



HAL
open science

La subjectivité dans le discours. Entre linguistique et traitement automatique des langues

Agata Jackiewicz

► **To cite this version:**

Agata Jackiewicz. La subjectivité dans le discours. Entre linguistique et traitement automatique des langues. Linguistique. Université Paris-Sorbonne, école doctorale 5 “ Concepts et langages ”, 2011. tel-01839499

HAL Id: tel-01839499

<https://hal.science/tel-01839499>

Submitted on 15 Jul 2018

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

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



Ecole doctorale V « Concepts et Langages »
Equipe « Sens Texte Informatique et Histoire » (STIH, EA 4509)

LA SUBJECTIVITÉ DANS LE DISCOURS

Entre linguistique et
traitement automatique des langues

Agata Jackiewicz

Dossier en vue de
l'Habilitation à Diriger des Recherches

Volume 1 : Synthèse

Jury :

Jacques Bres, Pr, Université Paul-Valéry (Montpellier 3), examinateur
Michel Charolles, Pr, Université Sorbonne Nouvelle (Paris 3), examinateur
Béatrice Daille, Pr, Université de Nantes, rapporteur
Patrick Dendale, MC HDR, Université d'Anvers, examinateur
Patrice Enjalbert, Pr, Université de Caen, examinateur
Béatrice Lamiroy, Pr, Université Catholique de Louvain (KUL), rapporteur
Franck Neveu, Pr, Université de Paris-Sorbonne, rapporteur, président
Olivier Soutet, Pr, Université de Paris-Sorbonne, directeur

Paris, le 26 novembre 2011

Remerciements

Mes premiers et plus sincères remerciements vont à Olivier Soutet, directeur de l'équipe STIH, qui a accepté de diriger ce projet d'HDR. Sa confiance et son soutien ont été aussi précieux que déterminants.

J'exprime mon plus grand respect et mes profonds remerciements aux rapporteurs qui m'ont fait l'honneur d'évaluer ce travail de recherche et de participer à sa soutenance : à Franck Neveu, en qualité du rapporteur interne, ainsi qu'à Béatrice Daille et à Béatrice Lamiroy, en tant qu'experts extérieurs.

Je suis reconnaissante à Patrice Enjalbert, Michel Charolles, Patrick Dendale et à Jacques Bres qui ont chaleureusement accepté de participer à ce jury. Je dois à Patrice Enjalbert des conseils avisés sur les premières versions de mes manuscrits.

J'exprime ma gratitude à Jean-Pierre Desclés pour son accueil au laboratoire LaLIC et ces longues années de collaboration.

Mes différents travaux se sont nourris de collaborations et d'échanges d'idées notamment avec Brigitte Grau, Jean-Luc Minel, Antoine Blais, Michel Charolles, Denis Vigier, Stéphane Ferrari, Thierry Charnois, Dominique Legallois, Frédéric Bilhaut, Max Silberztein... Qu'ils en soient tous remerciés.

Je dois beaucoup à l'ensemble des partenaires du projet ANR OntOpITex qui ont apporté un nouveau souffle à mon travail de recherche. Je les remercie très sincèrement pour cette collaboration à la fois passionnante, fructueuse et agréable.

Merci à l'ensemble des collègues et du personnel de l'ISHA pour leur présence chaleureuse et leur gentillesse.

J'ai une pensée pour mes étudiants, source d'énergie, d'enthousiasme et de stimulation. Je m'estime chanceuse de pouvoir croiser leurs chemins.

Enfin, j'adresse un mot affectueux à mes proches, ceux qui partagent mes états d'âme et illuminent ma vie: mes parents, mon mari Luc, mes fils Boris et Tom, mes amis : Dominique D., Anita, Véronique, Dominique S., Renn, Laure, Annick, Pascale, Karine, Marzena, Leila, Franco, Laurent, Jean-Paul, Lucine, Emmanuel... et tant d'autres. Votre présence et votre soutien me sont si chers !

