La distribution des différentes facettes sémantiques des MD polysémiques dans les combinaisons. Par exemple, des MD comme *alors*, *enfin*, *mais*, *bon* ont plusieurs fonctions en tant que MD : ces fonctions sont-elles toutes combinables avec d'autres MD ? Si oui, lesquelles ? Et dans quelle proportion par rapport à l'usage isolé ? Ceci peut avoir des conséquences sur la capacité de la combinaison à apparaître très fréquemment, à se lexicaliser et à développer un sens propre. Mon étude préliminaire sur *mais enfin* pose le problème de la signification globale de la forme orale réduite [mãfẽ], parfois graphiée *m'enfin* ou *menfin* et des fonctions de *mais* et *enfin* qui se combinent.

– La disponibilité et la qualité du signal sonore pour les données orales, ainsi que la possibilité d'aligner, de segmenter et d'annoter de manière quasi-automatique le signal pour un contexte suffisant autour des items étudiés. Pour l'analyse prosodique, dans mes travaux antérieurs, nous avons opté pour une fenêtre de cinq mots avant et cinq après le MD. Pour l'interprétation, il faut avoir accès à un contexte souvent plus large, car la portée des MD peut être assez variable.

– Le choix/développement d'un formalisme pour représenter de manière multidimensionnelle et hiérarchique la contribution sémantique des MD et de leur(s) combinaison(s).

A ce stade, en plus de la coordination générale, ma contribution principale portera sur l'analyse sémantique et prosodique des combinaisons de MD les plus fréquentes dans les corpus et des MD que ces combinaisons comportent (notamment à

travers une co-direction de thèse). L'objectif du projet est d'initier un processus de cartographie des MD complexes en français.

#### 4.1.2 Enchâssement de marqueurs de cause et de concession : oculométrie et potentiels évoqués

Ce projet concerne toujours les MD, et d'une certaine manière leur combinaison, mais dans une perspective différente. Il s'agit d'un projet expérimental auquel j'ai commencé à réfléchir à l'occasion d'une délégation CNRS<sup>239</sup> en 2019-2020. Les circonstances<sup>240</sup> ont fait que je n'ai pas pu réaliser les expériences comme prévu. Mon laboratoire de rattachement, l'ATILF, ne possède pas de plateforme expérimentale adéquate pour les réaliser, mais, depuis 2021, la MSH-Lorraine s'est dotée de la plateforme ECHO<sup>241</sup>, avec un poste d'eyetracking et un poste d'EEG que je pourrai utiliser<sup>242</sup>.

L'objectif des expériences proposées est d'enrichir la description de certains MD connecteurs en testant l'hypothèse d'une différence de « force », telle que par exemple Hybertie le suggère pour *donc* et *alors* (1996, p. 29-30)<sup>243</sup>. Elle explique la différence d'acceptabilité entre

(118) J'étais pressé, alors j'ai pris le sens interdit.

(119) # J'étais pressé, donc j'ai pris le sens interdit.

par le fait que *donc*, en introduisant une explication suffisante, introduirait une relation plus forte que *alors* entre les deux événements. *Alors* paraît être plus souple, dans la mesure où il a la capacité de se contenter d'un élément d'explication qui demande à être complété<sup>244</sup>. Un marqueur sera perçu comme fort quand il ne fonctionne qu'avec des prémisses correspondant à des conditions suffisantes. Un marqueur dit moins fort, comme *alors*, a un spectre plus large, il peut fonctionner aussi bien dans des raisonnements enthymématiques que dans des raisonnements stricts (comme le *si... alors* utilisé en mathématiques).

Dans certaines approches, par exemple Kehler (2002, p. 20-22), la conséquence (Result), l'explication (Explanation) et la concession (Violated Expectation ou Denial of Preventer) relèvent de la famille des relations cause-effet (voir Nazarenko 2000). En effet, ces relations communiquent l'existence d'une dépendance entre des événements/propositions. Dans tous les cas, on peut se demander si cette dépendance varie et s'il existe des items marquant cette variation. En d'autres termes, s'il existe des MD connectant plus ou moins fortement les événements/propositions reliés. Dans mon cas, je souhaite travailler sur les connecteurs *pourtant* et *quand même* marquant une relation de concession, comme dans l'exemple (120).

239 Lieu de la délégation : Institut des Sciences cognitives Marc Jannerod (Bron), L2C2 UMR 5304.

240 Période du covid et dissolution d'une partie du laboratoire.

241 <https://msh-lorraine.fr/nos-services/la-plateforme-echo>

242 Dans la perspective, entre autres, de ces expériences, l'ingénieur en place et moi avons d'ailleurs déposé une demande financement conjointe ATILF-MSH pour l'acquisition du logiciel et de la formation Tobii pour l'eyetracker. Cette demande a été acceptée en mai 2022.

243 D'autres approches, comme celle de Degand & Pander Maat (2003), cherchent à hiérarchiser l'emploi des connecteurs de cause, en fonction du degré d'implication du locuteur (*speaker involvement*).

244 Les analyses rhétoriques et argumentatives appelleraient cela un *enthymème* (voir Walton 2013, chap. 3), c'est-à-dire des schémas argumentatifs où il manque une prémisse.

(120) Paul n'avait pas beaucoup de temps, il est quand même/pourtant venu.

La dépendance se fait dans ce cas-là entre une proposition et sa conséquence attendue implicite (voir plus haut section 3.2.2, p. 134). Le fait que Paul n'ait pas beaucoup de temps laisse attendre le fait qu'il ne peut pas venir. Le contraste se construit sur la négation de la conséquence attendue (la négation de *il ne peut pas venir*). L'hypothèse que je souhaite tester est que *quand même* construit un contraste moins fort que *pourtant* avec la conséquence implicite. Il n'est pas évident de saisir cette différence sur des exemples comme (120) de structure « A *pourtant/quand même* B », car la souplesse de *quand même* lui permet d'assumer les fonctions de *pourtant*.

Si on fait l'hypothèse qu'une structure « A *pourtant* B » présente A comme une condition suffisante de non-B et que « A *quand même* B » présente A comme une des conditions possibles de non-B, on peut construire un environnement expérimental acceptable pour *quand même* mais trop souple pour *pourtant*. Concrètement, cela peut se faire en ajoutant dans un raisonnement une proposition qui crée un environnement enthymématique avec lequel *pourtant* fonctionnera mal voire pas. Le schéma de construction de tels environnements est représenté dans la Figure 22 et est illustré par l'exemple (121).

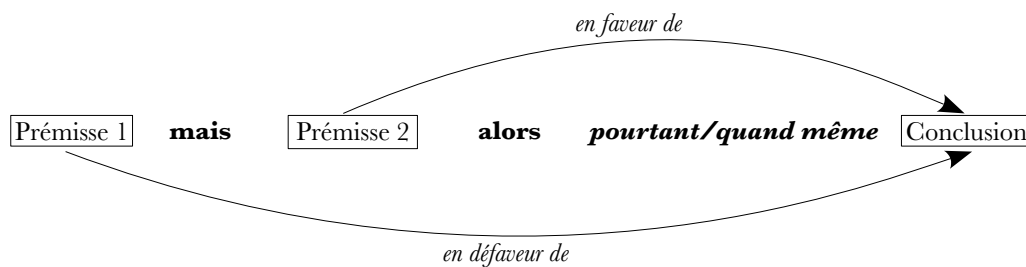


Figure 22. Structures des stimuli pour l'expérience sur *quand même* et *pourtant*

(121) a. Paul n'avait pas beaucoup de temps [P1], mais la réunion était très importante [P2], alors il est quand même venu [C].

b. % Paul n'avait pas beaucoup de temps [P1], mais la réunion était très importante [P2], alors il est pourtant venu [C].

La structure de la Figure 22 appliquée à l'exemple (121) présente *Paul n'avait pas beaucoup de temps* comme défavorisant la réalisation de la conclusion *Paul est venu*. En revanche, le *alors* présente la prémisse 2 *la réunion était très importante* comme favorisant la réalisation de la conclusion *Paul est venu*. Les connecteurs concessifs *pourtant* et *quand même* relient la prémisse 1 à la conclusion : *Paul n'avait pas beaucoup de temps* [P1], il est *quand même/pourtant* venu [C].

– Si *pourtant* demande à ce que P1 soit une condition suffisante, il présente la négation de la conclusion, c'est-à-dire « Paul n'est pas venu », comme difficilement contestable. Il y aurait donc une tension entre la présence de P2 et celle de *pourtant* qui peut engendrer une difficulté d'interprétation (ex. 121b).

– Si *quand même* est un connecteur moins fort qui demande seulement que P1 soit présentée comme une condition favorisant la négation de la conclusion, la tension devrait être moindre ou inexistante (ex. 121a).

On remarquera que ce dispositif exploite l'intuition fondamentale de Kehler, à savoir que la concession est étroitement liée à la causalité.

Du point de vue expérimental, le choix de l'eyetracking et de l'EEG se justifie par le fait que ces deux méthodes sont utilisées pour mettre en évidence des difficultés de traitement du signal.

L'utilisation de l'eyetracking repose notamment sur l'idée qu'une difficulté interprétative entraîne plus de fixations du regard sur la région problématique (ici la fin de la phrase à partir du mot testé) ainsi que des déplacements du regard plus nombreux à partir de ou vers la région problématique (voir van Gompel et al. 2007, partie 5).

L'utilisation de l'EEG se fera sur la base des mêmes stimuli que pour l'eyetracking. Les deux manifestations les plus fréquemment mentionnées sont la N400, qui correspond à une variation négative de potentiel dont le pic est observé en moyenne 400 ms après l'apparition du mot-cible, et la P600, qui correspond à une variation positive observée en moyenne dans un intervalle de 500-800 ms après le mot-cible. L'interprétation récente de la N400 fait ressortir le fait qu'elle est associée à une difficulté d'interprétation (Kutas & Federmeier 2011, p. 636). La P600 est quant à elle souvent associée à la difficulté de construire la cohérence discursive. En simplifiant, on pourrait dire que la N400 est une alerte sur une information inattendue et la P600 un signal du travail d'intégration de cette information dans un modèle cohérent (voir Brouwer et al. 2012).

Je fais l'hypothèse que les exemples du type de (121b) déclencheront une N400 et éventuellement une P600.

## **4.2 Catégorisation et représentation de la variation, suites**

Je souhaite également continuer à travailler sur la variation linguistique, la catégorisation et la représentation de variétés du français. À ce stade, se dégagent deux types de travaux qui supposent la constitution et l'exploitation de ressources particulières. La représentation et l'exploitation de l'oralité et d'autres dimensions de la variation liées à l'oralité dans les textes littéraires constituent un premier point (section 4.2.1), les ressorts linguistiques de l'humour constituent le deuxième point (4.2.2). Ces projets sont plus programmatiques que ceux sur les marqueurs discursifs, donc un peu moins documentés pour le moment, mais ils sont déjà engagés ou en discussion sur certains éléments.

### **4.2.1 L'oral représenté : théâtre, francophonie et littérature de jeunesse**

Ce point concerne principalement l'extension de ressources permettant l'étude de la variation linguistique dans les textes littéraires, en particulier la variation diamesique (la différence oral/écrit) en relation avec les dimensions diastratiques (les niveaux et registres) et diatopiques (variations régionales, francophonie, accents). Trois types de données retiennent mon intérêt : les textes de théâtre d'auteurs français, les textes relevant des aires dites francophones et les textes de littérature de jeunesse.

J'ai pour objectif d'étendre mon corpus de textes de théâtre en valorisant la base de données que m'a confiée C. Bernet suite à mon séjour de recherche à l'ENS de

Lyon à la fin de ma thèse et suite à son départ à la retraite. Cette base correspond en partie à un projet réalisé à l'INaLF<sup>245</sup> (projet Théâtre). Il apporte la dimension diachronique, car la base comporte des textes de théâtre allant du XVI<sup>ème</sup> siècle au XIX<sup>ème</sup> siècle. Concrètement, elle est composée de textes intégraux dans deux formats : un format .txt avec le texte brut et un format .xml de type TEI enrichi de balises de structure permettant notamment de coder les actes, les scènes, les personnages et les didascalies. En l'état, les textes sont répartis en trois groupes : le premier, intitulé *Théâtre classique*, comporte uniquement les textes de Corneille, Molière et Racine. Il s'agit d'un sous-ensemble du second groupe qui comporte des textes allant du XVI<sup>ème</sup> siècle au XIX<sup>ème</sup> siècle. Le troisième, intitulé *Pléiade*, constitue un ajout aux deux autres, il comporte 11 textes supplémentaires saisis, corrigés et balisés.

L'idéal serait de valoriser l'ensemble des textes, mais, dans un premier temps, j'envisage de sélectionner les textes les plus susceptibles de présenter des marques de variation linguistique (prioritairement les variétés « basses »), c'est-à-dire les textes relevant du genre comique et comportant des personnages populaires. Cela devrait me permettre, comme je l'ai fait dans mes premiers travaux, de travailler sur le réalisme des « parlures », sur leur rôle plus ou moins structurant au niveau textuel et sur la récurrence ou la nouveauté des procédés utilisés en fonction des périodes et des genres. Autrement dit, à travers l'utilisation des textes mentionnés, je souhaite rendre possible une stylistique diachronique de l'oralité populaire dans le théâtre comique français<sup>246</sup>.

Le projet concernant les deux autres ressources consiste à initier une intégration progressive et continue de textes littéraires francophones et de textes de littérature de jeunesse à une base comme la base *Frantext*<sup>247</sup>, pour laquelle des schémas d'annotation et une interface d'interrogation fonctionnels existent avec un certain nombre d'accords éditoriaux pour les textes encore sous droit. En effet, à ce jour, *Frantext* ne comporte que très peu voire pas de textes relevant de ces catégories, ce qui constitue un manque. Un tel enrichissement aurait de plus le mérite de diversifier cette base et de la décentrer d'une image de la littérature française majoritairement hexagonale et à destination du seul lectorat adulte.

La question de la littérature francophone soulève plus franchement celle de la variation diatopique et celle de la catégorisation d'une littérature différenciée de la littérature dite française et des critères de cette différenciation. On le sait, une telle catégorisation n'est pas politiquement et idéologiquement neutre, car elle est fortement liée à l'histoire coloniale. La sélection de textes nécessite donc une collaboration avec plusieurs collègues de littérature comparée spécialistes des différentes aires francophones (Afrique, Amérique du Nord, Antilles, Belgique, Maghreb, Suisse, etc.)

245 Institut National de la Langue Française (cet institut a été remplacé par l'ATILF).

246 L'analyse des marques d'oralité dans les textes du point de vue diachronique est une perspective de plus en plus présente. Je renvoie par exemple à des numéros assez récents comme celui de la revue *LINX* sur *L'énonciation et les marques d'oralité dans l'évolution du français* n°73 (Rodriguez Somolinos 2016) ou au volume collectif sur *Les Nouvelles voix d'accès au changement linguistique* (Ayes-Bennett et al. 2018, 2<sup>e</sup> partie, p. 181-370).

247 <https://www.frantext.fr>

Des discussions sont en cours au moment du dépôt de la présente synthèse. Cette base est gérée par mon laboratoire de rattachement, l'ATILF. Il est donc assez simple pour moi d'avoir accès à l'information et aux personnes qui administrent *Frantext*. Mais d'autres interfaces sont aussi envisageables.

Des bases comme le *Grand Corpus des littératures françaises et francophones du Moyen Âge au XX<sup>ème</sup> siècle<sup>248</sup> ou un projet comme *LiFraNum*<sup>249</sup> peuvent être d'une grande utilité pour recenser certains textes disponibles en version numérique.*

La littérature de jeunesse m'intéresse aussi tout particulièrement pour deux raisons principales : d'une part, elle intègre assez souvent des marques d'oralité et ouvre le champ des langues imaginaires et des principes de leur création, d'autre part, elle est un des lieux de manifestation du processus d'acquisition de la langue maternelle et elle peut être un point de repère pour l'analyse de ce processus. Là aussi, il y a nécessité d'un recensement des textes pertinents en fonction des contextes et des situations (contexte familial, contexte scolaire, lecture à haute voix par les adultes, lecture silencieuse par les enfants, etc.) et donc d'une collaboration avec des collègues spécialistes du domaine.

## 4.2.2 Humour et variation linguistique

Ce deuxième point concerne également la valorisation de ressources existantes et cible un ensemble de phénomènes en lien avec la variation linguistique et un genre de discours spécifique. Ces éléments ne sont pas sans lien avec le point précédent mais s'en distinguent tout de même en pratique.

Dans le cadre de cours que j'ai dispensés sur la variation linguistique<sup>250</sup>, j'ai récupéré des travaux d'étudiants qui consistaient à transcrire et à commenter des extraits oraux comportant des phénomènes variationnels marqués. Beaucoup de ces travaux portaient sur des caricatures, des imitations et des sketches d'humoristes et ce sont ces travaux que je souhaite exploiter. Chaque extrait transcrit dure environ 5 minutes. Un tri est nécessaire pour avoir des données « propres », c'est-à-dire un fichier son exploitable, une transcription correcte et des métadonnées suffisamment renseignées. Il paraît raisonnable de compter sur une centaine d'extraits. Une première étape est le lissage et la mise en forme des données pour créer une base exploitable et diffusable dans sa presque totalité, par exemple sur ORTOLANG<sup>251</sup>.