SOMMAIRE

SOMMAIRE	5
INTRODUCTION	7
1. Présentation du mémoire	7
2. Formation et parcours de recherche	8
3. Enseignement	10
4. L'énonciation : l'axe de l'intelligibilité des discours	11
QUESTIONS DE MÉTHODOLOGIE	13
1. Regard interdisciplinaire et démarche onomasiologique	13
2. Ce qu'ils sont : leur nature, leurs fonctions et leurs marques	14
3. Entre l'empirique et le théorique	15
4. Corpus, normes et usages	16
5. Des notions aux marques, des marques aux ressources	18
6. Pour une capitalisation des données linguistiques	19
7. Vers une « ontologie » de l'énonciation	20
REGARDS SUR LA SUBJECTIVITÉ, AU TRAVERS DES...	23
I. CONNAISSANCES CAUSALES	25
A. BREF ÉTAT DES LIEUX	25
B. LA CAUSALITÉ PRISE EN CHARGE	26
1. Problématique	26
2. Une lecture en contexte	27
3. Prendre en charge une connaissance causale	29
4. Double dimension de la causalité « langagière »	35
5. Conclusion	36
II. ORGANISATIONS TEXTUELLES	37
A. ÉTAT DES LIEUX	37
B. LES SÉRIES LINÉAIRES DANS LE DISCOURS	40
1. Introduction	40
2. Les séries linéaires : observation en corpus	41
3. Les séries linéaires à la lumière des analyses linguistiques	42
4. Nature cognitive et logique de l'objet « série »	44
5. Opérations, structures et configurations de marques	46
6. Conclusion	47
III. DISCOURS RAPPORTÉS	49
A. ÉTAT DES LIEUX	49
B. RESPONSABILITÉ ET PRISE EN CHARGE DANS LE DISCOURS RAPPORTÉ	52
1. Problématique	52
2. Choix préliminaires	52
3. De la prise en charge à la responsabilité	54
4. Responsabilité énonciative dans différents modes d'intervention de l'énonciateur rapportant	56
5. Typologie des citations	59

6. Pour conclure	60
IV. JUGEMENTS D'ÉVALUATION	61
A. BREF ÉTAT DES LIEUX	61
B. CONSTITUANTS DÉTACHÉS ET JUGEMENTS D'ÉVALUATION	63
1. Expressions qui ciblent des personnes	63
2. Des expressions aux patrons	64
3. Explorations outillées	66
4. Caractérisation ontologique, branchée sur le social et le culturel	69
5. Conclusion	70
BILAN ET PERSPECTIVES	71
BIBLIOGRAPHIE	73
ANNEXES	81

INTRODUCTION

1. Présentation du mémoire

Le présent mémoire a pour principal objet de présenter de manière synthétique mon parcours de recherche et d'enseignement mené en linguistique appliquée au traitement automatique des langues (TAL). Il prend également acte d'un certain nombre d'incursions dans les champs voisins du TAL que sont la recherche d'information (RI), le Web sémantique (WS), l'ingénierie des connaissances (IC).

Mon cheminement a abouti, l'évolution propre des disciplines concernées aidant, à la thématisation d'un objet, encore relativement peu étudié en linguistique et longtemps resté implicite dans les travaux appliqués du TAL, il s'agit de l'énonciation textuelle. Dans cet espace, je dégage une perspective de recherche autant que les grandes lignes d'une démarche originale. Il s'agit d'une élaboration en cours de constitution, aux influences et aux héritages multiples, dont je chercherai à montrer la cohérence, les retombées méthodologiques, ainsi que des prolongements envisagés.

Comme la plupart des recherches à caractère appliqué, mes travaux ne se situent pas à l'intérieur d'un cadre théorique unique, mais empruntent à divers champs de savoir et croisent plusieurs méthodes. Je me réclame à la fois de la linguistique de l'énonciation, de la linguistique textuelle et de la linguistique informatique. Je pratique une linguistique « outillée », sur corpus, qui propose des contributions linguistiques à des problématiques et à des applications informatiques qui relèvent du traitement automatique des langues.

Reconstituer mon itinéraire de recherche sur une quinzaine d'années, jalonné d'expériences variées, au sein des laboratoires CAMS (Centre d'Analyse et de Mathématiques Sociales) et LaLIC (Langage, Logique, Informatique et Cognition) et souvent en collaboration avec des chercheurs provenant d'autres laboratoires (LATTICE, GREYC, CRISCO, PRAXILING...), m'est apparu comme un exercice stimulant mais parfois difficile. Cette forme de synthèse oblige à revenir aux motivations premières, parfois enfouies, souvent résultant des opportunités de projets ou de rencontres, et à reconstruire le fil conducteur de ces années de recherche.

Le travail de synthèse de l'HDR inclut également une auto-évaluation du parcours accompli. Et de ce fait permet de revenir sur ses difficultés et certaines de ses lacunes, auxquelles j'ai souhaité apporter une réponse. Il est important de considérer le travail accompli à la lumière des conditions scientifiques, matérielles et humaines dans lesquelles il s'est déroulé. Le spectre large des problématiques abordées va de concert avec des parcours sélectifs, où la problématisation n'est jamais exhaustive et où la contribution proposée doit être située dans un contexte scientifique et temporel précis.

Je tenterai également de réinterroger ces différentes expériences à la lumière de mes travaux récents. A plusieurs titres, l'approche linguistique des jugements d'évaluation soulève des questions proches à celles que j'ai rencontrées en travaillant sur la causalité. Dans les deux cas, le caractère interprétatif des contenus analysés est avéré.

La formule choisie ici est une synthèse sélective de mes travaux, focalisée essentiellement sur les contributions à l'étude de la subjectivité dans le discours. J'ai fait le choix de présenter en les détaillant les résultats les plus significatifs. Ce document permet en effet de réunir des études n'ayant jamais été mises en perspective ensemble. A chaque étape, je mentionne brièvement les travaux qui ne font pas l'objet d'un développement.

Après une brève présentation de mon parcours universitaire, j'exposerai ce qui constitue le fil conducteur de mon travail de recherche et les principaux choix théoriques effectués. Les réflexions méthodologiques précèdent et « indexent » quatre sections qui présentent quatre regards particuliers sur la texture énonciative des discours. L'ensemble s'achève par un bref bilan et l'exposé des directions que je souhaite prendre dans les années à venir. La section des Annexes inclut mon curriculum vitae détaillé.

2. Formation et parcours de recherche

Arrivée en France en juin 1986, après avoir validé ma première année de Statistiques et d'économétrie à l'Université de Gdansk (Pologne), j'ai préparé un DUT d'informatique à l'IUT de Rodez (en Aveyron). J'ai opté ensuite pour les Sciences du Langage (en licence et en maîtrise) alliées à un Diplôme Universitaire des Sciences Cognitives pour l'ingénierie de la Connaissance, à l'Université de Toulouse-Le Mirail. C'est Andrée Borillo qui m'a accueillie dans ces deux formations et initiée à une linguistique tournée vers le formel et le TAL. J'ai réalisé sous sa direction une première étude linguistique sur les adverbiaux temporels et programmé un outil informatique permettant de systématiser le codage de ces adverbiaux. Cette formation a fortement orienté mon parcours ultérieur et ma manière de travailler. Premièrement, j'y ai découvert l'intérêt des approches pluridisciplinaires (alliant linguistique, psychologie cognitive, informatique, neuropsycholinguistique). Deuxièmement, j'ai bénéficié d'un enseignement de l'informatique conçu pour le public de sensibilité littéraire. Cette expérience a influencé ma propre manière de travailler avec les étudiants littéraires, des années plus tard, notamment dans le cadre de la licence *Langue française et Techniques informatiques* (à Paris 4). Enfin, j'ai acquis des bases méthodologiques permettant un travail rigoureux et reproductible sur des observables linguistiques (collecte des observables, codage et organisation des données).

Quittant Toulouse pour Paris, en 1990, j'ai intégré le DEA de Sciences Cognitives, piloté conjointement par l'EHESS, Paris VI et l'Ecole Polytechnique. Cette formation à dominante Intelligence Artificielle m'a apporté également une ouverture en philosophie, en logique et en sociologie. En partant des travaux sur la méréologie de Stanislas Lesniewski¹, mon mémoire (dirigé par Jean-Pierre Desclés, au CAMS : UMR 17) a porté sur la modélisation linguistique et la représentation logique et informatique de la relation partie tout. Il a été mené au sein la société Décision International (DI), dans un contexte appliqué, où il a apporté une brique dans la constitution d'une méthode d'acquisition des connaissances METODAC. La fin des années 80 et le début des années 90 ont été marqués par un vif intérêt pour des systèmes à base de connaissances et des systèmes experts. La problématique de l'extraction et la modélisation des connaissances à partir des textes a conduit chercheurs et entreprises à proposer des méthodes (intégrées souvent à des ateliers) de génie cognitif et logiciel. Le travail sur la modélisation linguistique de la relation partie-tout et sur les ressources associées a été exploité dans le système SEEK (LaLIC) et ses évolutions (SEEK-Java, LaLIC), dans le projet E-VEIL (Xistos et Paris 4), ainsi que dans le système CAMELEON (IRIT).