Ceci a pour objectif de permettre un travail sur la catégorisation et la stylisation de certaines particularités saillantes et efficaces du point de vue du genre comique. Le domaine de l'humour fait l'objet d'un intérêt croissant en raison de son importance sociale et des propriétés linguistiques et cognitives qu'il met en jeu, comme le montrent des travaux généraux (Attardo 2020) ou plus ciblés sur une variété, le français marseillais (Gasquet-Cyrus 2004) ou sur un genre, le monologue (Defays 2011).

Ce sont des données très riches du point de vue de la variation linguistique qui ajoutent une dimension diaphasique (variation individuelle) dans le cas des monologues et des *stand up*, et du point de vue de la représentation symbolique de la variation, car elles

248 Il s'agit d'une base de données relativement importante qui comporte des textes numérisés et interrogeables mais qu'il faut évaluer et trier pour pouvoir travailler finement.

<https://datec.lavoisier.fr/ressources-electroniques/arts/classiques-garnier-numerique/grand-corpus-des-litteratures-francaises-et-francophones-du-moyen-age-au-xxe-siecle/908>

249 Le projet LiFraNum est projet pluridisciplinaire qui a pour objectif de constituer un corpus de productions littéraires francophones nativement numériques. Lien : <https://marge.univ-lyon3.fr/projet-lifranum>

250 Voir document I, partie CV (p. 13). J'espère pouvoir transposer en partie ces cours dans la prochaine maquette.

251 Plate-forme d'Outils et de Ressources linguistiques pour un Traitement Optimisé de la LANGUE française. <https://www.ortolang.fr/fr/accueil>

manifestent verbalement des stéréotypes sociaux. De plus, en reprenant ces données récemment, j'ai confirmé ma première impression qu'elles constitueraient aussi un corpus très approprié pour l'étude des MD, en particulier ceux qui permettent de gérer l'interaction.

# 5

## Références





## 5.1 Références des corpus utilisés

Les exemples non référencés sont des exemples construits.

<b>Allier</b>	Corpus recueilli dans le cadre de la thèse de S. Giron. Entretiens semi-dirigés. Voir Giron (2004).
<b>Arg-microtexts</b>	Corpus de textes courts en allemand et leur traduction en anglais pour l'analyse de l'argumentation (réponses à des questions polémiques). <a href="https://github.com/peldszus/arg-microtexts/blob/master/topics_triggers.md">https://github.com/peldszus/arg-microtexts/blob/master/topics_triggers.md</a>
<b>CFPP (CFP)</b>	<i>Corpus de Français Parlé Parisien</i> . Entretiens sur Paris et sa région. Voir Branca-Rossoff et (2011). <a href="http://cfpp2000.univ-paris3.fr">http://cfpp2000.univ-paris3.fr</a>
<b>CLAPI</b>	<i>Corpus de Langue Parlée en Interaction</i> (extraits). Situations d'interaction variées. <a href="http://clapi.ish-lyon.cnrs.fr">http://clapi.ish-lyon.cnrs.fr</a>
<b>C-ORAL-ROM (COR)</b>	Parole spontanée plus ou moins formelle, partie française. Voir Cresti et al. (2004).
<b>CORPAIX</b>	CORpus d'AIX-en-Provence, équipe DÉLIC.
<b>CRFP</b>	<i>Corpus de Référence du Français Parlé</i> . Situations variées avec prise en compte du niveau d'étude des locuteurs. <a href="http://www.up.univ-mrs.fr/delic/corpus/index.html">http://www.up.univ-mrs.fr/delic/corpus/index.html</a>
<b>DECODA (DEC)</b>	Dialogues d'appels au service clientèle de la RATP. Voir Béchet et al. (2012). <a href="http://decoda.univ-avignon.fr">http://decoda.univ-avignon.fr</a>
<b>DÉCLICS</b>	Corpus du projet <i>Dispositif d'Études CLIniques sur les Corpus de Santé</i> . Consultations médicales, entretiens cliniques, présentations cliniques. Voir Blasco (2021, p. 30-37).
<b>ESLO</b>	Enquêtes Sociolinguistiques à Orléans (ESLO1) (interactions dans des situations diverses). <a href="http://eslo.huma-num.fr">http://eslo.huma-num.fr</a>
<b>ESTER</b>	Extraits d'émissions radiophoniques, distribué par ELRA. <a href="https://catalogue.elra.info/en-us/repository/browse/ELRA-S0241">https://catalogue.elra.info/en-us/repository/browse/ELRA-S0241</a>
<b>ETAPE</b>	Extraits d'émissions radiophoniques. Pour un descriptif, voir Nouvel (2012, p. 110-116).
<b>FLEURON (FLE)</b>	FLEURON (extraits) : interactions étudiant/personnel administratif. <a href="https://apps.atilf.fr/fleuron2">https://apps.atilf.fr/fleuron2</a>
<b>Fra80</b>	Corpus de <i>Français des Années 80</i> . Entretiens semi-dirigés. Voir <i>Cahiers du Français des Années Quatre-Vingts</i> , hors-série 1 (1989), CREDIF, ENS de Saint-Cloud.
<b>Frantext</b>	Base de données de textes français du IX <sup>ème</sup> siècle au XXI <sup>ème</sup> siècle. <a href="https://www.frantext.fr">https://www.frantext.fr</a>
<b>frcapop</b>	Corpus <i>FRançais CANadien POPulaire</i> . Sélection d'entretiens des corpus Sankoff-Cedergren et Montréal 74 sur la base des profils des locuteurs. Voir Dagnat (2006 [6]).
<b>FRE</b>	<i>French Oral Narrative</i> . Contes lus. <a href="http://frenchoralnarrative.qub.ac.uk">http://frenchoralnarrative.qub.ac.uk</a>
<b>HUS</b>	Corpus constitué par M. Husianycia pour sa thèse. Enregistrements de réunions de travail. Voir Husianycia (2011).
<b>Le Monde</b>	Textes du quotidien français <i>Le Monde</i> , années 1987-1998.
<b>Montréal 84</b>	Corpus de français parlé à Montréal, qui fait suite au corpus Sankoff-Cedergren. Entretiens semi-dirigés. Voir Thibault & Vincent (1990).
<b>OFROM</b>	Corpus <i>Oral de FRançais de Suisse ROMande</i> . <a href="http://www11.unine.ch">http://www11.unine.ch</a>

<b>Sankoff-Cedergren</b>	Corpus de français parlé à Montréal. Entretiens semi-dirigés. Voir Sankoff et al. (1976) et Sankoff & Laberge (1978).
<b>Sci-DTB</b>	<i>Discourse Dependency TreeBank for Scientific Abstracts</i> . Résumés d'articles scientifiques en anglais. Voir Yang & Li (2018). <a href="https://github.com/PKU-TANGENT/SciDTB">https://github.com/PKU-TANGENT/SciDTB</a>
<b>TCOF</b>	<i>Traitement de Corpus Oraux en Français</i> . Interactions enfants-adultes et adultes-adultes. <a href="http://www.cnrtl.fr/corpus/tcof">http://www.cnrtl.fr/corpus/tcof</a>
<b>tremblay</b>	Corpus constitué de cinq pièces de l'écrivain Michel Tremblay, numérisé et annoté dans le cadre de Dagnat (2006 [6]).
<b>TUFS</b>	Corpus réalisé à la Tokyo University of Foreign Studies (partie française).
<b>VALIBEL</b>	Corpus de français belge, géré par le Centre VALIBEL-Discours et Variation. <a href="https://uclouvain.be/fr/instituts-recherche/ilc/valibel/corpora.html">https://uclouvain.be/fr/instituts-recherche/ilc/valibel/corpora.html</a>

## 5.2 Références bibliographiques personnelles

Les numéros entre crochets sont identiques à ceux utilisés dans le document 1 (CV) et dans le document 3 (Recueil de mes travaux).

Les travaux jugés représentatifs sont indiqués par un astérisque.

- [1] 2000 **Dagnat M.**  
« Le 'coup de stylet' de Michel Tremblay : la langue populaire dans *Les Belles-sœurs* ». *Études canadiennes/Canadian Studies* 49, p. 67-83.
- [2] 2002a **Dagnat M.**  
« La tentation du corpus. Du théâtre au roman, de la parole à l'écriture chez Michel Tremblay ». *Canadian Studies in Europe/Etudes Canadiennes en Europe* 2, Mazaryk University Press, p. 51-61.
- [3] 2002b **Dagnat M.**  
*Michel Tremblay. Le « joual » dans Les Belles-Sœurs*. Paris : L'Harmattan.
- [4] 2003 **Dagnat M.**  
« Michel Tremblay sur le fil de l'art ». In Piccione M.-L. (dir.), *Voies vers l'extrême, voix de l'extrême*. Bordeaux : MSH d'Aquitaine, p. 139-150.
- [5] 2005 **Dagnat M.**  
« 'Feuilleté énonciatif' et hybridité générique ». In Hazaël-Massieux M.-C., Bertrand M. (éd.), *Langues et identités narratives dans les littératures de l'ailleurs (Antilles, Réunion, Québec)*. Aix-en-Provence : Presses de l'Université de Provence, p. 35-89.
- [6] 2006 **Dagnat M.**  
*L'oral comme fiction*, thèse de doctorat de Sciences du langage et Ph.D. d'Études françaises (cotutelle), Aix-en-Provence et Montréal, Université de Provence (Aix-Marseille I), Laboratoire Parole & Langage et Université de Montréal, Département d'Études françaises, 2006, 2 volumes (616 p. + 170 p. d'annexes).
- [7]\* 2007 **Dagnat M.**  
« L'oral au pied de la lettre : raisons et déraisons graphiques ». *Études françaises* 43/1, p. 83-100.
- [8]\* 2008a **Dagnat M.**  
« La catégorisation de la variation linguistique ». *Grenzgänge. Beiträge zu einer modernen Romanistik* 15/30, p. 11-41.

- [9]\* 2008b **Dagnat M.**  
« Constructionnalité des parataxes conditionnelles ». In Durand J., Habert B., Laks B. (éd.), *Actes du 1er Congrès Mondial de Linguistique Française (CMLF 2008)*. Paris : Institut de Linguistique Française, EDP Sciences, p. 2467-2482. Session Syntaxe.
- [10]\* 2008c **Dagnat M.**  
« Profils linguistiques et structure textuelle ». In Heiden S., Pincemin B. (éd.), *Actes des 9<sup>e</sup> Journées d'Analyse Statistique des Données Textuelles (JADT)* vol. 1. Lyon : Presses Universitaires de Lyon, p. 369-379.
- [11] 2008d Jayez J. & **Dagnat M.**  
« The interpretation of continuative cues in SDRT ». In Benz A., Kühnlein P., Stede M. (eds), *Proceedings of the Workshop Constraints in Discourse (CID 3)*. University of Potsdam (July 30th-August 1st 2008), p. 53-60.
- [12]\* 2009 Jayez J. & **Dagnat M.**  
« One More Step and You'll Get Pseudo-Imperatives Right ». In Riester A., Solstad T. (eds), *Proceedings of Sinn und Bedeutung 13*, University of Stuttgart, p. 247-260.
- [13]\* 2010a **Dagnat M.**  
« Support linguistique de la métalepse ». In Ablali D., Kastberg Sjöblom M. (éd.), *Linguistique et littérature : Cluny, 40 ans après* vol. 2, Besançon. Presses Universitaire de Franche Comté (édition électronique), 16 p.
- [14]\* 2010b **Dagnat M.** & Jayez J.  
« La cohésion parataxique : une approche constructionnelle ». In Béguelin M.-J., Avanzi M., Corminboeuf G. (éd.), *La Parataxe, Structures, marquages et explications discursives*. Berne : Peter Lang, vol. 2, p. 61-93.
- [15] 2011 Jayez J. & **Dagnat M.**  
« Discourse 'major continuatives' in a non-monotonic framework ». In Yoo H.-Y., Delais-Roussarie E. (dir.), *Actes Interface Discours Prosodie / Proceedings Discourse Interface Prodody (IDP 09)*, p. 223-238 (édition électronique).
- [16]\* 2012a **Dagnat M.**, Colotte V., Bartkova K. & Bonneau A.  
« Continuations intra et interphrastiques du français : premiers résultats expérimentaux ». In *Actes du 3e Congrès Mondial de Linguistique Française (CMLF 2012)*, Lyon, Institut de Linguistique Française, EDP Sciences, p. 1471-1485. Session Phonétique, phonologie et interfaces.
- [17] 2012b Bartkova K., Bonneau A., Colotte V. & **Dagnat M.**  
« Productions of 'Continuation contours' by French speakers in L1 (French) and L2 (English) ». In *Proceedings of the 6<sup>th</sup> International Conference on Speech Prosody (Shangai)*, Tongji University Press, vol. 1, p. 426-429.
- [18] 2012c Jayez J. & **Dagnat M.**  
« The Semantics of French Continuative Rises in SDRT ». In Benz A., Stede M., Kühnlein P. (eds), *Constraints in Discourse 3, Representing and inferring discourse structure*. Amsterdam/Philadelphia : John Benjamins, p. 109-135.
- [19]\* 2012d **Dagnat M.**  
« Textual Stratification and functions of orality in theatre ». In Brumme J., Espunya A. (eds.), *The Translation of Fictive Dialogue*. Amsterdam : Rodopi, p. 81-99.
- [20]\* 2013 Bartkova K. & **Dagnat M.**  
« Cross Linguistic Study of French and English Prosody ». *The Phonetician* 107-108, 2013/I-II, International Society of Phonetic Sciences, p. 35-52.
- [21] 2014a Colinet M., Danlos L., **Dagnat M.** & Winterstein G.  
« Emplois de la préposition *pour* suivie d'une infinitive : description, critères formels et annotation en corpus ». In *Actes du 4e Congrès Mondial de Linguistique*

*Française*, Berlin, Institut de Linguistique Française, EDP Sciences, p. 3041-3058. Session Sémantique.

- [22] 2014b Bartkova K. & **Dargnat M.**  
« Automatic Extraction of the general tendency of F0 patterns. Cross Linguistic Study on laboratory data ». In Campbell N., Gibbon D., Hirst D. (eds), *Social and Linguistic Speech Prosody, Proceedings of the 7<sup>th</sup> International Conference on Speech Prosody* (Trinity College, Dublin), Science Foundation Ireland, p. 848-852.
- [23]\* 2015a **Dargnat M.**  
« La compositionnalité en question(s) ». *Verbum*, vol. XXXVII/2, p. 197-206.
- [24]\* 2015b **Dargnat M.**, Bartkova K. & Jouvét D.  
« Discourse Particles In French: Prosodic Parameters Extraction and Analysis ». In Dediu A.-H., Martin-Vide C., Vicsi K. (eds), *Statistical Language and Speech Processing* Cham : Springer, p. 39-49.
- [25]\* 2016a Bastien A., Bartkova K. & **Dargnat M.**  
« How to be a Discourse Particle ? ». In Barnes J., Brugos A., Shattuck-Hufnagel S., Veilleux N. (eds), *Proceedings of Speech Prosody 2016*, Boston : International Speech Communication Association, p. 858-863.
- [26]\* 2016b Combettes B. & **Dargnat M.**  
« Grammaire de constructions et diachronie : intensivité et concession ». In *Actes du 5e Congrès Mondial de Linguistique Française (CMLF 2016)*, Tours, SHS Web of Conference 27, IDP Sciences, 17 p. Session Histoire du français : perspectives diachronique et synchronique.
- [27] 2017a Bartkova K., **Dargnat M.**, Jouvét D. & Lee L.  
« Annotations de particules de discours en français sur une large variété de corpus ». In L. Danlos L., Fort K., Guillaume B., Kahane S. (éd.), *Actes de l'atelier « ACor4French – Les corpus annotés du français »*, TALN 2017, Orléans, p. 10-17.
- [28]\* 2017b Jouvét D., Bartkova K., **Dargnat M.** & Lee L.  
« Analysis and Automatic Classification of Some Discourse Particles on a Large Set of French Spoken Corpora ». In Camelin N., Estève Y., Martin-Vide C. (eds), *Statistical Language and Speech Processing* Cham : Springer, p. 32-43.
- [29] 2018 Lee L., Bartkova K., **Dargnat M.** & Jouvét D.  
« Prosodic and Pragmatic Values of Discourse Particles in French ». In Botinis A. (ed.), *Proceedings of ExLing 2018 - 9th Tutorial and Research Workshop on Experimental Linguistics*, Paris 7 Denis Diderot, p. 77-80.
- [30] 2019a Huber L., Toussaint Y., Roze C., **Dargnat M.** & Braud C.  
« Alignement de structures discursives et argumentatives par fouille de graphes et de redescriptions ». In Couceiro M., Napoli A. (éd.), *Actes des XXVF rencontres de la Société Francophone de Classification (SFC)*, Nancy, p. 83-88.
- [31] 2019b Lee L., Bartkova K., Jouvét D., **Dargnat M.** & Keromnes Y.  
« Can prosody meet pragmatics ? Case of discourse particles in French ». In *Proceedings of the International Congress of Phonetic Sciences (ICPhS2019)*, Melbourne, paper 575, 5 p.
- [32]\* 2019c Huber L., Toussaint Y., Roze C., **Dargnat M.** & Braud C.  
« Aligning Discourse and Argumentation Structures using Subtrees and Redescription Mining ». In Stein B., Wachsmuth H. (eds), *Proceedings of the 6th Workshop on Argument Mining*, ACL 2019, Florence, Association for Computational Linguistics, p. 35-40.
- [33]\* 2020a Combettes B. & **Dargnat M.**  
« Le souvenir du contexte : rémanence et grammaticalisation des marqueurs discursifs ». *Le Français moderne*, 88/2, p. 282-294.