J'ai effectué un détour par le monde de l'édition (société EC2, Edition Colloques Conseil), où j'ai travaillé sur la terminologie de l'intelligence artificielle. Ce travail a été publié par EC2 sous forme d'un *Vocabulaire d'Intelligence Artificielle* en 1991. J'ai également fait des vacances au CNRS sur des questions de résumé automatique. Puis, c'est à l'Université Paris-Sorbonne (au laboratoire LaLIC) et en convention CIFRE avec la société EDIAT (devenue ensuite CR2A-DI) que j'ai entrepris en 1994² une thèse, soutenue en janvier 1998, sur l'expression de la causalité dans les textes et son filtrage automatique. Ce travail a été dirigé par Jean-Pierre Desclés. En septembre de la même année, j'ai été nommé maître de conférences, à l'UFR ISHA de l'Université Paris 4, en restant au laboratoire LaLIC.

J'y ai travaillé durant plusieurs années sur la caractérisation et l'identification automatique dans des textes de certaines notions ou relations (causalité, relation partie-tout, citation, sériation...) par la méthode d'exploration contextuelle. Pendant cette période, j'ai collaboré étroitement avec plusieurs membres du groupe « Exploration contextuelle » (Desclés 1997), le mode de travail étant collectif et centré sur des projets appliqués (résumé automatique, filtrage de textes selon des points de vue...). J'ai noué au sein du LaLIC plusieurs collaborations enrichissantes : avec Leïla Nait Baha et Philippe Laublet (sur le projet RAP), avec Jean-Luc Minel et Gustavo Crispino (sur la navigation textuelle), avec Antoine Blais (sur le résumé), etc. Ces études ont donné lieu à la création de ressources linguistiques de grand volume (relation partie-tout, causalité, sériation, citation) organisées dans des bases de données. Ces ressources ont été exploitées dans les différents systèmes informatiques (SERAPHIN, SAFIR, ContextO, RAP, E-Veil et EXCOM) développés au laboratoire LaLIC. Mes différents travaux ont tous trouvé des prolongements dans des thèses et

¹ Exposée dans *Fondements de la théorie générale des ensembles* (1916). Voir aussi (Miéville 1984).

² La convention Cifre exigeant la nationalité française, le démarrage de la thèse a été conditionné par la date de la naturalisation.

divers travaux de recherche réalisées au LALIC, mais aussi dans d'autres laboratoires. A cette époque, j'ai également participé à des collaborations avec des équipes à l'étranger (en Pologne, en Tunisie, en Bulgarie, en Belgique et en Corée du Sud). J'ai effectué plusieurs séjours, alliant enseignements en école d'été et séances de travail. J'y ai pris part à des travaux contrastifs sur plusieurs langues ou faisant appel à des cadres théoriques et des méthodes différents.

A partir de 2002, j'ai fait le choix de m'orienter vers la linguistique textuelle, afin de travailler sur les problèmes de cohérence et de cohésion des discours ; questions que j'ai rencontrés à de multiples reprises lors des travaux appliqués (notamment sur le résumé automatique et sur la fouille par points de vue). J'ai entamé des études sur des unités et des organisations discursives de grain supérieur à la phrase, qui ont porté successivement sur : (i) les séries linéaires, (ii) les discours rapportés et (iii) les opinions. Le projet REGAL dirigé par le LIMSI (groupe LIR, Brigitte Grau) m'a permis de participer à une recherche collective qui a associé des méthodes numériques avec des méthodes symboliques (cadres de discours) pour construire un système destiné à une segmentation thématique de textes. Au sein du laboratoire LATTICE, j'ai collaboré avec Michel Charolles sur la problématique les cadres de discours, ainsi qu'avec Denis Vigier sur les adverbiaux cadratifs de nature énonciative. En participant au projet ANR SFA « Spatial Framing Adverbials : linguistic and psycholinguistic approaches » (Lattice, ENS et Paris 3), je me suis familiarisée avec la problématique de la grammaticalisation et du détachement.

Dans ce contexte est née la revue électronique *Discours*, dont j'ai assurée la direction durant ses deux premières années d'existence. La revue *Discours* publie des études sur toutes les formes d'expression linguistique à même de contribuer à la cohérence des discours et d'influer sur leur traitement aussi bien par des humains que par des programmes informatiques. Une des spécificités de la revue est de mettre en ligne, à disposition de la communauté, des ressources linguistiques élaborées en amont des travaux publiés, tels que des corpus annotés, tableaux de codage et outils d'analyse.