- [34] 2020b Lee L., Jouvét D., Bartkova K., Keromnes Y. & **Dargnat M.**  
« Étude comparative de corrélats prosodiques de marqueurs discursifs français et anglais selon leur fonction pragmatique ». In *Journées d'Etudes sur la Parole* (JEP 2020), Nancy, p. 335-343.
- [35] 2020c Lee L., Jouvét D., Bartkova K., Keromnes Y. & **Dargnat M.**  
« Correlation between prosody and pragmatics: case study of discourse markers in French and English ». In *Proceedings of Interspeech 2020* (Shanghai), 5 p.
- [36] 2020d Huber L., Memmadi C., **Dargnat M.** & Toussaint Y.  
« Do sentence embeddings capture discourse properties of sentences from scientific abstract ? » In *1<sup>st</sup> Workshop on Computational Approaches to Discourse (CODI), Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, 10 p.
- [37] 2020e **Dargnat M.** & Jayez J.  
« Projection and main content ». In Abeillé A., Bonami O. (eds), *Constraint-based Syntax and Semantics. Papers in Honor of Danièle Godard*. Stanford : CSLI Publications, p. 101-126.
- [38]\* 2020f **Dargnat M.**  
« Subjectivité et projection : le cas des particules discursives ». In *Actes du 7<sup>e</sup> Congrès Mondial de Linguistique Française*, Montpellier, SHS Web of Conference 78, IDP Sciences, 14 p. Session Sémantique.
- [39]\* 2021a **Dargnat M.**  
Section XIV-2-4 « Les juxtapositions à interprétation conditionnelle » et section XVIII-4-3 « Interjections et particules de discours ». In Abeillé A., Godard D., Delaveau A. (dir.), *Grande Grammaire du Français*. Arles/Paris : Actes Sud/Imprimerie Nationale, p. 1569-1572 et p. 2015-2025.
- [40]\* 2021b **Dargnat M.** & Jayez J.  
« Mais et associés ». *Lexique* 29, p. 211-228.
- [41]\* 2021c Combettes B. & **Dargnat M.**  
« Grammaticalisation du tour corrélatif *d'autant que* comme mot du discours ». In A. Roig A., Schnedecker C. (éd.), *La connexion corrélatrice, De la phrase au discours*. Paris : Classiques Garnier, p. 47-69.
- [42]\* 2022a **Dargnat M.**  
« *Mais enfin* : construction et association ». *Langages* 225, p. 49-63.
- [43]\* 2022b **Dargnat M.**  
« Explication et argumentation ». In Blasco M. (éd.), *Parler à l'hôpital : écouter ce qui est dit, décrypter ce qui se dit*. Münster : NODUS Publikationen, p. 88-204.

## 5.3 Références générales

### A

- Abécassis M. 2003. « Le français populaire : a valid concept ? » *Marges linguistiques* 6, p. 116-132.
- Abeillé A & Borsley R. D. 2006. « La syntaxe des comparatives corrélatives en anglais et en français ». *Faits de langue* 28/1, p. 21-33.
- Abeillé A. & Godard D. (dir.) 2021. *Grande Grammaire du Français*. 2 tomes. Arles : Actes Sud.
- Abraham W. (ed.) 1991. *Discourse Particles*. Amsterdam-Philadelphia : John Benjamins.
- Accuosto P. 2021. *Mining arguments in scientific abstracts. Application to argumentative quality assessment*. Ph.D., Universitat Pompeu Fabra, Barcelona.
- Adam J.-M. 1997. *Le style dans la langue : une reconception de la stylistique*. Lausanne : Delachaux et Niestlé.
- Adam J.-M. 2019. « La notion de texte ». *Encyclopédie Grammaticale du Français*. 75 p. [En ligne](#)
- Adam J.-M. 2000. *Textes et genres de discours*. Paris : Nathan.
- Aijmer K. 2002. *English Discourse Particles. Evidence from a corpus*. Amsterdam-Philadelphia : John Benjamins.
- Aijmer K. 2013. *Understanding Pragmatic Markers*. Edinburgh : Edinburgh University Press.
- Aijmer K. & Simon-Vandenberghe A.-M. 2011. « Pragmatic markers ». In Zienkowski J., Östman J.-O., Verschueren J. (eds), *Discursive Pragmatics*, Amsterdam-Philadelphia : John Benjamins, p. 223-247.
- Ameka F. K. 1992. « Interjections : The universal yet neglected part of speech ». *Journal of Pragmatics* 18, p. 101-118.
- Ameka F. K. 2006. « Interjections ». In Brown K. (ed.), *Encyclopedia of Language and Linguistics*. London : Elsevier, p. 743-746.
- Amossy R. 2000. *L'argumentation dans le discours. Discours politique, littérature d'idées, fiction*. Paris : Nathan.
- Amossy R. 2005. « The argumentative dimension of discourse ». In van Eemeren F.H., Houtlosser P. (eds.), *Practices of Argumentation*. Amsterdam-Philadelphia : John Benjamins, p. 87-98.
- Amossy R. 2009. « Argumentation in Discourse : A socio-discursive Approach to Arguments ». *Informal Logic* 29/3, p. 252-267.
- Andersen G. & Fretheim T. (eds). 2000. *Pragmatic Markers and Propositional Attitude*. Amstersam-Philadelphia : John Benjamins.
- Anderson J. R. 1976. *Language, memory, and thought*. Hillsdale : Erlbaum.
- Anis. J. 1981. « Écrit/oral : discordances, autonomes, transpositions ». *Études de linguistique appliquée* 42, p. 7-22.
- Anis J. 1999. « Chats et usages graphiques ». In Anis J. (éd.), *Internet, communication et langue française*. Paris : Hermès Sciences Publications, p. 71-90.
- Anscombre J.-C. & Ducrot O. 1997 [1983]. *L'argumentation dans la langue* 3<sup>e</sup> éd. Sprimont : Mardaga.
- Asher N. 1993. *Reference to Abstract Objects in Discourse*. Cham : Springer.
- Asher N. & Lascarides A. 2003. *Logics of conversation*. Cambridge : Cambridge University Press.
- Asher N. & Vieu L. 2005. « Subordinating and coordinating discourse relations ». *Lingua* 115, 591-610.
- Attardo S. 2020. *The Linguistics of Humor. An Introduction*. Oxford : Oxford University Press.
- Auchlin A. 1981. « Réflexions sur les marqueurs de structuration de la conversation ». *Études de Linguistique Appliquée* 44, p. 88-103.
- Ayres-Bennett W., Carlier A., Glikman J., Rainsford T. M., Siouffi G., Skupien Dekens C. (dir.) 2018. *Nouvelles voies d'accès au changement linguistique*. Paris : Classiques Garnier.
- Azar M. 1999. « Argumentative Text as Rhetorical Structure : An Application of Rhetorical Structure Theory ». *Argumentation* 13, p. 97-114.

### B

- Bakhtine M. 1984 [1979]. *Esthétique de la création verbale*. Paris : Gallimard. Textes originaux datant de 1952-1953.
- Bakhtine M. (Volochinov) 1977 [1929]. *Le Marxisme et la philosophie du langage. Essai d'application de la méthode sociologique en linguistique*. Paris : Les éditions de minuit.
- Barthes R. 1987 [1982]. « L'effet de réel ». In Genette G., Todorov T. (dir.), *Littérature et réalité*. Paris : Editions du Seuil, p. 81-90.

- Bartkova K., Delais-Roussarie E., Santiago F. 2012. « *Prosotran* : a tool to annotate prosodically non-standard data ». *Proceedings of Speech Prosody 2012*, Shanghai, p. 55-58.
- Bartkova K. & Jouvét D. 2013. « Automatic detection of the prosodic structures of Speech Utterances ». *Proceedings of the 15th International Conference on Speech and Computer (SPECOM)*. Berlin : Springer, p. 1-8.
- Bastien A. 2015. *Prosodie et fonctions pragmatiques des marqueurs discursifs à base verbale : 'écoute', 'écoutez', 'tu comprends', 'vous comprenez', 'tu sais', 'vous savez', 'tu vois', 'vous voyez'*. Master 2 de l'Université de Lorraine, 65 p.
- Bazzanella C. 1990. « Phatic Connectives as Interactional Cues in Spoken Italian ». *Journal of Pragmatics* 14, p. 629-647.
- Beaver D., Roberts C., Simons M. & Tonhauser J. 2017. « Questions under discussion : Where information structure meets projective content ». *Annual Review of Linguistics* 3, p. 265-284.
- Bechet F., Maza B., Bigouroux N., Bazillon T., El-Beze M., De Mori R., Arbillot E.. 2012. « DECODA: a call-centre human-human spoken conversation corpus. » *Proceedings of LREC 2012*, p. 1343-1347.
- Beeching K. 2002. *Gender, politeness and pragmatic particles in French*. Amsterdam : John Benjamins.
- Beeching K. 2016. *Pragmatic markers in British English. Meaning in social interaction*. Cambridge : Cambridge University Press.
- Béguelin M.-J., Avanzi M., Corminboeuf G. (éds) 2010. *La Parataxe*, 2 tomes. Berne : Peter Lang.
- Benveniste E. 1966. *Problèmes de linguistique générale*, 2 tomes. Paris : Gallimard.
- Benz A. & Jasinskaja K. 2017. « Questions Under Discussion: From Sentence to Discourse ». *Discourse processes* 54(3), p. 177-186.
- Berrendonner A. 1983. « Connecteurs pragmatiques et anaphores ». *Cahiers de linguistique française* 7, p. 215-246.
- Bertrand R. & Chanet C. 2005. « Fonctions pragmatiques et prosodie de *enfin* en français spontané ». *Revue de Sémantique et Pragmatique* 17, p. 41-68.
- Beyssade C. 2015. « French ». In Jungbluth K., Da Milano F. (eds), *Manual of Deixis in Romance Languages*. Berlin : De Gruyter, p. 167-188.
- Beyssade C. Delais-Roussarie E., Doetjes J., Marandin J.-M. & Rialland A. 2004. « Prosody and Information in French ». In Corblin F., de Swart H. (eds), *Handbook of French Semantics*. Stanford : CSLI Publications, p. 477-499.
- Beyssade C. & Marandin J.-M. 2007. « French intonation and attitude attribution ». In Pas-cal D., McReady E., Palmer A. & Reese B. (eds). *Proceedings of the 2004 Texas Linguistics Society Conference : Issues at the Semantics-Pragmatics Interface*. [En ligne](#)
- Bezuidenhout A. 2003. « Procedural Meaning and the Semantics/Pragmatics Interface ». In Bianchi C. (ed.) 2003), *The Semantic/pragmatics distinction*, Stanford : CSLI publications, p. 101-131.
- Bhatia V. 2013. *Analysing Genre. Language Use in Professional Settings*. London-New York : Routledge.
- Bhatia S., Chaudary P. & Dey N. 2016. *Opinion Mining in Information Retrieval*. Singapore : Springer (ebook).
- Biber D. & Reppen R. (eds). 2015. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge : Cambridge University Press.
- Biber D. & Conrad S. 2019. *Register, genre, and style*. Cambridge : Cambridge University Press.
- Biet C. & Triau C. 2006. *Qu'est-ce que le théâtre ?* Paris : Gallimard.
- Blakemore D. 1987. *Semantic constraints on relevance*. Oxford: Blackwell.
- Blakemore D. 1992. *Understanding utterances. Introduction to Pragmatics*. Oxford and Cambridge : Blackwell.
- Blakemore D. 2002. *Relevance and Linguistic Meaning. The Semantics and Pragmatics of Discourse Markers*. Cambridge : Cambridge University Press.
- Blanche-Benveniste C. 2000. *Approches de la langue parlée en français*. Paris : Ophrys.
- Blanche-Benveniste C. & Jeanjean C. 1987. *Le français parlé. Édition et transcription*. Paris : Didier érudition.
- Blanche-Benveniste C. et al. 1990. *Le français parlé, études grammaticales*. Paris : CNRS éditions.
- Blasco M. (éd.) 2021. *Parler à l'hôpital. Écouter ce qui est dit, décrypter ce qui se dit*. Münster : Nodus Publikationen.
- Blondeau H., Tremblay M., Bertrand A. & Michel E. 2021. « A new milestone for the study of variation in Montréal French: The Hochelaga-Maisonneuve sociolinguistic survey ». *Corpus* 22, 16 p. [En ligne](#)
- Bolinger D. 1977. *Meaning and Form*. London-New York : Longman.
- Bonami O. 1999. *Les constructions du verbe. Le cas des groupes prépositionnels argumentaux*. Thèse de l'Université Paris 7.
- Booth. W. C. 1983. *The Rhetoric of Fiction*. Chicago : University of Chicago Press.
- Bordas E. 2008. « *Style* », *un mot et des discours*. Paris : Kimé.
- Boulanger L. 2001. *Pièces à conviction. Entretiens avec Michel Tremblay*. Montréal : Leméac.



- Bourdieu P. 1983. « Vous avez dit populaire ? ». *Actes de la Recherche en Sciences Sociales* 46, p. 98-105.
- Branca-Rosoff S., Fleury S., Lefeuvre F. & Pires M. 2011. « Constitution et exploitation d'un corpus de français parlé parisien ». *Corpus* 10, p. 81-98.
- Braud C. 2015. *Identification automatique des relations de discours implicites à partir de corpus annotés et de données brutes*. Thèse de l'Université Paris 7. [En ligne](#)
- Breck E., Cardie C. 2022. « Opinion Mining and Sentiment Analysis ». In Mitkov R. (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford : Oxford University Press, chap. 43, p. 1031-1053.
- Brémond C. 2002. *Les « petites marques du discours »*. Le cas du marqueur méta-discursif bon en français. Thèse de l'Université de Provence.
- Bres J. 2005. « Savoir de quoi on parle : dialogue, dialogal, dialogique ; dialogisme, polyphonie... » In Bres J., Haillet P. P., Mellet S., Nølke H. & Rosier L. (dir.), *Dialogisme et polyphonie. Approches linguistiques*. Bruxelles : De Boeck Duculot, p. 47-61.
- Brezina V. 2018. *Statistics in Corpus Linguistics : A Practical Guide*. Cambridge : Cambridge University Press.
- Brinton L. J. 1996. *Pragmatic Markers in English. Grammaticalization and Discourse Functions*. Berlin-New York : Mouton de Gruyter.
- Brinton L. J. 2010. « Discourse markers ». In Jucker A. H., Taavitsainen I. (eds.), *Historical pragmatics*. Berlin : De Gruyter Mouton, p. 285-314.
- Brinton L. J. 2017. *The Evolution of Pragmatic Markers in English. Pathways of change*. Cambridge : Cambridge University Press.
- Brinton L. J. & Traugott E. C. 2005. *Lexicalization and language change*. Cambridge : Cambridge University Press.
- Brouwer H., Fitz H. & Hoeks J. 2012. « Getting real about semantic illusions : Rethinking the functional rôle of the P600 in langage comprehension. *Brain Research* 1446, p. 127-143.
- Bruxelles S., Dos Reis Nunes G., Gouaze J., Fouquier E., Ducrot O., Remis A. 1976. « Mais occupe-toi d'Amélie ». *Actes de la recherche en sciences sociales* 2/6, p. 47-62.
- Buchi É. 2000. « Approche diachronique du marqueur métadiscursif français *quoi* (« La pragmatization d'un réévaluatif, *quoi* »). In Englebort A. et al. (éd.), *Actes du XXII<sup>e</sup> Congrès International de Linguistique et de Philologie Romanes* (Bruxelles, 23-29 juillet 1998), Tübingen : Niemeyer, p. 81-91.
- Buchi É. & Städler T. 2010. « La pragmatization de l'adverbe *enfin* du point de vue des romanistes ». In *Actes du CMLF 2008*, p. 159-171.
- Budzynska K. & Villata S. 2016. « Argument Mining ». *IEEE Intelligent Informatics Bulletin* 17/1, 6 p.
- Bühler K. 2011 [1934]. *Theory of Language. The representational function of langage*. Amsterdam-Philadelphia : John Benjamins.
- Bybee J. L. 2010. *Language, Usage, and Cognition*. Cambridge : Cambridge University Press.