Mon travail sur les relations intersubjectives dans le discours rapporté a débouché sur une première formalisation des paramètres énonciatifs sous forme d'ontologie, en vue de son exploitation dans des applications relevant du Web sémantique (OntoDiscours). Cette voie a été jugée prometteuse et est actuellement poursuivie dans le cadre du projet OntOpiTex. La problématique de l'expression linguistique des opinions s'est imposée progressivement dès 2008, pour aboutir en octobre 2009 à un projet ANR, OntOpiTex, réunissant les laboratoires LaLIC, GREYC, CRISCO, ainsi que les entreprises TKM (Grenoble) et NOOPSIS (Caen). Ce projet porte sur l'extraction automatique et la caractérisation des opinions et des jugements d'évaluation dans des textes. Son originalité est de s'appuyer sur des modélisations linguistiques et ontologiques concernant d'une part les notions évaluatives et d'autre part les objets cibles de l'évaluation. Le travail sur le langage de l'évaluation, tout en bénéficiant des acquis des recherches sur la prise en charge et le discours rapporté, a ouvert un champ d'exploration plus large. Les réflexions collectives menées dans ce consortium, portent notamment sur les rapports entre les jugements, les normes et les paradigmes évaluatifs.

Une deuxième piste née des recherches sur la subjectivité dans le discours rapporté m'a guidée vers l'étude des discours en circulation. Le travail a porté sur l'analyse et la détection automatique de formules politiques à fort impact. Une étude complète de faisabilité a été réalisée en collaboration avec la société Noopsis. Grâce à cette expérience notamment, j'ai pu nouer une collaboration avec une équipe de chercheurs du laboratoire PRAXILING (Montpellier 3) qui travaillent sur des pratiques médiatiques (François Perea et Laurent Fauré). Actuellement, je participe à la création d'un Observatoire des pratiques médiatiques émergentes (OPME), piloté par PRAXILING et soutenu par la MSH de Montpellier. L'objectif principal du projet est une contribution à l'observation des pratiques d'usage et de production médiatiques dans la vie sociale et institutionnelle. Cette problématique est abordée selon trois angles en interaction étroite : (i) les productions socioprofessionnelles, (ii) les dispositifs technologiques, (iii) les pratiques des usagers et le travail de réception. Dans ce cadre, j'assume la responsabilité d'un axe particulier : « Interface, données et métadonnées ». Son objet est l'analyse des données enregistrées sur le terrain et des recueils de productions afférentes (écrites, audio-visuelles, numériques, en ligne ou non). Ma participation personnelle directe concerne le balisage et l'extraction des énoncés d'opinion et d'évaluation dans les médias. Ce projet ouvre un nouveau volet à mes travaux, il s'agit de la dimension sociétale de l'analyse des discours.

Si mon parcours ne m'a pas conduit « physiquement » dans plusieurs lieux et laboratoires, j'ai toujours cherché à apprendre de ce qui se fait ailleurs. Je dois donc beaucoup à de nombreux

chercheurs avec qui je n'ai pas collaboré directement et qui de par leurs écrits ont nourri mes réflexions et mes recherches.

3. Enseignement

Mes activités de recherche sur l'énonciation et les organisations discursives sont cohérentes avec les enseignements et la direction des mémoires que j'assure dans le cadre du Master « Information et communication » *Ingénierie de la langue pour la gestion intelligente de l'information (ILGI)* à Paris 4. En Master 1, je suis chargée d'un cours intitulé « *Approches de l'énonciation et linguistique de corpus* » et d'un cours de « *Sémantique cognitive* ». En Master 2, je propose un enseignement sur les « *Composantes textuelles et structures pour le TAL* ». Dans ces trois unités, les étudiants sont amenés à acquérir des savoirs fondamentaux en linguistique, qui sont abordés dans la perspective des multiples applications de l'ingénierie de la langue. Ces applications sont vues non seulement d'un point de vue technique (quelles connaissances linguistiques pour quelles procédures ?) mais aussi d'un point de vue sociétal. A quels besoins précis souhaitons-nous répondre ? La « résolution » de certains problèmes sociétaux passe en effet par des applications de type TAL (la détection du plagiat, des contenus racistes ou diffamatoires...). En retour, que pouvons-nous raisonnablement attendre des analyses outillées menées sur de vastes corpus textuels ? Quels peuvent en être des apports pour le champ des humanités (représentations culturelles, problématiques sociales...) ? L'important pour moi est de positionner clairement ces enseignements dans un champ plus vaste, celui des sciences humaines et sociales.