## C

- Cabrio H. & Villata S. 2018. « Five Years or Argument Mining : a Data-driven Analysis ». *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI-18)*, p. 5427-5433.
- Cadiot A., Ducrot O., Fradin B. & Nguyen T. B. 1985. « Enfin, marqueur métalinguistique », *Journal of Pragmatics* 9, p. 199-239.
- Calvet 1993. *La sociolinguistique*, Paris, Presses Universitaires de France.
- Cappelen H. & Lepore E. 2002. « Indexicality, Binding, Anaphora and Priority Truth ». *Analysis* 62.4, p. 271-281.
- Carston R. 2016. « The heterogeneity of procedural meaning ». *Lingua* 175/176, p. 154-166.
- Chanet C. 2003. « Fréquence des marqueurs discursifs en français parlé : quelques problèmes de méthodologie ». *Recherches sur le Français Parlé* 18, 25 p. [En ligne](#)
- Charaudeau D. & Maingueneau D. (dir.) 2002. *Dictionnaire d'analyse du discours*. Paris : Seuil.
- Charolles M. 2020. « Discourse topics and digressive markers ». *Journal of Pragmatics* 161, p. 57-77.
- Chenetier-Alev 2004. *L'oralité dans le théâtre contemporain : Herbert Achternbusch, Pierre Guyotat, Valère Novarina, Jon Fosse, Daniel Danis, Sarah Kane*, thèse de l'université Paris 3 . Lille : ANRT.
- Chervel André. 1979. « Rhétorique et grammair : petite histoire du circonstanciel ». *Langue française* 41, p. 5-19.
- Chierchia G. & MacConnell-Ginet S. 1990. *Meaning and Grammar : An Introduction to Semantics*. Cambridge : MIT Press.
- Choi-Jonin I. & Delais-Roussarie E. 2006. « L'association de propositions sans marque segmentale en français parlé : étude syntactico-sémantique et prosodique ». *Faits de langue* 28, p. 83-94.
- Clark B. 1993. « Relevance and pseudo-imperatives ». *Linguistics and Philosophy* 16/1, p. 79-121.
- Clark B. 2013. *Relevance Theory*. Cambridge : Cambridge University Press.
- Cogard K. 2001. *Introduction à la stylistique*. Paris : Flammarion.

- Colombat B. & Lahaussais A. 2019. *Histoire des parties du discours*. Louvain : Peeters.
- Combettes B. 1983. *Pour une grammaire textuelle. La progression thématique*. Paris/Gembloux et Bruxelles : Duculot et De Boeck.
- Combettes B. 1994. « Une approche diachronique des connecteurs et modalisateurs ». *Pratiques* 84, p. 55-67.
- Combettes B. 2002. « Analyse linguistique des textes et stylistique ». *Langue française* 135, p. 95-113.
- Combettes B. 2007. « Discontinuité et cohérence discursive : le cas des ajouts après le point ». *Cahiers de praxématique* 48, p. 111-134.
- Constant M., Gülşen Eryiğit G., Johanna Monti J., van der Plas L., Ramisch C., Rosner M., Todirascu A. 2017. « Multiword expression processing : A survey ». *Computational Linguistics* 43/4, p. 837-893.
- Corblin F. 1999. « Une approche cognitive des conditionnelles ». *Linx* 41, 15 p. [En ligne](#)
- Corblin F. 2002. *Représentation du discours et sémantique formelle : Introduction et applications au français*. Paris : Presses Universitaires de France.
- Corblin F. & de Swart H. (eds) 2004. *Handbook of French Semantics*. Stanford: CSLI Publications.
- Corminboeuf G. 2007. « Les constructions hypothétiques non marquées ». *Encyclopédie Grammaticale du Français*, 18 p. [En ligne](#)
- Corminboeuf G. 2008. « 'Tu m'embrasses encore, et c'est mon pied dans les pompons ! Comment construit-on le sens ? ». *Discours* 3 (en ligne), 26 p.
- Corminboeuf G. 2009. *L'expression de l'hypothèse en français. Entre hypotaxe et parataxe*. Bruxelles : DeBoeck/Duculot.
- Cornish F. 1999. *Anaphora, Discourse and Understanding Evidence from English and French*. Oxford: Oxford University Press.
- Cresti E., Do Nascimento F. B., Moreno-Sandoval A., Veronis J., Martin P. & Choukri, K. « The C-ORAL-ROM CORPUS. A multilingual resource of spontaneous speech for romance languages ». Proceedings of LREC 2004 (Lisbon), 4 p.
- Crible L. & Degand, L. 2019. « Domains and functions: A two-dimensional account of discourse markers ». *Discours : revue de linguistique, psycholinguistique et informatique* 24, 35 p. [En ligne](#)
- Cruse A. 1986. *Lexical Semantics*. Cambridge : Cambridge University Press.
- Cruse A. 2000. *Meaning in Language. An introduction to Semantics and Pragmatics*. Oxford : Oxford University Press.
- Culicover P. 1972. « OM sentences ». *Foundations of Language* 8, p. 199-236.
- Culicover P. & Jackendoff R. 1997. « Semantic subordination despite syntactic coordination ». *Linguistic Inquiry* 28, p. 195-217.
- ## D
- Daudaravičius V. & Marcinkevičienė R. 2004. « Gravity Counts for the Boundaries of Collocations ». *International Journal of Corpus Linguistics* 9/2, p. 321-348.
- Dancygier B. 1998. *Conditionals and Prediction. Time, Knowledge and Causation in Conditional Constructions*. Cambridge : Cambridge University Press.
- Daneš F. 1974. « Functional sentence perspective and the organization of the text ». In Daneš F. (ed). *Papers on Functional Sentence Perspective*. Prague : Academia /The Hague: Mouton, p. 106-128.
- Danlos L., Colinet M., Steinlin J. 2015. « FDTB1, première étape du projet 'French Discourse Treebank' : repérage des connecteurs de discours en corpus ». *Discours* 17, 22 p. [En ligne](#)
- Davoine J.-P. 1980. « Les connecteurs phatiques, tu penses !, penses-tu !, remarque, écoute ». In Kerbrat-Orecchioni C. (éd.), *Le discours polémique*, Lyon : Presses Universitaires de Lyon, p. 83-107.
- Debausieux J.-M. 2004. « Les conjonctions de subordination : mots grammaticaux ou mots de discours ? Le cas de *parce que* ». *Revue de Sémantique et Pragmatique*, p. 51-67.
- Defays J.-M. 2011. « Prolégomènes à l'étude du monologue ». *Humoresques* 33, p. 23-38.
- Degand L. & Pander Maat H. 2003. « A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale ». In Verhagen A., van de Weijer J. (eds.), *Usage-based approaches to Dutch*. Utrecht : LOT, p. 175-199.
- Degand L., Cornillie B. & Pietrandrea P. (eds) 2013. *Discourse Markers and Modal Particles. Categorization and description*. Amsterdam-Philadelphia : John Benjamins.
- Degand L. & Evers-Vermeul J. 2015. « Grammaticalization or pragmaticalization of discourse markers ? More than a terminological issue ». *Journal of Historical Pragmatics* 16, p. 59-85.
- Degand L. & Pander Maat H. 2001. « Scaling causal relations and connectives in terms of speaker involvement ». *Cognitive Linguistics* 12/3, p. 211-245.
- Degand L. & Zufferey S. 2016. « French Connectives ». In Chapelle C. A. (ed.). *The Encyclopedia of Applied Linguistics*. John Wiley & Sons, 6 p. Online publication.
- Delattre P. 1966. « Les dix intonations de base du français ». *The French Review* 40/1, p. 1-14.
- Delort L. 2004. « Relations subordonnantes et coordonnantes pour la désambiguïsation du

- discours ». Workshop SDRT, TALN 2004, Fès, 10 p.
- Delvin J., Chang M.-W., Lee K., Toutanova K.. 2019. « BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding ». In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*. Minneapolis : Association for Computational Linguistics, p. 4171–4186.
- De Mulder W. 2020. « Lexicalisation – grammaticalisation – pragmaticalisation ». In Marchello-Nizia C. et al. (éd.), *Grande Grammaire Historique du Français*. Berlin-Boston : De Gruyter Mouton, section 45.5, p. 1820-1826.
- Denturck E. 2008. *Étude des marqueurs discursifs, l'exemple de quoi*. Master in de Taal-en Letterkunde. Universiteit Gent.
- Dér C. I. 2010. « On the status of discourse markers ». *Acta Linguistica Hungarica*, 57 (1), p. 3-28.
- Desagulier G. 2015. « Le statut de la fréquence dans les grammaires de constructions : simple comme bonjour ». *Langages* 197, p. 99-128.
- Desagulier G. 2017. *Corpus Linguistics and Statistics with R. Introduction to Quantitative Methods in Linguistics*. Cham : Springer.
- Desmets M., Hamon S. & Lavieu B. 2003. « Les grammaires HPSG ». *Linx* 48, p. 57-76.
- Detges U. & Waltereit R. 2016. « Grammaticalization and pragmaticalization ». In Fischer S., Gabriel C., *Manual of Grammatical Interfaces in Romance*. Berlin : de Gruyter Mouton, p. 635-657.
- Deulofeu H.-J. 1989. « Les couplages de constructions verbales en français parlé : effet de cohésion discursive ou syntaxe de l'énoncé ». *Recherches sur le français parlé* 9, p. 111-141.
- Deulofeu H.-J. 1999. *Recherche sur les formes de la prédication dans les énoncés assertifs en français contemporain. Le cas des énoncés introduits par le morphème que*. Thèse d'état de l'Université Paris 3.
- De Vogüé S. 1999. « Le champ des subordonnées dites conditionnelles du français : conditions, éventualités, suppositions et hypothèses ». *Linx* 41, 22 p. [En ligne](#)
- Di Cristo A. 1998. « Intonation in French ». In Hirst D., Di Cristo A. (eds), *Intonation Systems. A Survey of Twenty Languages*. Cambridge : Cambridge University Press, p. 195-218.
- Di Cristo A. 2004. « La prosodie au carrefour de la phonétique, de la phonologie et de l'articulation formes-fonctions ». *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence (TIPA)* 23, p. 67-211.
- Di Cristo A. 2013. *La prosodie de la parole*. Bruxelles : De Boeck.
- Diewald G. 2012. « Grammaticalization and pragmaticalization. » In Heine B., Narrog H. (eds), *The Oxford Handbook of Grammaticalization*. Oxford Handbooks Online, 8 p.
- Diewald G. 2013. « “Same same but different” Modal particles, discourse markers and the art (and purpose) of categorization ». In Degand L., Cornillie B., Pietrandrea P. (eds), *Discourse Markers and Modal Particles. Categorization and description*. Amsterdam-Philadelphia : John Benjamins, p. 19-46.
- Dijk T. A. van 2014. *Discourse and knowledge : A sociocognitive approach*. Cambridge : Cambridge University Press.
- Dostie G. 2004. *Pragmaticalisation et marqueurs discursifs. Analyse sémantique et traitement lexicographique*. Liège : De Boeck/Duculot.
- Dostie G., Léard J.-M. 1985. « Particularités syntaxiques et sémantiques des conditionnelles en québécois ». *Revue Québécoise de Linguistique* 15/1, p. 119-152.
- Ducrot O. 1972. *Dire et ne pas dire. Principes de sémantique linguistique*. Paris : Hermann.
- Ducrot O. 1983. « Opérateurs argumentatifs et visée argumentative ». *Cahiers de Linguistique Française* 5, p. 7-36.
- Ducrot O. 1984. *Le dire et le dit*. Paris : Éditions de Minuit.
- Ducrot O. et al. 1980. *Les mots du discours*. Paris : Les éditions de minuit.
- Dugua C. & Kanaan-Caillol L. 2021. « Du recueil à l'outillage des corpus oraux : comment accéder à la variation ? Présentation ». *Corpus* 22, 8 p. [En ligne](#)

## E

- Eco U. 1985 [1979]. *Lector in fabula. Le rôle du lecteur ou la Coopération interprétative dans les textes narratifs*. Paris : Grasset.
- van Eemeren F.H., Garssen, B., Krabbe E.C.W., Snoeck Henkemans F.A., Verheij B., Wage-mans J.H.M. 2013. *Handbook of Argumentation Theory*. Springer.
- Eensoo E. & Valette M. 2014. « Approche textuelle pour le traitement automatique du discours évaluatif ». *Langue française* 184, p. 109-124.
- Eensoo E., Nouvel D., Martin A., Valette M. 2016. « Combiner analyses textométriques, apprentissage supervisé et représentation vectorielle pour l'analyse de la subjectivité ». *11<sup>e</sup> Défi Fouille de Texte (DEFT'2015)*, Caen, 9 p. [En ligne](#)
- Eibe F., Hall M. A. & Witten I. H. 2016. *The WEKA Workbench. Online Appendix for 'Data Mining: Practical Machine Learning Tools and Techniques'*. Morgan Kaufmann, Fourth Edition. [En ligne](#)

Erman B. & Kostinas U.-B. 1993. « Pragmatization: The case of *ba'* and *you know* ». *Studier I Modernsprakvetenskap* (Stockholm studies in modern philology) 10, p. 76-93.

## F

- Fagard B. 2010. « *E vida, olha...* : Imperatives as discourse markers and grammaticalization paths in Romance : A diachronic corpus study ». *Languages in Contrast* 10 (2), p. 214-267.
- Fauconnier G. (1985). *Mental Spaces*. Cambridge : Cambridge University Press.
- Ferguson C. A. 1959. « Diglossia ». *Word* 15, p. 325-340.
- Fernandez-Vest J. 1994. *Les particules énonciatives dans la construction du discours*. Paris : Presses Universitaires de France.
- Fetzer A. 2012. « Contexts in interaction : Relating pragmatic wastebaskets ». In Finkbeiner R. et al. (eds.), *What is a context? Linguistic approaches and challenges*. Amsterdam-Philadelphia : John Benjamins, p. 105-127.
- Fetzer A. 2017. « Context ». In Huang Y. (ed.), *The Oxford Handbook of Pragmatics*. Oxford : Oxford University Press, p. 259-276.
- Fillietaz I. & Roulet E. 2002. « The Geneva Model of Discourse Analysis: an interactionist and modular approach to discourse organization ». *Discourse Studies* 4(3), p. 369-392.
- Fillmore J. C. 1988. « The mechanisms of 'Construction Grammar' ». *Berkeley Linguistic Society* 14, 35-55.
- Finegan E. 1995. « Subjectivity and subjectivization : an introduction ». In Stein D., Traugott E. (eds), *Subjectivity and subjectivization. Linguistic perspectives*. Cambridge : Cambridge University Press, p. 1-15.
- Finkbeiner R., Meibauer J. & Schumacher P. (eds.) 2012. *What is a context? Linguistic approaches and challenges*. Amsterdam-Philadelphia : John Benjamins.
- von Fintel K. & Iatridou S. 2007. « Anatomy of a Modal Construction ». *Linguistic Inquiry* 38/3, p. 445-483.
- Fischer K. & Drescher M. 1996. « Methods for the Description of Discourse Particles : Contrastive Analysis ». *Language Sciences* 18/3, p. 853-861.
- Fischer K. 2000. *From Cognitive Semantics to Lexical Pragmatics. The Functional Polysemy of Discourse Particles*. Berlin/New York : de Gruyter.
- Fischer K. 2006. « Towards an understanding of the spectrum of approaches to discourse particles : introduction to the volume ». In *Approaches to Discourse Particles*, Fischer K. (ed.). Amsterdam : Elsevier, p. 1-20.
- Fischer K. (ed.) 2006. *Approaches to Discourse Particles*. Amsterdam : Elsevier.
- Fisette J. 1996. *Pour une pragmatique de la signification*, Montréal, XYZ éditeur.
- Fléchon G., Frassi P. & Polguère A. 2012. « Les pragmatèmes ont-ils un charme indéfinissable ? » In Ligas P., Frassi P. *Lexiques. Identités. Cultures*, QuiEdit, p. 81-104.
- Foolen A. 1996. « Pragmatic particles ». In Verschueren J. et al., *Handbook of Pragmatics*. Amsterdam-Philadelphia : John Benjamins, p. 1-24.
- Forget D. 1979. « Quel est le français standard au Québec ? ». In Thibault P. (dir.) *Le français parlé : études sociolinguistiques*. Carbondale-Edmonton : Linguistic Research Inc, p. 153-161.
- Frank-Job B. 2006. « A dynamic-interactional approach to discourse markers ». In Fischer K. (ed.), *Approaches to discourse particles*. Amsterdam : Elsevier., p. 395-413.
- Franke M. 2008. « Pseudo-imperatives and other cases of conditional conjunction and conjunctive disjunction ». In Fabricius Hansen C., Ramm W. (eds), *'Subordination' versus 'Coordination' in Sentence and Text. A cross-linguistic perspective*. Amsterdam-Philadelphia : John Benjamins, p. 255-279.
- Fraser B. 1999. « What are discourse markers ? ». *Journal of Pragmatics* 31, p. 931-952.
- Fraser B. 2009. « An account of discourse markers ». *International Review of Pragmatics* 1, p. 293-320.
- Frédéric M. 1997. *La stylistique française en mutation ?* Bruxelles : Académie Royale de Belgique.
- Freeman J. B. 1991. *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure*. Foris Publications.
- Freeman J. B. 2011. *Argument Structure: Representation and Theory*. Cham : Springer.
- Frege G. 1892. « Über Sinn und Bedeutung ». *Zeitschrift für Philosophie und philosophische Kritik* 100, p. 25-50.
- Fried M. & Östman J.O. 2004. « A thumbnail sketch of Construction Grammar ». In Fried M., Östman J.O. (éds.), *Construction Grammar in a Cross-Language Perspective*. Amsterdam : John Benjamins, p. 11-86.

## G

- Gadet F. 2003. *La variation sociale en français*. Paris : Ophrys.
- Gadet F. 2020. « Langue et variation ». In *Encyclopédie Grammaticale du Français*. [En ligne](#)
- Gadet F. 2021. « Qu'est-ce que le français ? ». In Abeillé A., Godard D. (dir.), *Grande Grammaire du Français*. Arles : Actes Sud, tome 1, p. xix-xxii.

- Galburn E. & Miettinen P. 2012. « From black and white to full color : extending redescription mining outside the Boolean world ». *Statistical Analysis and Data Mining : The ASA Data Science Journal* 5(4), p. 284-303.
- Galburn & Miettinen 2017. *Redescription Mining*. Cham : Springer.
- Gardes-Tamine J. 1996. *La rhétorique*. Paris : Armand Colin.
- Gardes-Tamine J. 1997. *La stylistique*. Paris : Armand Colin.
- Gasquet-Cyrus M. 2004. *Pratiques et représentations de l'humour verbal. Étude sociolinguistique du cas marseillais*. Thèse de l'université de Provence.
- Gauvin L. 2000. *Langagement, l'écrivain et la langue au Québec*. Montréal : Boréal.
- Gazdar G. 1981. « Unbounded Dependencies and Coordinate Structure ». *Linguistic Inquiry* 12, p. 155-184.
- Genette G. 1972. *Figures III*. Paris : Éditions du Seuil.
- Genette G. 2004. *Métalepse. De la figure à la fiction*. Paris : Seuil.
- Gervais A. (dir.). 2000. *Emblématiques de l'« époque du joual », Jacques Renaud, Gérald Godin, Michel Tremblay, Yvon Deschamps*. Montréal : Lanctôt éditeur.
- Ghezzi C. 2014. « The development of discourse and pragmatic markers ». In Ghezzi C., Molinelli P. (eds), *Discourse and Pragmatic Markers from Latin to the Romance Languages*. Oxford : Oxford University Press, p. 10-26.
- Ghezzi C. & Molinelli P. 2014. « Discourse and Pragmatic Markers from Latin to the Romance Languages : New Insights ». In Ghezzi C., Molinelli P. (eds), *Discourse and Pragmatic Markers from Latin to the Romance Languages*. Oxford : Oxford University Press, p. 1-9.
- Ghezzi C. & Molinelli P. (eds) 2014. *Discourse and Pragmatic Markers from Latin to the Romance Languages*. Oxford : Oxford University Press.
- Gicquel B. 1999. *Stylistique littéraire et informatique*. Arras : Artois Presses Université.
- Gilbert M. A. 1997. *Coalescent Argumentation*. Mahwah : Lawrence Erlbaum Associates Publishers.
- Ginzburg J. 2012. *The Interactive Stance. Meaning for Conversation*. Oxford : Oxford University Press.
- Giron S. 2004. *Inventaire et classement des constructions verbales dans un corpus de français parlé dans l'Allier*. Thèse de l'Université Clermont-Ferrand 2.
- Givón T. 2005. *Context as Other Minds*. Amsterdam-Philadelphia : John Benjamins.
- Glikman J. 2005. *La parataxe. Définition et évolution*. Mémoire de DEA de l'Université Paris 10.
- Goldberg A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago : University of Chicago Press.
- Goldberg A. 2006. *Constructions at Work: the nature of generalization in language*. Oxford : Oxford University Press.
- Gomez-Txurruka I. 2003. « The natural language conjunction *and* ». *Linguistics and Philosophy* 26, p. 255-286.
- Greimas A. J. 1966. *Sémantique structurale. Recherche de méthode*. Paris : Larousse.
- Grice P. 1975 [1989]. « Logic and conversation ». In Davidson, D., Harman, G. (eds), *The Logic of Grammar*. Ecino (CA) : Dickenson, p. 64-75.
- Grice P. 1989. *Studies in the way of words*. Cambridge/London : Harvard University Press.
- Gries S. T. 2013. « 50-something years of work on collocations: what is or should be next... » *International Journal of Corpus Linguistics* 18 (1), p. 137-165.
- Gries S. T. 2019. « 15 years of collocations: some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures) ». *International Journal of Corpus Linguistics* 24 (3), p. 385-412.
- Grieve J. 1995. *Dictionary of contemporary French connectors*. Londres : Routledge.
- Gross G. & Nazarenko A. 2004. « Quand la langue cause : paramètres d'une analyse linguistique ». *Intellectica. Revue de l'Association pour la Recherche Cognitive* 38, p. 15-41.
- Groupe d'Entrevernes. 1979. *Analyse sémiotique des textes. Introduction, théorie, pratique*. Lyon : Presses Universitaires de Lyon.
- Groupe λ. 1975. « Car, parce que, puisque ». *Revue Romane* 10/2, p. 248-280.
- Gruber H. & Redecker G. (eds) 2014. *The Pragmatics of Discourse Coherence. Theories and applications*. Amsterdam-Philadelphia: John Benjamins.
- Gruber H. & Redecker G. 2014. « Introduction : the pragmatics of discourse coherence ». In Gruber H., Redecker G. (eds), *The Pragmatics of Discourse Coherence. Theories and applications*. Amsterdam-Philadelphia: John Benjamins, p. 1-20.
- Gülich E. 1970. *Makrosyntax der Gliederungssignale im gesprochenen Französisch*. Munich : Wilhelm Fink. (cité par Hansen 1998).
- Gutzmann D. 2013. « Expressives and beyond : An introduction to varieties of use-conditional meaning ». In Gutzmann, D., Gärtner, H.-M. (eds), *Beyond expressives : Explorations in use-conditional meaning*. Leiden : Brill, p. 1-58.
- Gutzmann D. 2015. *Use-conditional meaning. Studies in multidimensional semantics*. Oxford : Oxford University Press.
- Gutzmann D. & Turgay K. (eds). 2019. *Secondary Content. The Semantics and Pragmatics of Side Issues*. Leiden-Boston : Brill.

## H

- Habert B., Nazarenko A. & Salem A. 1997. *Les linguistiques de corpus*. Paris : Armand Colin/Masson.
- Hafed Dahou A. 2021. *Discourse Marker Identification in French Spoken Corpora : Using Rule-Based Method and Machine Learning*. Mémoire de stage de M2 TAL, IDMC Nancy.
- Halliday M.A.K. & Hasan, R. 1976. *Cohesion in English*. London : Longman.
- Hancil S. (éd.) 2011. *Marqueurs discursifs et subjectivité*. Rouen : Presses Universitaires de Rouen et du Havre.
- Hancil S. & König E. (eds). 2014. *Grammaticalization – Theory and Data*. Amsterdam-Philadelphia: John Benjamins.
- Hansen M.-B. M. 1997. « *Alors and donc in spoken French. A reanalysis* ». *Journal of Pragmatics* 28, p. 153-187.
- Hansen M.-B. M. 1998a. *The Function of discourse particles*. Amsterdam-Philadelphia : John Benjamins.
- Hansen M.-B. M. 1998b. « La grammaticalisation de l'interaction ou Pour une approche polysémique de l'adverbe *bien* ». *Revue de Sémiotique et Pragmatique* 4, 111-138.
- Hansen M.B. M. 2005. « From prepositional phrase to hesitation marker. The semantic and pragmatic evolution of French *enfin* ». *Journal of Historical Pragmatics* 6/1, p. 37-68.
- Hansen M.-B. M. 2008. *Particles at the Semantics/pragmatics Interface: Synchronic and Diachronic Issues, a Study with Special Reference to the French Phrasal Adverbs*, Elsevier.
- Hansen M.B. M. à par. « Les pragmatèmes ». In Buchi E. (éd.), *Manuel d'étymologie lexicale romane*. Berlin : De Gruyter Mouton, 16 p.
- Hazaël-Massieux M.-C. 1993. *Écrire en créole. Oralité et écriture aux Antilles*. Paris : Éditions L'Harmattan.
- Heiden S. 2002. *Weblex, Manuel utilisateur, version 4.1*, Lyon, CNRS/ENS-LSH.
- Heiden S., Magué J.-P., & Pincemin B. 2010. « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement ». *JADT 2010 : 10th International Conference on the Statistical Analysis of Textual Data*, 12 p. [En ligne](#).
- Heine B. 2013. « On discourse markers: Grammaticalization, pragmaticalization, or something else ? ». *Linguistics* 51 (6), p. 1205-1247.
- Heine B., Kaltenböck G., Kuteva T. & Long H. 2021a. « On the rise of discourse markers ». In Haselow A., Hancil S. (eds), *Studies at the Grammar-Discourse Interface: Discourse markers and discourse-related grammatical phenomena*. Amsterdam-Philadelphia : John Benjamins, p. 23-55.
- Heine B., Kaltenböck G., Kuteva T. & Long H. 2021b. *The Rise of Discourse Markers*. Oxford : Oxford University Press.
- Hilpert M. 2014. *Construction Grammar and its Application to English*. Edinburgh : Edinburgh University Press.
- Hirschberg J. & Litman D. 1993. « Empirical Studies on the Disambiguation of Cue Phrases ». *Computational Linguistics* 19, p. 501-530.
- Hockett C. F. 1960. « The Origin of Speech ». *Scientific American* 203, p. 88-111.
- Hoffmann T. 2013. « Abstract phrasal and clausal constructions. In Heine B., Narrog H. (eds), *The Oxford Handbook of Grammaticalization*. Oxford : Oxford University Press, p. 230-245.
- Hoffman T. & Trousdale G. (eds) 2013. *The Oxford Handbook of Construction Grammar*, Oxford : Oxford University Press.
- Horn L. R. 1985. « Toward a New Taxonomy for Pragmatic Inference: Q-Based and R-Based Implicature ». In Schiffrin D. (ed.), *Meaning, Form, and Use in Context: Linguistic Applications (GURT '84)*. Washington : Georgetown University Press, p. 11-42.
- Horn L. R. 1989. *A Natural History of Negation*. Chicago : Chicago University Press.
- Houdebine A.-M. (dir.) 2002. *L'imaginaire linguistique*. Paris : L'Harmattan.
- Huang Y. 2015. *Pragmatics*, 2<sup>nd</sup> ed. Oxford : Oxford University Press.
- Huang Y. (ed.) 2017. *The Oxford Handbook of Pragmatics*. Oxford : Oxford University Press.
- Husianyia M. 2011. *Caractérisation de types de discours dans des situations de travail*. Thèse de l'Université Nancy 2, 504 p.
- Hybertie C. 1996. *La conséquence en français*. Paris : Ophrys.
- Hymes D. 2001 [1977]. *Foundations of Sociolinguistics. An Ethnographic Approach*. Oxon-New York : Routledge.

## I

- Iten C. 2005. « *Although revisited* », 33 p. [En ligne](#).

## J

- Janssen T.M.V. 2012. « Compositionality: its historic context ». In Werning M., Hinzen W., E. Machery (eds), *The Oxford Handbook of Compositionality*. Oxford : Oxford University Press, p. 19-46.
- Jayez J. 2010. « Projective meaning and attachment ». In *Logic, Language and Meaning. Revised selected papers of the 17th Amsterdam Colloquium, Amsterdam 2009*, n° 6042 in Lecture Notes in

- Artificial Intelligence. Berlin : Springer, p. 325-334.
- Jayez J. & Rossari C. 2001. « The discourse level sensitivity of consequence discourse markers in French ». *Cognitive Linguistics* 12, p. 275-290.
- Jouve V. 1992. *L'effet-personnage dans le roman*. Paris : Presses Universitaires de France.
- Jucker A. & Ziv Y. (eds) 1998. *Discourse Markers, Description and theory*. Amsterdam-Philadelphia: John Benjamins.
- Jun S.-A. & Fougeron C. 2000. « A phonological model of French intonation ». In Botinis A. (ed.), *Intonation : Analysis, Modelling and Technology*. Dordrecht : Kluwer, p. 209-242.
- Kerbrat-Orecchioni C. 2009 [1980]. *L'énonciation*. Paris : Armand Colin.
- Kleiber G. 2006. « Sémiotique de l'interjection ». *Langages* 161, p. 10-23.
- Knott A. & Sanders T. 1998. « The Classification of Coherence Relations and their Linguistic Markers : An Exploration of Two Languages ». *Journal of Pragmatics* 30/2, p. 135-175.
- Koch P. & Osterreicher W. 2001. « Gesprochene Sprache und geschriebene Sprache. Langage parlé et langage écrit ». In Holtus G., Metzelin M., Schmitt C. (eds), *Romanistischen Linguistik*. Tübingen : Max Niemeyer, p. 584-627.
- Kutas M. & Federmeier K. D. 2011. « Thirty years and counting : Finding meaning in the N400 component of the event-related brain potential ». *Annual Review of Psychology* 62, p. 621-647.

## K

- Kahane S. & Mazziotta N. 2015. « Quel classement syntaxique pour les « marqueurs discursifs », « mots-phrases » et autres « inserts » prédicatifs et locutifs ? ». *Travaux de linguistique* 71, p. 7-42.
- Kaltenböck G., Heine G. & Kuteva T. 2011. « On thetical grammar ». *Studies in Language* 35 (2), p. 848-893.
- Kaplan R. 1999. « The meaning of *ouch* and *oops* ». Transcription d'une conférence. [En ligne](#)
- Karagjosova E. 2003. « Modal particles and the common ground: Meaning and function of German *ja*, *doch*, *eben/halt* and *auch* ». In Kühnlein P., Rieser H., Zeevat H. (eds), *Perspectives on Dialogue in the New Millennium. Pragmatics and Beyond*. Amsterdam : John Benjamins, p. 335-349.
- Karttunen L. 2016. « Presupposition : What went wrong ». In Monorey, M., Little, C.-R., Collard, J., Burgdorf, D. (eds), *Proceedings of Semantics and Linguistic Theory (SALT)* 26, p. 705-731.
- Kauffer M. & Keromnes Y. (eds). 2019. *Theorie und Empirie in der Phraseologie. Approches théoriques et empiriques en phraséologie*. Tübingen : Stauffenburg.
- Kay P. 1995. « Construction Grammar ». In Verschueren J. et al. (eds), *Handbook of Pragmatics*. Amsterdam-Philadelphia : John Benjamins, p. 171-177.
- Kay P. 2002. « An Informal Sketch of a Formal Architecture for Construction Grammar ». *Grammars* 5, p. 1-19.
- Kerbrat-Orecchioni C. 2000. « Quelle place pour les émotions dans la linguistique au XX<sup>e</sup> siècle ? Remarques et aperçus ». In Plantin C., Doury M., Traverso V. (éd.), *Les émotions dans les interactions*. Lyon : Presses Universitaires de Lyon, p. 34-74.

## L

- Laberge S. 1971. *Étude de la variation des pronoms sujets définis et indéfinis dans la français parlé à Montréal*. Montréal : Département d'anthropologie de l'université de Montréal.
- Lacheret-Dujour A. & Beaugendre F. 1999. *La prosodie du français*. Paris : CNRS Éditions.
- Lacheret A. 2007. « Prosodie-discours : une interface à multiples facettes ». *Nouveaux Cahiers de linguistique française* 28, p. 7-40.
- Ladd D. R. 2008. *Intonational Phonology*. 2nd ed. Cambridge : Cambridge University Press.
- Langendoen D. T. & Savin H. B. 1971. « The projection problem for presupposition ». In Fillmore, J., Langendoen, D. T. (eds), *Studies in Linguistic Semantics*. New York : Chicago-San Francisco : Holt, Rinehart and Winston, p. 54-60.
- Larthomas P. 2001 [1972]. *Le Langage dramatique, sa nature, ses procédés*. Paris : Presses Universitaires de France.
- Lascarides A. & Copestake A. 1999. « Default Representation in Constraint-Based Frameworks ». *Computational Linguistics* 25, p. 55-105.
- Laurendeau P. 1985. « Description du marqueur d'opérations *coudon* dans le cadre d'une théorie énonciative ». *Revue québécoise de linguistique appliquée* 15/1, p. 79-117.
- Léard J.-M. 1991. « Pour une description synthétique du québécois : le cas de *coudon* et des mots du discours ». *Revue québécoise de linguistique théorique et appliquée* 10/3, p. 133-152.
- Leech G. N. 1989 [1983]. *Principles of Pragmatics*. London : Longman.
- Legallois D. 2016. « La notion de construction ». *Encyclopédie grammaticale du français*. [En ligne](#).