Mes activités d'enseignante et mes recherches propres se nourrissent en particulier grâce à la direction des mémoires de master et des projets d'intégration (voir la partie Annexes). Je travaille essentiellement avec les étudiants en master professionnel. Pour cette raison, peu de travaux ont donné lieu à des publications scientifiques. En revanche, l'insertion professionnelle des étudiants, notamment dans le domaine de la veille et de l'e-réputation, est un indicateur significatif de la pertinence des savoirs transmis et des problématiques abordées.

Le deuxième point que je souhaite souligner est l'importance que revêt pour moi le fait de construire le sujet de recherche avec participation active de l'étudiant, en s'appuyant fortement sur son bagage propre. Une formation initiale en lettres modernes s'avère bénéfique pour travailler en finesse sur des aspects sémantiques. Etre issu du giron des sciences politiques permet d'aborder avec facilité les aspects sociétaux de la veille. Une formation d'ingénieur facilite le recours éclairé à des technologies, ce qui mène plus naturellement vers une conception d'outils d'analyse linguistique efficaces et adaptés. Pouvoir travailler avec des publics aussi différents est de toute évidence très stimulant pour un enseignant. De plus, ayant expérimenté durant plusieurs années un mode de travail collectif, avec des groupes de deux, trois ou quatre étudiants, chargés de sujets présentant une complémentarité à l'intérieur d'un même domaine notionnel, j'ai constaté une nette amélioration de la qualité des travaux rendus. La réflexion commune et la mutualisation de certaines tâches permettent de diminuer de manière notable le sentiment d'insécurité qu'éprouvent souvent les étudiants face à ce premier travail personnel de grande ampleur. On perçoit une adhésion plus forte à la problématique abordée et, au final, une plus grande assurance au moment de la présentation des travaux réalisés.

Les traitements informatiques ouvrent des perspectives nouvelles pour l'enseignement de la linguistique. Il devient possible de mettre en place de véritables travaux dirigés où un phénomène linguistique choisi (le pronom *nous*, l'antonomase du nom propre, les verbes de parole, les marques d'intégration linéaire...) est d'abord observé à l'intérieur d'un corpus, puis extrait, encodé, et discuté sur la base des analyses qualitatives et quantitatives. Les travaux des étudiants deviennent de ce fait entièrement traçables. Ce faisant, il s'agit d'apprendre aux étudiants une attitude confiante mais critique face aux outils et aux méthodes du TAL. Savoir s'approprier ces outils, en utilisateur éclairé, pour être capable d'apprécier justement la pertinence et la portée des résultats fournis par les programmes. Car les résultats issus des traitements informatiques peuvent facilement faire illusion. Il est important pour cela que les méthodes classiques et éprouvées de la linguistique soient connues et maîtrisées par les étudiants.

J'assure également depuis ma thèse des enseignements d'informatique destinés à des étudiants en sciences humaines et notamment aux étudiants en licence de *Langue française et techniques informatiques (LFTI)*. Il s'agit des cours d'algorithmique et de programmation impérative, ainsi que d'un cours sur les bases de données relationnelles. Ici, l'apprentissage d'une pensée rigoureuse,

d'une démarche structurée, modulaire et systématique prépare les étudiants à pouvoir aborder ensuite, de manière formelle notamment, des problématiques linguistiques complexes. Ce cursus généraliste a pour objectif de former des étudiants orientés vers l'analyse du langage et de la langue française qui soient également des utilisateurs experts des outils mis à disposition dans ce domaine par les techniques informatiques et les nouvelles technologies.

Il est à noter que cette formation existe à l'université de Paris-Sorbonne depuis 2000 et fait partie des premières licences bi-disciplinaires. Dès ma prise de fonctions, j'ai participé à l'élaboration des maquettes et à la mise en place de deux deug bi-disciplinaires à Paris 4 : *Langue française et Informatique* et *Langues et Informatique* (avec l'Anglais, le Russe ou l'Allemand). Cette expérience pilote a été l'occasion d'une réflexion approfondie sur les enjeux, les possibilités et les difficultés d'une formation interdisciplinaire en informatique, logique, mathématique et linguistique.