- Legallois D. & Gréa P. 2003. « La Grammaire de construction ». *Cahiers du Crisco* 21, p. 5-27.
- Lehmann C. 1985. « Grammaticalization: synchronic variation and diachronic change ». *Lingua e stile* 20 (3), p. 303-18.
- Lehmann C. 2015 [1982]. *Thoughts on Grammaticalization* (3<sup>rd</sup> ed.). Berlin : Language Science Press.
- Lenepveu V. 2007. « Toutefois et néanmoins, une synonymie partielle ». *Syntaxe et Sémantique* 8, p. 91-106.
- Levelt W. J. M. 1989. *Speaking : From Intention to Articulation*. ACL-MIT Press Series in Natural Language Processing : MIT Press.
- Lewis T. 2011. « A discourse constructional approach to the emergence of discourse markers in English ». *Linguistics* 49 (2), p. 415-443.
- Li S., Wang L., Cao Z., & Li W. 2014. « Text-level discourse dependency parsing ». In *Proceedings of the 52<sup>nd</sup> Annual Meeting of the ACL*, vol. 1, p. 25-35.
- Lopez Villegas C. 2019. « Discourse and Pragmatic Markers in Spoken Written Everyday Life ». [En ligne](#)
- Lyons J. 1977. *Semantics*, 2 volumes. Cambridge : Cambridge University Press.
- Analysis*, 2 vol., 2<sup>nd</sup> ed. Oxford: Wiley & Blackwell, p. 189-221.
- Mathieu Y. 2003. « La Grammaire de construction ». *Linx* 48, p. 43-56.
- Mayaffre D. 2006. « Philologie et/ou herméneutique numérique : nouveaux concepts pour de nouvelles pratiques ? ». *XXVII<sup>e</sup> Colloque d'Albi Langages et Signification*, p.15-25. [En ligne](#)
- Meibauer J. 2012. « What is a context ? Theoretical and empirical evidence ». In Finkbeiner R., Meibauer J., Schumacher P. (eds.), *What is a context? Linguistic approaches and challenges*. Amsterdam-Philadelphia : John Benjamins, p. 9-32.
- Mel'čuk I. 2013. « Tout ce que nous voulions savoir sur les phrasèmes, mais... » *Cahiers de lexicologie* 102, p. 129-149.
- Mesbahi L., Jouvét D., Bonneau A., Fohr D., Illina I. & Laprie Y. 2011. « Reliability of non-native speech automatic segmentation for prosodic feedback ». *Proceedings of SLaTE 2011*.
- Meschonnic H. 1987. « Littérature et oralité ». *Présence francophone* 31, p. 9-30.
- Métrich R., Faucher E. & Courdier G. 1998-2002. *Invariables difficiles, Dictionnaire allemand-français des particules, connecteurs, interjections et autres mots de la communication*, 4 vol. Nancy : Groupe de lexicographie franco-allemande de l'Université de Nancy 2 et ATILF.
- Meyer M. 1995. *De la problématique. Philosophie, sciences et langage*. Bruxelles : Mardaga.
- Moeschler J. & Reboul A. 1994. *Dictionnaire encyclopédique de pragmatique*. Paris : Seuil.
- Molinié G. & Cahné P. (dir.). 1994. *Qu'est-ce que le style ?* Paris : Presses Universitaires de France.
- Molinié G. 1998. *Sémiosylistique, L'effet de l'art*. Paris : Presses Universitaires de France.
- Molino J. 1989. « Interpréter ». In Reichler C. (dir.), *L'interprétation des textes*. Paris : Les Éditions de Minuit, p. 9-52.
- Molino J. 1994. « Pour une théorie sémiologique du style ». In Molinié G., Cahné P. (dir.), *Qu'est-ce que le style ?* Paris : Presses Universitaires de France, p. 213-261.
- Mougeon F. 1995. *Quel français parler ? Initiation au français parlé au Canada et en France*. Toronto : Éditions du Gref, Centre universitaire Glendon, Université York.
- Morel M.-A. 1996. *La Concession en français*. Paris : Ophrys.
- Morel M.-A. & Danon-Boileau L. 1998. *Grammaire de l'intonation*. Paris : Ophrys.
- Müller S. 2005. *Discourse Markers in Native and Non-native English Discourse*. Amsterdam-Philadelphia : John Benjamins.
- Müller S. 2021. « HPSG and Construction Grammar ». In Müller S., Abeillé A., Borsley R. D., Koenig J.-P. (eds), *Head-Driven Phrase*

## M

- Mahrer R. 2017. *Phonographie. La représentation écrite de l'oral en français*. Berlin-Boston : Walter de Gruyter.
- Maingueneau D. 1991. *Pragmatique pour le texte littéraire* : Paris, Dunod.
- Mann W. & Thompson S. 1988. « Rhetorical Structure Theory : Toward a functional theory of text organization ». *Text* 8(3), p. 243-281.
- Marandin J.-M. 2006. « Contours as Constructions ». *Constructions* 1, p. 28 p. [En ligne](#)
- Marchello-Nizia C. 2006. *Grammaticalisation et changement linguistique*. Bruxelles : De Boeck.
- Martin P. 2009. *Intonation du français*. Paris : Armand Colin.
- Martin P. 2015. *The Structure of Spoken Language. Intonation in Romance*. Cambridge : Cambridge University Press.
- Martin P. 2018 *Intonation, structure prosodique et ondes cérébrales. Introduction à l'analyse prosodique*. Paris : ISTE Éditions.
- Marchello-Nizia C., Combettes B., Prévost S. & Scheer T. (éds). 2020. *Grande grammaire historique du français*. Berlin : de Gruyter Mouton.
- Maschler Y. & Schiffrin D. 2015. « Discourse Markers, Language, Meaning and Context ». In Tannen D., Hamilton H. E., Schiffrin D. (eds), *The Handbook of Discourse*



*Structure Grammar: The Handbook*. Berlin : Language Science Press, chap. 32, p. 1497-1553.  
Müller S., Abeillé A., Borsley R. D. & Koenig J.-P. (eds). 2021. *Head-Driven Phrase Structure Grammar: The Handbook*. Berlin: Language Science Press. [En ligne](#)

## N

- Narrog H. & Heine B. (eds) 2012. *The Oxford Handbook of Grammaticalization*. Oxford : Oxford University Press.  
Nazarenko A. 2000. *La cause et son expression en Français*. Paris : Ophrys.  
Némo F. 2000. « *Enfin, encore, toujours* entre indexicalité et emplois ». *Actes du XXII<sup>e</sup> Congrès international de linguistique et de philologie romanes* (Bruxelles, juillet 1998). Tübingen : Max Niemeyer Verlag, p. 499-511.  
Nölke H. 1990. « Les adverbiaux contextuels : problèmes de classification ». *Langue française* 88, p. 12-27.  
Nouvel D. 2012. *Reconnaissance des entités nommées par exploration de règles d'annotation. Interpréter les marqueurs d'annotation comme instructions de structuration locale*. Thèse de l'université François Rabelais.

## O

- Onea E. 2016. *Potential Questions at the Semantics-Pragmatics Interface*. Brill.  
Östman, J.-O. 1981. *'You know' : A Discourse Functional Approach*. Amsterdam-Philadelphia : John Benjamins.  
Östman, J.-O. 1995. « Pragmatic particles twenty years after ». In Wärvik, B., Tanskanen, S.-K., Hiltunen R. (eds). *Organization in discourse. Proceedings from the Turku Conference*. University of Turku, Turku, Finland, p. 95-108.

## P

- Padilla Cruz M. 2016. « Three Decades on Relevance Theory ». In Padilla Cruz M. (ed.), *Relevance Theory. Recent Developments, Current Challenges and Future Directions*. Amsterdam-Philadelphia : John Benjamins, 32 p.  
Paillard D. 1998. « Les mots du discours comme mots de langue. » *Le Gré des langues* 14, p. 10-41.  
Paillard D. 2011. « Marqueurs discursifs et scène énonciative ». In Hancil S. (éd.), *Connecteurs discursifs*. Rouen : Presses Universitaires de Rouen, p. 13-39.  
Paillard D. 2017. « Scène énonciative et types de marqueurs discursifs ». *Langages* 207, p. 5-16.  
Pavis P. 1996. *Dictionnaire du théâtre*, Paris, Dunod.

- Peldszus A. & Stede M. 2013. « From argument diagrams to automatic argument mining : A survey ». *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1), p. 1-31.  
Peldszus A. & Stede M. 2015. « An annotated corpus of argumentative microtexts ». *First European Conference on Argumentation: Argumentation and Reasoned Action*, Portugal, Lisbon, June 2015.  
Peldszus A. & Stede M. 2016. « Rhetorical structure and argumentation structure in monologue text ». *Proceedings of the 3rd Workshop on Argumentation Mining*, ACL, Berlin, August 2016, p. 103-112.  
Perelman C. 1969. *L'empire rhétorique*. Bruxelles : Presses de l'Université de Bruxelles.  
Perelman C. & Olbrechts-Tyteca L. 1958. *Traité de l'argumentation. La nouvelle rhétorique*. Bruxelles : Presses de l'Université de Bruxelles.  
Péroz P. 1992. *Systématique des valeurs de bien en français contemporain*. Genève : Droz.  
Petit M. 2010. « Discrimination prosodique et représentation du lexique : les connecteurs discursifs ». *Études de linguistique appliquée* 157, p. 75-93.  
Petitjean A. 1999. « Approches sémio-linguistiques du texte théâtral ». In Lautel A., Castellana M., *Le théâtre du sens, actes du S.I.R.F.A.S*. Arras : Artois Presse Université, p.43-56.  
Pier J. & Schaeffer J.-M. (dir.) 2005. *Métalepses. Entorses au pacte de la représentation*. Paris : Éditions de l'EHESS.  
Pierrehumbert J. & Hirschberg J. 1990. « The meaning of intonational contours in the interpretation of discourse ». In Cohen P. R., Morgan J., Pollack M. E. (eds), *Intentions in communication*. Cambridge : MIT Press, p. 271-311.  
Pilehvar M. T. & Camacho-Collados J. 2021. *Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning*. Morgan & Claypool.  
Pillet E. 2011a. « Paroles pour rire ». *Humoresques* 33, p. 5-15.  
Pillet E. 2011b. « Humour et décalages : Gad Elmaleh ». *Humoresques* 33, p. 59-69.  
Plantin C. 1990. *Essais sur l'argumentation. Introduction à l'étude linguistique de la parole argumentative*. Paris : éditions Kimé.  
Plantin C. 2002. « Argumentation studies and discourse analysis. The French situation and global perspectives ». *Discourse Studies* 4/3, p. 343-368.  
Plantin C. 2003. « Argumentation studies in France. A new legitimacy ». In van Eemeren F. H. et al. (eds), *Anyone who has a view. Theoretical contributions to the study of argumentation*. Dordrecht : Kluwer, p. 173-187.

- Plantin C. 2011. *Les bonnes raisons des émotions. Principes et méthodes pour l'étude du discours émotionné*. Bern : Peter Lang.
- Plourde M. (dir.) 2000. *Le français au Québec, 400 ans d'histoire et de vie*. Montréal : Fides/Les Publications du Québec, Conseil de la Langue Française.
- Polanyi L. 1985. « A Theory of Discourse Structure and Discourse Coherence ». *Chicago Linguistics Society* 21, p. 306-322.
- Polanyi L. & van den Berg M. H. 1996. « Discourse Structure and Discourse Interpretation ». In Dekker P., Stokhof M., *Proceedings of the Tenth Amsterdam Colloquium*, ILLC/Department of Philosophy, University of Amsterdam, p. 113-131.
- Polanyi L., van den Berg M., Ahn D. 2003. « Discourse Structure and Sentential Information Structure. An Initial Proposal ». *Journal of Logic, Language and Information* 12, p. 337-350.
- Pollard C. & Sag I. 1994. *Head-driven Phrase Structure Grammar*. Chicago : University of Chicago.
- Pollock J. L. 1995. *Cognitive Carpentry : a Blueprint for How to Build a Person*. Cambridge, MIT Press.
- Poplack S. 2021. « Le corpus comme portail pour l'étude de la variation (socio)linguistique », *Corpus* 22, 16 p. [En ligne](#)
- Portes C., Bertrand R. & Espesser R. 2007. « Contribution to a grammar of intonation in French. Form and function of three rising patterns ». *Nouveaux Cahiers de Linguistique Française* 28, p. 155-162.
- Portes C., Beyssade C., Michelas A., Marandin J.-M., Champagne-Lavau M. 2014. « The Dialogical Dimension of Intonational Meaning: Evidence from French ». *Journal of Pragmatics* 74, p. 5-29.
- Post B. 2000. *Tonal and phrasal structures in French intonation*. The Hague : Holland Academic Graphics.
- Potts C. 2005. *The Logic of conventional implicatures*. Oxford : Oxford University Press.
- Potts C. 2007. « The expressive dimension ». *Theoretical Linguistics* 33, p. 255-268.
- Prasad R., Dinesh N., Lee A., Miltsakaki E., Robaldo L., Joshi A., Webber B. 2008. « The Penn Discourse Treebank 2.0 ». In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. Stroudsburg : Association for Computational Linguistics, p. 2961-2968. [En ligne](#)
- Prevost S. & Dufresne M. 2020. « Mécanismes et processus à l'œuvre ; causes et motivations ». In Marchello-Nizia C. et al. (éd.), *Grande Grammaire Historique du Français*. Berlin-Boston : De Gruyter Mouton, section 2.5, p. 82-89.+
- Prevost S. & Fagard B. 2007. « Grammaticalisation et lexicalisation : la formation d'expression complexes ». *Langue française* 156, p. 3-8.

## R

- Radvansky G. A. & Zacks J. M. 2014. *Event cognition*. Oxford : Oxford University Press.
- Rastier 2001. *Arts et sciences du texte*. Paris : Presses Universitaires de France.
- Rastier F. 2005. « Enjeux épistémologiques de la linguistique de corpus ». In Williams G. (dir.), *La Linguistique de corpus*. Rennes : Presses Universitaires de Rennes. Aussi [en ligne](#)
- Rastier F. 2011. *La mesure et le grain. Sémantique de corpus*. Paris : Champion.
- Razgoulieva A. 2002. « Combinaison des connecteurs mais enfin ». *Cahiers de Linguistique Française* 24, p. 143-168.
- Redeker G. 1990. « Ideational and pragmatic markers of discourse structure ». *Journal of Pragmatics* 14, p. 367-381.
- Renkema J. 2009. *The texture of discourse*. Amsterdam-Philadelphia : John Benjamins.
- Renkema J. & Schubert C. 2018. *Introduction to Discourse Studies* (new edition). Amsterdam-Philadelphia : John Benjamins.
- Riegel M., Pellat J.-C. & Rioul R. 2009. *Grammaire méthodique du français* 4<sup>e</sup> éd. Paris : Presses Universitaires de France.
- Roberts C. 1989. « Modal Subordination and Pronominal Anaphora in Discourse ». *Linguistics and Philosophy* 12, p. 683-721.
- Roberts C. 2012 [1998]. « Information structure in discourse : Towards an integrated formal theory of pragmatics ». *Semantics and Pragmatics* 5 (6), p. 1-69.
- Rodriguez Somolinos A. 2011. « Présentation : les marqueurs de discours – Approches contrastives ». *Langages* 184, p. 3-12.
- Rodriguez Somolinos A. 2016. « Énonciation et marques d'oralité dans l'évolution du français. Présentation ». *LINX* 73, p. 9-16.
- Rossari C. 1994. *Les opérations de reformulation*. Berne : Peter Lang.
- Rossari C. 2000. *Connecteurs et relations de discours : des liens entre cognition et signification*. Nancy : Presses Universitaires de Nancy.
- Rossi M. 1999. *L'intonation, le système français : description et modélisation*. Paris : Ophrys.
- Roulet E. 1987a. « Complétude interactive et connecteurs reformulatifs ». *Cahiers de Linguistique Française* 8, p. 111-140.
- Roulet E. 1987b. *L'Articulation du discours en français contemporain*. Bern : Peter Lang.
- Roulet E., Filliettaz L., Grobet A. 2001. *Un modèle et un instrument d'analyse de l'organisation du discours*. Berne : Peter Lang.
- Roze C. 2009. *Base lexicale des connecteurs discursifs du français*. Mémoire de M2 recherche en lin-

- guistique informatique de l'Université Paris 7.
- Roze C. 2013. *Vers une algèbre des relations de discours*. Doctorat de linguistique, Université Paris 7.
- Roze C., Danlos L. & Muller P. 2012. « LEX-CONN : A French Lexicon of Discourse Connectives. *Discours* 10, 15 p. [En ligne](#)
- Russell B. 2007. « Imperatives in conditional conjunctions ». *Natural Language Semantics* 15, p. 131-166.
- Ryngaert J.-P. 1993. *Lire le théâtre contemporain*. Paris : Dunod.
- ## S
- Saint-Gérand J. 1993. *Morales du style*. Toulouse : Presses Universitaires du Mirail.
- Sag I. 2012. « Sign-Based Construction Grammar. An informal synopsis ». In Boas H.C., Sag I. *Sign-Based Construction Grammar*. Stanford: CSLI Publications. [En ligne](#)
- Sanders T., Demberg V., Hoek J., Scholman M. C.J., Asr. F. T., Zufferey S. & Evers-Vermeul J. 2021. « Unifying dimensions in coherence relations: How various annotation frameworks are related ». *Corpus Linguistics and Linguistics Theory* 17/1, p. 1-71. [En ligne](#)
- Sanders T., Sanders J. & Sweetser E. 2009. « Causality, cognition and communication: A mental space analysis of subjectivity in causal connectives ». In Sanders T., Sweetser E. (eds), *Causal Categories in Discourse and Cognition*. Berlin-New York : Mouton de Gruyter, p. 19-59.
- Sanders J., Sanders T. & Sweetser E. 2012. « Responsible subjects and discourse causality. How mental spaces and perspective help identifying subjectivity in Dutch backward causal connectives ». *Journal of Pragmatics* 44, p. 191-213.
- Sanders T., Spooren W. & Noordman L. 1992. « Toward a taxonomy of coherence relations ». *Discourse processes* 15(1), p. 1-35.
- Sankoff D., Sankoff G., Laberge S. & Topham M. 1976. « Méthodes d'échantillonnage et utilisation de l'ordinateur dans l'étude de la variation grammaticale ». *Cahiers de linguistique de l'Université du Québec* 6, p. 85-125.
- Sankoff D. & Laberge S. 1978. « The Linguistic Market and the Statistical Explanation of Variability ». In Sankoff D. (ed.), *Linguistic Variation, Models and Methods*. New York-San Francisco-London, Academic Press, p. 239-250.
- Sasamoto R. & Wilson D. (eds). 2016. « Little words : communication and procedural meaning ». *Lingua* 175-176, p. 1-166.
- Savelli M. 1993. *Contribution à l'analyse macro-syntaxique : les constructions « siamoises » de type « plus V1 plus V2 »*. Thèse de l'Université de Provence.
- Scholman M., Asr F. T., Zufferey S. & Evers-Vermeul J. 2018. « Unifying dimensions in coherence relations: How various annotation frameworks are related ». *Corpus Linguistics and Linguistic Theory*, 71 p. (en ligne open access).
- Schiffrin D. 1987. *Discourse Markers*. Cambridge : Cambridge University Press.
- Schoonjans S. 2013. « Modal particles : Problems in defining a category ». In Degand L., Cornillie B., Pietrandrea P. (eds), *Discourse Markers and Modal Particles. Categorization and description*. Amsterdam-Philadelphia : John Benjamins, p. 133-161.
- Schourup L. 1982. *Common discourse particles in English conversation*. New York : Garland.
- Schourup L. 1999. « Tutorial overview. Discourse markers ». *Lingua* 107, p. 227-265.
- Schubert C. & Sanchez-Stockhammer C. (eds). 2016. *Variational Text Linguistics. Revisiting Register in English*. Berlin-Boston : De Gruyter Mouton.
- Searle J. 1969. *Speech Acts*. Cambridge : Cambridge University Press. (voir aussi 1972 [1969], traduction en français. Paris : Hermann).
- Simons M., Tonhauser J., Beaver D. & Roberts C. 2011. « What projects and why ». In Li, N., Lutz, D. (eds), *Semantics and Linguistic Theory (SALT) 20*, eLanguage, p. 309-327.
- Sperber D. & Wilson D. 1995 [1986]. *Relevance, Communication and Cognition* (2nd ed.). Oxford-Cambridge : Blackwell.
- Stalnaker R. C. 1974. « Pragmatic presuppositions ». In Munitz, M., Unger, P. (eds), *Semantics and Philosophy*. New York : New York University Press, p. 197-214.
- Stede M., Afantenos S., Peldszus A., Asher N. & Perret J. 2016. « Parallel Discourse Annotations on a Corpus of Short Texts ». *Proceedings LREC'16*, Slovenia, ELRA, p. 1051-1058.
- Stede M., Mamprin S., Peldszus A., Herzog A., Kaupat D., Chiarcos C. & Warzecha S. 2016. *Textannotation: Potsdamer Kommentarkorpus 2.0*. Potsdam: Universitätsverlag Potsdam, Potsdam Cognitive Science Series 8.
- Stede M. & Schmitz B. 2000. « Discourse Particles and Discourse Functions ». *Machine Translation* 15, p. 125-147.
- Stede M. & Schneider J. 2019. *Argumentation Mining*. San Rafael (Ca) : Morgan & Claypool Publishers.
- Stede M., Taboada M. & Das D. 2017. *Annotation Guidelines for Rhetorical Structure*. Version 1.0. <https://www.sfu.ca>
- Stefanowitsch A. & Gries S. T. 2003. « Collocations: Investigating the interaction of

- words and constructions ». *International Journal of Corpus Linguistics* 8/2, p. 209-243.
- Stein D. & Traugott E. (eds) 1995. *Subjectivity and subjectivization. Linguistic perspectives. Cambridge* : Cambridge University Press.
- Stenius E. 1967. « Mood and language-game ». *Synthese* 17(1), p. 254-274.
- Swales J. M. 2013. *Research Genres. Exploration and Applications. Cambridge* : Cambridge University Press.
- Sweetser E. 1988. « Grammaticalization and Semantic Bleaching ». *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, p. 389-405. [En ligne](#)
- Sweetser E. 1990. *From etymology to pragmatics. Metaphorical and cultural aspects of semantic structure. Cambridge* : Cambridge University Press.
- ## T
- Taboada M. & Mann W. 2006. « Rhetorical Structure Theory: Looking Back and Moving Ahead ». *Discourse Studies* 8/3, p. 423-459.
- Tannen D., Hamilton H. E. & Schiffrin D. (eds) 2015. *The Handbook of Discourse Analysis*, 2 vol., 2<sup>nd</sup> ed.. Oxford: Wiley & Blackwell.
- Thibault P. & Vincent D. 1990. *Un corpus de français parlé. Montréal 84 : historique, méthodes et perspectives de recherche. Sainte-Foy* : Département de langues et linguistique, université Laval.
- Tindale C. W. 2009. « L'argumentation rhétorique et le problème de l'auditoire complexe ». *Argumentation et analyse du discours* 2, 15 p. [En ligne](#)
- Todorov T. 1984. *Mikhaïl Bakhtine. The Dialogic Principle. Manchester* : Manchester University Press. Version en français: *Mikhaïl Bakhtine. Le Principe dialogique. Suivi de Écrits du Cercle de Bakhtine. Paris* : Editions du Seuil, 1981.
- Tonhauser J., Beaver D., Roberts C. & Simons M. 2013. « Toward a taxonomy of projective content ». *Language* 89, p. 66-109.
- Tonhauser J., Beaver D.I. & Degen J. 2018. « How projective is projective content ? Gradient in projectivity and at-issueness ». *Journal of Semantics* 35, p. 495-542.
- Torck D. 1996. *Aspects de la causalité discursive en français oral contemporain. Thèse de L'université Libre d'Amsterdam. Studies in Language and Language Use* 23. Amsterdam : IFOTT.
- Toulmin S. E. 1958. *The Uses of Argument. Cambridge University Press.*
- Traugott E. 1982. « From propositionnal to textual and expressive meanings : somme semantic-pragmatic aspects of grammaticalization ». In *Perspectives on historical linguistics*, Lehmann, W. P., Malkiel, Y. (eds). Amsterdam-Philadelphia : John Benjamins, p. 245-272.
- Traugott E. 1995. « Subjectivization in grammaticalization ». In Stein D., Traugott E. (eds), *Subjectivity and subjectivization. Linguistic perspectives. Cambridge* : Cambridge University Press, p. 31-54.
- Traugott E. C. 2007. « Discussion article : Discourse markers, modal particles, and contrastive analysis, synchronic and diachronic ». *Catalan Journal of Linguistics* 2, p. 139-157.
- Traugott E. C. 2008. « Grammaticalization, constructions and the incremental development of language: Suggestions from the development of Degree Modifiers in English ». In Eckardt R., Jäger G., Veenstra T. (eds), *Variation, Selection, Development-Probing the Evolutionary Model of Language Change. Berlin-New-York* : Mouton de Gruyter, p. 219-250.
- Traugott E. C. 2010. « (Inter)subjectivity and (Inter)subjectification: A Reassessment ». In Davidse K., Vandelanotte L., Cuyckens H. (eds), *Subjectification, Intersubjectification and Grammaticalization. Berlin* : Mouton de Gruyter, p. 29-74.
- Traugott E. C. 2014. « Toward a constructional framework for research on language change. » In Hancil S., König E. (eds), *Grammaticalization – Theory and Data. Amsterdam-Philadelphia: John Benjamins*, p. 87-105.
- Traugott E. & Dasher R. 2002. *Regularity in semantic change. Cambridge* : Cambridge University Press.
- Trousdale G. 2012. « Grammaticalisation, constructions and the grammaticalization of constructions ». In Davidse K. et al. (eds), *Grammaticalization and Language Change: New reflections. Amsterdam-Philadelphia* : John Benjamins, p. 167-198.
- Trousdale G. 2015. « Multiple inheritance and constructional change ». In De Smet H., Ghesquière L., Van de Velde F. (eds), *On Multiple Source Constructions in Language Change. Amsterdam-Philadelphia* : John Benjamins, p. 19-42.
- Tsai Y.-T. 2020. *A Corpus-Based Comparative Analysis of Argumentation and Rhetorical Structure Theory. Mémoire de Master 2 TAL, IDMC, Nancy.*
- ## U
- Ubersfeld A. 1996. *Lire le théâtre III. Le dialogue de théâtre. Paris* : Belin.
- ## V
- Valette M. & Slodzian M. 2008. « Sémantique des textes et recherche d'information ». *Revue*

- française de linguistique appliquée* vol. XIII, p. 118-133.
- Van Baar T. 1996. « Particles ». In Devriendt B., Goossens L., van der Auwera J. (eds), *Complex structures: A functionalist perspective*. Berlin-New York : Mouton de Gruyter, p. 259–301.
- Van Dijk, T. A. 2008. *Discourse and context: a socio-cognitive approach*. Cambridge : Cambridge University Press.
- Van der Sandt R.A. 1992. « Presupposition projection as anaphora resolution ». *Journal of Semantics* 9, p. 333-377.
- Van Dijk T. A. 2014. *Discourse and Knowledge. A sociocognitive approach*. Cambridge : Cambridge University Press.
- Van Gompel R. P. G., Fischer M. H., Murray W. S. & Hill R. L. (eds). 2007. *Eye Movements: A Window on Mind and Brain*. Amsterdam : Elsevier.
- Vázquez Molina J. 2016. « Des connecteurs argumentatifs aux opérateurs discursifs. Présentation ». *Scolia* 30, p. 7-14.
- Verschueren J., Östman J.-O., Bloomaert J. & Bulcaen C. (eds). 1996. *Handbook of Pragmatics*. Amsterdam-Philadelphia: John Benjamins
- Villata S. 2018. *Explainable, Trustable and Emphatic Artificial Intelligence from Formal Argumentation Theory to Argumentation for Humans*. Computer science. Université Côte d’Azur.
- Vincent D. 1993. *Les ponctuations de la langue et autres mots du discours*. Québec : Nuit Blanche Éditeur.
- Vincent D. & Sankoff D. 1992. « Punctors : a pragmatic variable ». *Language Variation and Change* 4, p. 205-216.
- Walton D. N. 2013. *Methods of Argumentation*. Cambridge : Cambridge University Press.
- Walton D. N. & Krabbe E. C. W. 1995. *Commitment in Dialogue*. Albany : State University of New York Press.
- Walton D. N., Reed C. & Macagno F. 2008. *Argumentation Schemes*. Cambridge : Cambridge University Press.
- Webber B., Knott A. & Joshi A. 2001. « Multiple Discourse Connectives in a Lexicalized Grammar for Discourse ». In Bunt H., Muskens R., Thijsse E., *Computing Meaning* vol. 2, Kluwer, p. 229.
- Werning M., Hinzen W. & Machery E. (eds). 2012. *The Oxford Handbook of Compositionality*. Oxford : Oxford University Press.
- Wharton T. 2003. « Interjections, language and the showing-saying continuum ». *Pragmatics and Cognition* 11 (1), p. 39-91.
- Wharton T. 2016. « That bloody *so-and-so* has retired : Expressives revisited ». *Lingua* 175-176, p. 20-35.
- Williams G. (dir.) 2005. *La Linguistique de corpus*. Rennes : Presses Universitaires de Rennes.
- Wilson D. 2017. « Relevance Theory ». In Huang Y. (ed.), *The Oxford Handbook of Pragmatics*. Oxford : Oxford University Press, p. 79-99.
- Wilson D. & Carston R. 2019. « Pragmatics and the challenge of ‘non-propositional’ effects ». *Journal of Pragmatics* 145, p. 31-38.
- Wilson D. & Sperber D. 1993. « Linguistic form and relevance ». *Lingua* 90, p. 1-25. [repris dans Wilson & Sperber 2012, p. 149-168].
- Wilson D. & Sperber D. 2012. *Meaning and Relevance*. Cambridge : Cambridge University Press.
- Winograd T. 1975. « Frame representations and the declarative/procedural controversy ». In Bobrow D., Collins A. (eds), *Representation and understanding : Studies in cognitive science*. New York: Academic Press, p. 185-210.
- Winterstein G. 2010. *La dimension probabiliste des marqueurs de discours : nouvelles perspectives sur l’argumentation dans la langue*. Thèse de l’Université Paris 7, 231 p.
- Winterstein G. 2012. « What *but*-sentences argue for: An argumentative analysis of *but* ». *Lingua* 122 (15), p. 1864-1885.
- Wachsmuth H., Da San Martino G., Kiesel D., & Stein B. 2017. « The impact of modeling overall argumentation with tree kernels ». In *Proceedings of the 2017 EMNLP Conference, ACL, Copenhagen*, p. 2379-2389.
- Waltereit R. 2001. « Modal particles and their functional equivalents : A speech-act-theoretic approach ». *Journal of Pragmatics* 33, p. 1391-1417.
- Waltereit R. 2007. « A propos de la genèse diachronique des combinaisons de marqueurs. L’exemple de *bon ben* et *enfin bref* ». *Langue française* 154, p. 94-109.
- Waltereit R. 2012. « Grammaticalization and discourse ». In Heine B., Narrog H. (eds), *The Oxford Handbook of Grammaticalization*. Oxford : Oxford Handbooks Online, 8 p.
- Walton D. N. 1992. *The place of emotions in argument*. University Park : The Pennsylvania State University Press.
- Yan X. & Han J. 2002. « gSpan : graph-based substructure pattern mining ». In *Proceedings of 2002 IEEE International Conference on Data Mining* (Maebashi City, Japan), p.721–724.
- Yang A. & Li S. 2018. « SciDTD : Discourse Dependency TreeBank for Scientific Abs-

## W

## Y

- tracts ». In *Proceedings of the 56<sup>th</sup> Annual Meeting of the ACL*, vol. 2, p. 444-449.
- Yoon J. & Gries S. 2016. *Corpus-based approaches to Construction Grammar*. Amsterdam: John Benjamins.
- Yule G. (ed.) 2017. *The Study of Language* 6<sup>th</sup> edition. Cambridge : Cambridge University Press.
- Zimmermann M. 2011. « Discourse particles ». In Maienborn C., von Stechow, K., Portner P. (eds), *Semantics : An International Handbook of Natural Language Meaning*, vol. 2. Berlin : de Gruyter, p. 2011-2038.
- Zufferey S. 2014. « Givenness, procedural meaning and connectives : The case of French *puisque* ». *Journal of Pragmatics* 62/1, p. 121-135.

## Z

- Zienkowski J., Östman J.-O. & Verschueren J. (eds) 2011. *Discursive Pragmatics*. Amsterdam-Philadelphia: John Benjamins.



# 6

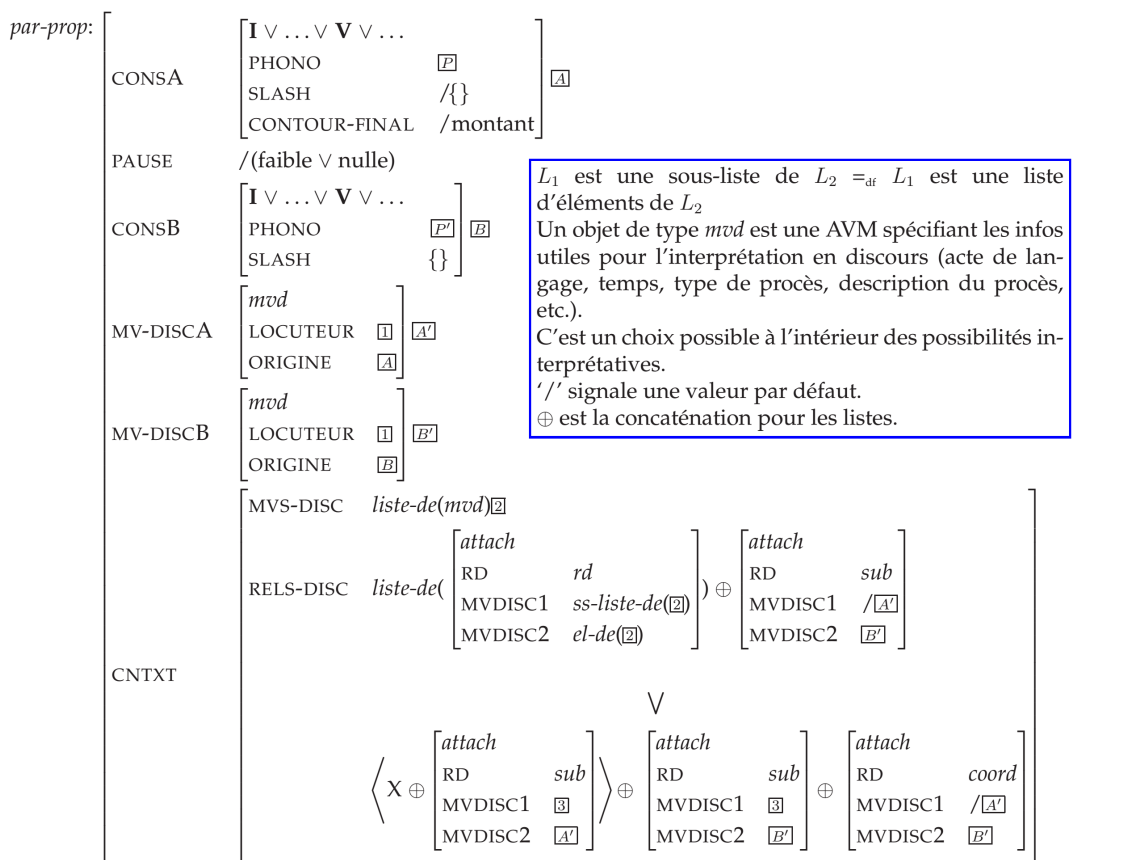
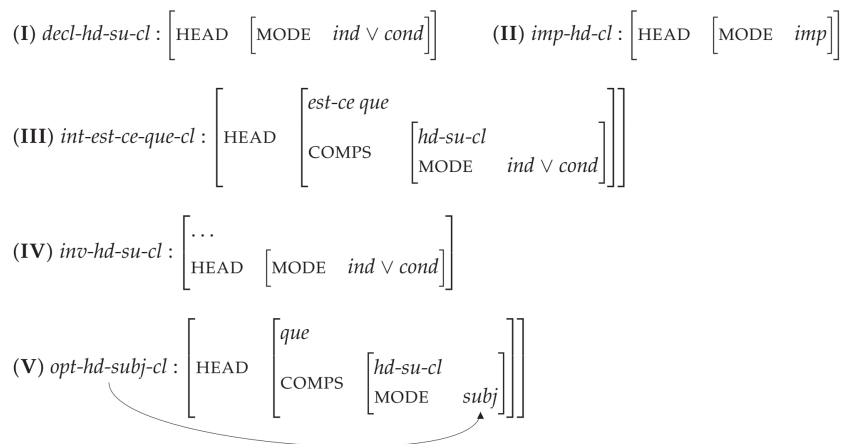
## Annexes