4. L'énonciation : l'axe de l'intelligibilité des discours

Mon domaine de recherches se situe au croisement de la linguistique et du traitement automatique des langues. Le fil directeur de mes travaux est l'énonciation appréhendée à l'échelle du discours. Le questionnement qui traverse les différentes études que j'ai menées depuis ma thèse de doctorat est comment la subjectivité affecte et informe les objets de discours ? Comment le sujet énonciateur gère et assume le fait d'être également sujet connaissant, sujet éthique, sujet social et sujet communicant ? Je propose une sorte de « décryptage de l'énonciation à l'usage du TAL », montrant comment décoder les instructions laissées au lecteur au moyen de certaines marques énonciatives.

Pourquoi choisir comme destinataire privilégié l'univers du traitement automatique des langues ?

Dans la hiérarchie, aujourd'hui devenue classique en TAL, des niveaux d'analyse à laquelle il est nécessaire de soumettre un texte pour permettre son interprétation, le niveau énonciatif (généralement inclus dans l'intension du terme pragmatique³) constitue l'étape ultime. On y associe habituellement le recours à certains repères contextuels et co-textuels (spatiaux, temporels, interpersonnels...) pour calculer la référence, ainsi que l'incidence de l'implicite (du présupposé ou du sous-entendu) sur les valeurs de succès des actes de discours. On exemplifie un certain nombre de cas d'école, pour montrer la nécessité, et surtout la difficulté, de tenir compte des connaissances extra-linguistiques. L'inventaire des systèmes informatiques qui assurent ces différents types d'analyse s'arrête, pour sa part, au niveau sémantique, pour lequel relativement peu de réalisations effectives peuvent être exhibées. La composante énonciative des productions textuelles, quant à elle, n'est (pratiquement ?) jamais considérée comme objet de traitement à part entière.

Or, l'énonciation est une dimension fondamentale des langues et des discours.

« L'énonciation constitue le pivot de la relation entre la langue et le monde : d'un côté elle permet de représenter dans l'énoncé des faits, mais d'un autre côté elle constitue elle-même un fait, un événement unique défini dans l'espace et le temps » (Charaudeau et Maingueneau 2002)

Définie, avec (Benveniste 1974, p.80), comme « la mise en fonctionnement de la langue par un acte individuel d'utilisation », l'énonciation prend en charge un énonçable en le plongeant dans un système de référence intersujets, en relation avec l'espace-temps. Les opérations de référencement (permettant de référer à la réalité extralinguistique) et les opérations de modalisation (permettant à l'énonciateur de se situer par rapport à ce qu'il dit et à son interlocuteur) s'en trouvent étroitement liées (Culioli 1990).

Plusieurs linguistes (Culioli, Ducrot, Authier-Revuz, Kerbrat-Orecchioni, Desclés...) ont contribué au développement d'une approche étendue de l'énonciation, fondée sur la description détaillée des relations entre l'énoncé et les éléments constitutifs du cadre énonciatif. Dans cette perspective, l'énonciation est vue comme un phénomène langagier complexe qui témoigne également de la façon dont le sujet parlant s'approprie la langue pour organiser, présenter et conduire son discours, s'emparer du discours d'autrui, réaliser des actes de langage, s'avouer être à l'origine des opinions et des jugements etc.

³ On a souvent tendance à intégrer les théories de l'énonciation dans les courants pragmatiques.

« ... la problématique de l'énonciation (la nôtre) peut être ainsi définie : c'est la recherche des procédés linguistiques (shifteurs, modalisateurs, termes évaluatifs, etc.) par lesquels le locuteur imprime sa marque à l'énoncé, s'inscrit dans le message (implicitement ou explicitement) et se situe par rapport à lui (problème de la 'distance énonciative'). (Kerbrat-Orecchioni 2002, p. 36)

Les discours que nous cherchons à analyser et à traiter automatiquement demandent à être considérés tels qu'ils sont : des événements énonciatifs produits dans des situations concrètes, non seulement pour représenter une « réalité » extra-linguistique, mais aussi pour entrer en dialogue (au sens large du terme) avec son interlocuteur et son propre dire. Ils relèvent toujours d'une activité sociale qui implique un moment, un lieu, une visée, des partenaires d'un certain type. Les diverses attitudes que l'énonciateur donne à voir dans ses productions langagières (s'impliquer, prendre position, évaluer ou juger...) construisent et alimentent son image. En même temps, elles jouent, à des degrés divers, sur la manière dont sera interprété le contenu (l'état du monde) dénoté par ses paroles.