## Annexe 1 : Structures de traits des constructions juxtaposées

### Annexe 1a : construction *juxta-prop* (juxtaposition propositionnelle)



**Annexe 1b** : construction *juxta-cond* (juxtaposition propositionnelle à interprétation conditionnelle)

<i>juxta-cond</i> :	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;">CONSA</td> <td style="padding: 5px;"><math>[decl-hd-su-cl \vee imp-hd-cl \vee inv-hd-su-cl \vee opt-hd-subj-cl</math></td> <td style="padding: 5px;"><math>/{\}}]</math></td> <td style="padding: 5px;"><math>\boxed{A}</math></td> </tr> <tr> <td style="padding: 5px;">PAUSE</td> <td colspan="3" style="padding: 5px;"><math>/(faible \vee nulle)</math></td> </tr> <tr> <td style="padding: 5px;">CONSB</td> <td style="padding: 5px;"><math>[decl-hd-su-cl \vee int-hd-su-cl</math></td> <td style="padding: 5px;"><math>/{\}}]</math></td> <td style="padding: 5px;"><math>\boxed{B}</math></td> </tr> <tr> <td style="padding: 5px;">MV-DISCA</td> <td style="padding: 5px;"><math>[mvd</math></td> <td style="padding: 5px;"><math>LOCUTEUR \boxed{1} \boxed{A'}</math></td> <td rowspan="2" style="padding: 5px;"><math>X \Leftrightarrow Y</math> signale qu'on modifie la structure <math>X</math> en se conformant aux déclarations attribut-valeur de <math>Y</math>.</td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;"><math>ORIGINE \boxed{A}]</math></td> <td></td> </tr> <tr> <td style="padding: 5px;">MV-DISCB</td> <td style="padding: 5px;"><math>[mvd</math></td> <td style="padding: 5px;"><math>LOCUTEUR \boxed{1} \boxed{B'}</math></td> <td rowspan="2" style="padding: 5px;"><math>X \Leftrightarrow Y</math> signale qu'on modifie la structure <math>X</math> en se conformant aux déclarations attribut-valeur de <math>Y</math>.</td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;"><math>ORIGINE \boxed{B}]</math></td> <td></td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;"><math>MVS-DISC</math></td> <td style="padding: 5px;"><math>liste-de(mv-disc)\boxed{2}</math></td> <td></td> </tr> <tr> <td style="padding: 5px;">CNTXT</td> <td style="padding: 5px;"><math>RELS-DISC</math></td> <td style="padding: 5px;"><math>liste-de(</math></td> <td style="padding: 5px;"><math>\oplus</math></td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;"></td> <td style="padding: 5px;"><math>\begin{matrix} attach \\ RD \\ MVDISC1 \\ MVDISC2 \end{matrix}</math></td> <td style="padding: 5px;"><math>\begin{matrix} attach \\ RD \\ MVDISC1 \\ MVDISC2 \end{matrix}</math></td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;"></td> <td style="padding: 5px;"><math>\begin{matrix} rd \\ ss-liste-de(\boxed{2}) \\ el-de(\boxed{2}) \end{matrix}</math></td> <td style="padding: 5px;"><math>\begin{matrix} condition \\ \boxed{A'} \Leftrightarrow [FORCE-ILLOC \ ind\acute{e}t] \\ \boxed{B'} \end{matrix}</math></td> </tr> </table>	CONSA	$[decl-hd-su-cl \vee imp-hd-cl \vee inv-hd-su-cl \vee opt-hd-subj-cl$	$/{\}}]$	$\boxed{A}$	PAUSE	$/(faible \vee nulle)$			CONSB	$[decl-hd-su-cl \vee int-hd-su-cl$	$/{\}}]$	$\boxed{B}$	MV-DISCA	$[mvd$	$LOCUTEUR \boxed{1} \boxed{A'}$	$X \Leftrightarrow Y$ signale qu'on modifie la structure $X$ en se conformant aux déclarations attribut-valeur de $Y$ .		$ORIGINE \boxed{A}]$		MV-DISCB	$[mvd$	$LOCUTEUR \boxed{1} \boxed{B'}$	$X \Leftrightarrow Y$ signale qu'on modifie la structure $X$ en se conformant aux déclarations attribut-valeur de $Y$ .		$ORIGINE \boxed{B}]$			$MVS-DISC$	$liste-de(mv-disc)\boxed{2}$		CNTXT	$RELS-DISC$	$liste-de($	$\oplus$			$\begin{matrix} attach \\ RD \\ MVDISC1 \\ MVDISC2 \end{matrix}$	$\begin{matrix} attach \\ RD \\ MVDISC1 \\ MVDISC2 \end{matrix}$			$\begin{matrix} rd \\ ss-liste-de(\boxed{2}) \\ el-de(\boxed{2}) \end{matrix}$	$\begin{matrix} condition \\ \boxed{A'} \Leftrightarrow [FORCE-ILLOC \ ind\acute{e}t] \\ \boxed{B'} \end{matrix}$
CONSA	$[decl-hd-su-cl \vee imp-hd-cl \vee inv-hd-su-cl \vee opt-hd-subj-cl$	$/{\}}]$	$\boxed{A}$																																								
PAUSE	$/(faible \vee nulle)$																																										
CONSB	$[decl-hd-su-cl \vee int-hd-su-cl$	$/{\}}]$	$\boxed{B}$																																								
MV-DISCA	$[mvd$	$LOCUTEUR \boxed{1} \boxed{A'}$	$X \Leftrightarrow Y$ signale qu'on modifie la structure $X$ en se conformant aux déclarations attribut-valeur de $Y$ .																																								
	$ORIGINE \boxed{A}]$																																										
MV-DISCB	$[mvd$	$LOCUTEUR \boxed{1} \boxed{B'}$	$X \Leftrightarrow Y$ signale qu'on modifie la structure $X$ en se conformant aux déclarations attribut-valeur de $Y$ .																																								
	$ORIGINE \boxed{B}]$																																										
	$MVS-DISC$	$liste-de(mv-disc)\boxed{2}$																																									
CNTXT	$RELS-DISC$	$liste-de($	$\oplus$																																								
		$\begin{matrix} attach \\ RD \\ MVDISC1 \\ MVDISC2 \end{matrix}$	$\begin{matrix} attach \\ RD \\ MVDISC1 \\ MVDISC2 \end{matrix}$																																								
		$\begin{matrix} rd \\ ss-liste-de(\boxed{2}) \\ el-de(\boxed{2}) \end{matrix}$	$\begin{matrix} condition \\ \boxed{A'} \Leftrightarrow [FORCE-ILLOC \ ind\acute{e}t] \\ \boxed{B'} \end{matrix}$																																								

**Exemples de contraintes**

- (C1)  $\boxed{A} [HEAD [MODE cond]] \Rightarrow \boxed{B} [decl-hd-su-cl [HEAD [MODE cond]]]$
- (C2)  $\boxed{A} [HEAD [MODE ind [TEMPS \neg imparfait]]] \Rightarrow \boxed{B} [decl-hd-su-cl [HEAD [MODE ind]]]$
- (C3)  $\boxed{A} [HEAD [TEMPS \neg plus-que-parfait]]]$
- (C4)  $\boxed{A} [HEAD [TEMPS imparfait]] \Rightarrow \boxed{A} [HEAD [TEMPS imparfait] SEM [VAL-MODALE irr\acute{e}el]]]$
- (C5)  $\boxed{A} [HEAD [TEMPS imparfait]] \Rightarrow \boxed{B} [HEAD [TEMPS imparfait \vee cond-pass\acute{e}] SEM [VAL-MODALE irr\acute{e}el]]]$

## Annexe 2 : Contrastes des valeurs de pente pour l'analyse des contours prosodiques continuatifs

(extrait de Dagnat et al. 2012a [16], figures 3 et 4, p. 1478)

### Légende

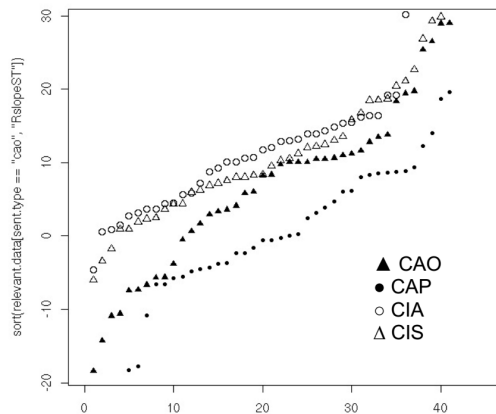
\* : la différence observée est significative. Seuils : \*\*\*<1%<\*\*<1%<\*<5%

*RslopeST* : pente de la droite de régression calculée à partir des hauteurs en semi-tons sur le segment-cible

*Ordonnées* : valeurs de *RslopeST* ; *abscisses* : rangement des observations par ordre croissant (chaque point correspond à une observation)

### Modalités de lecture des tableaux

Quand le résultat de la colonne de droite ( $\Pr(>|z|)$ ) est suivi d'un ou plusieurs astérisques, cela veut dire que le résultat du contraste testé est inférieur à un certain seuil de significativité (indiqué dans la légende, ci-dessus). Plus  $\Pr$  est faible, plus l'hypothèse qu'il n'y ait pas de différence (par ex.  $\text{cap-cao}=0$ ) entre les deux configurations contrastées a des chances de ne pas être vérifiée. Donc, plus  $\Pr$  est faible, plus les configurations contrastées peuvent être considérées comme différentes (du point de vue de la pente de régression en semi-tons calculée sur les deux dernières syllabes), autrement dit plus la différence observée peut être considérée comme n'étant pas due au hasard.

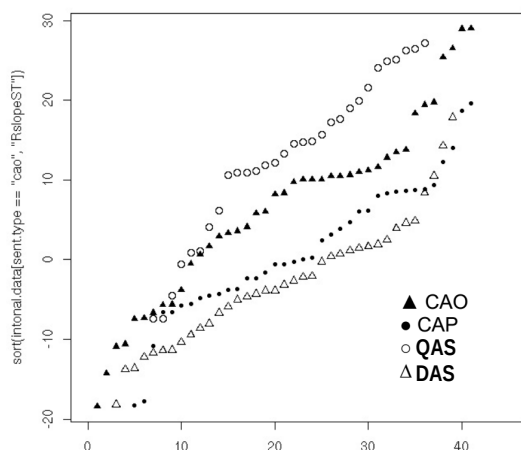


### Linear Hypotheses

	Estim.	Std. Er.	z value	Pr(> z )
CAP-CAO= 0	-7.9892	2.1006	-3.803	0.000769 ***
CIA-CAO= 0	4.8406	2.6946	1.796	0.269306
CIS-CAO= 0	3.9481	2.6541	1.488	0.438261
CIA-CAP= 0	12.8298	2.6974	4.756	<1e-04 ***
CIS-CAP= 0	11.9373	2.6570	4.493	<1e-04 ***
CIS-CIA= 0	-0.8925	2.1703	-0.411	0.975893

1e-04 = 1/10 000

Annexe 2a. Contraste des pentes dans les configurations CAO-CAP-CIS-CIA



### Linear Hypotheses

	Estim.	Std. Er.	z value	Pr(> z )
CAP-CAO= 0	-7.8688	2.9639	-2.655	0.0399 *
QAS-CAO= 0	-8.3980	2.9639	-2.833	0.0240 *
DAS-CAO= 0	0.5316	3.0429	0.175	0.9981
QAS-CAP= 0	-0.5293	2.9639	-0.179	0.9980
DAS-CAP= 0	8.4004	3.0429	2.761	0.0296 *
QAS-DAS= 0	8.9296	3.0429	2.935	0.0170 *

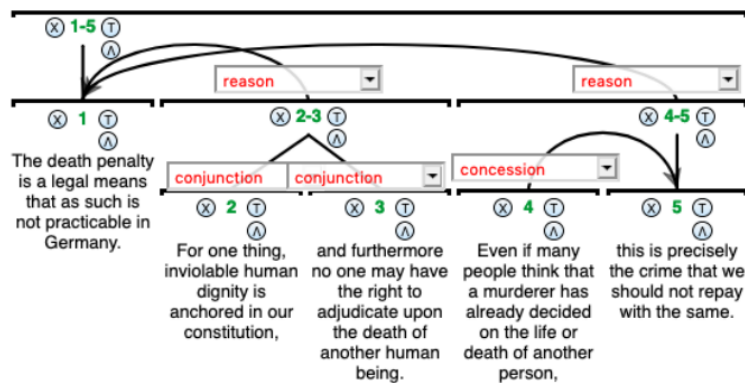
Annexe 2b. Contraste des pentes dans les configurations CAO-CAP-QAS-DAS

### Annexe 3 : Graphes initiaux, RST et argumentation

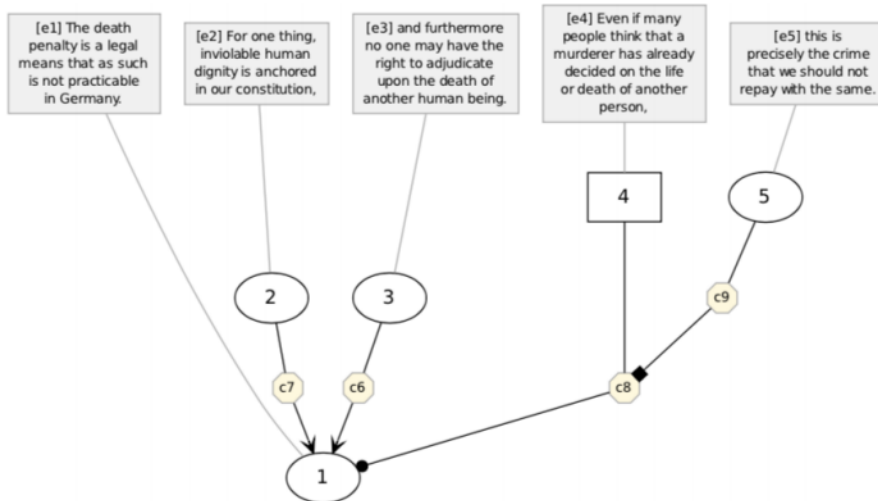
Exemple (53)

*Topic* : Should Germany introduce the death penalty?

The death penalty is a legal means that as such is not practicable in Germany [1]. For one thing, inviolable human dignity is anchored in our constitution [2], and furthermore no one may have the right to adjudicate upon the death of another human being [3]. Even if many people think that a murderer has already decided on the life or death of another person [4], this is precisely the crime that we should not repay with the same [5]. (arg-microtexts, micro\_b\_006)



3a. Graphe initial de la RST



3b. Graphe initial de la structure argumentative

## **Annexe 4 : Liste des propriétés des *pragmatic markers* de Brinton**

<b>Phonological and lexical characteristics</b>	
a	Pragmatic markers are often « small » items, although they may also be phrasal or clausal; they are sometimes phonologically reduced.
b	Pragmatic markers may form a separate tone group, but they may also form a prosodic unit with preceding or following material.
c	Pragmatic markers do not constitute a traditional word class, but are most closely aligned to adverbs, conjunctions, or interjections.
<b>Syntactic characteristics</b>	
d	Pragmatic markers occur either outside the syntactic structure or loosely attached to it.
e	Pragmatic markers occur preferentially at clause boundaries (initial/final) but are generally movable and may occur in sentence-medial position as well.
f	Pragmatic markers are grammatically optional but at the same time serve important pragmatic functions (and are, in a sense, pragmatically non-optional).
<b>Semantic characteristics</b>	
g	Pragmatic markers have little or no propositional/conceptual meaning, but are procedural and non-compositional.
<b>Functional characteristics</b>	
h	Pragmatic markers are often multifunctional, having a range of pragmatic functions.
<b>Sociolinguistic and stylistic characteristics</b>	
i	Pragmatic markers are predominantly a feature of oral rather than written discourse; spoken and written pragmatic markers may differ in form and function.
j	Pragmatic markers are frequent and salient in oral discourse.
k	Pragmatic markers are stylistically stigmatized and negatively evaluated, especially in written or formal discourse.
l	Pragmatic markers may be used in different ways and in different frequencies by men and women.

Brinton (1996, p. 33-35 et 2007, p. 3-7)

## NOTES

## NOTES



## NOTES