Je défends l'idée que l'exploitation avertie des propriétés énonciatives des discours peut conduire à améliorer sensiblement la qualité des résultats obtenus par différentes applications informatiques (de résumé, de filtrage, de recherche d'information, etc.).

Plus généralement, les rapports entre l'informatique (le TAL, le WS, la RI) et la linguistique ne seront véritablement féconds que si l'on sort de la vision des textes comme lieu de représentation des connaissances, avec d'un côté un monde de choses et d'activités « muettes » et de l'autre des représentations textuelles détachées de leur source énonciative, qui en offriraient une image

Quand on réfléchit en termes d'énonciation, on a accès à des phénomènes linguistiques d'une grande finesse, où se mêlent étroitement la référence au monde et l'incontournable inscription de l'énonciateur dans son propre discours. En faisant de l'énonciation l'axe d'intelligibilité des discours, pour reprendre la formule de Maingueneau⁴, on se dote de moyens permettant d'analyser les dispositifs de parole en fonctionnement : avec les conditions énonciatives qui traversent ce qui est dit, où ce qui est dit renvoie à ses propres conditions d'énonciation (la relation au destinataire, les supports matériels, le mode de circulation des énoncés...). De plus, une telle vision des productions linguistiques met en avant et exploite ce qui constitue leur spécificité, par rapport à d'autres supports susceptibles de traitements formels. Rappelons que tout discours, et seul le discours, se construit à travers l'énonciation.

Du fait de l'énonciation, l'objectivité et la subjectivité dans les productions textuelles se trouvent intimement liées. Pour ma part, j'ai cherché à décrire finement les différentes manières dont peut s'exprimer la subjectivité des énonciateurs au sein des objets discursifs aussi différents que les énoncés causaux, les énumérations, les discours rapportés et les jugements évaluatifs. J'ai tenté de rendre compte de l'intrication de ces deux dimensions (objective et subjective, référentielle et modale), notamment pour ce qui concerne la causalité et l'évaluation.

⁴ <http://www.vox-poetica.org/t/articles/maingueneau.html>

QUESTIONS DE MÉTHODOLOGIE

Une discipline d'application, comme celle que je pratique, associe nécessairement des connaissances et des méthodes issues de domaines variés. Cela offre des possibilités inédites en matière d'analyse et de traitements. Mais, ses motivations, ses pratiques, ses résultats et ses difficultés sont parfois moins lisibles, plus difficiles à cerner et à apprécier. La clarté méthodologique et la précision des critères d'analyse deviennent alors fondamentales.

Peut-être convient-il de nuancer cette dichotomie franche entre linguistique appliquée et linguistique fondamentale, comme invite à le faire Dominique Maingueneau (ci-dessus). Y compris en ce qui concerne mes propres travaux. J'espère pouvoir montrer que les visées applicatives non seulement ne m'éloignent pas des préoccupations théoriques, mais au contraire offrent des angles d'approche pertinents, permettant de progresser dans la connaissance des faits linguistiques choisis pour l'étude.

« Avec la révolution apportée par l'informatique il sera de plus en plus difficile de séparer linguistique appliquée et non appliquée : nombre de recherches sur le langage qui relèvent de la recherche fondamentale sont d'ores et déjà menées en fonction de leur retombées prévisibles sur les 'industries de la langue' » (Maingueneau 2009, p.164).

Les problèmes que j'ai mis en relief dans cette partie me semblent symptomatiques des recherches en linguistique informatique en général. J'ai isolé les points de méthodologie de la présentation des recherches à proprement parler, pour rendre plus lisibles ces dernières et surtout parce que, pour l'essentiel, ces points leurs sont communs.

1. Regard interdisciplinaire et démarche onomasiologique

La linguistique que je pratique est celle qui part des besoins et tente d'y répondre. C'est non seulement une linguistique appliquée, mais aussi et surtout une linguistique qui collabore, qui se frotte à d'autres disciplines, qui accepte de renoncer au confort de ses habitudes quant aux procédés ou aux méthodes classiques, éprouvés et validés.

C'est un regard interdisciplinaire et onomasiologique que je tente de poser sur les objets que j'étudie.

La langue ne donne pas corps à la notion de causalité, mais permet de l'exprimer dans une grande diversité d'acceptions possibles. De même, la notion d'opinion prend de multiples significations et les formes de surface qui permettent de l'exprimer sont extrêmement variées. Le discours rapporté, objet plus proprement linguistique, n'est pas pour autant plus aisé à appréhender, car les pratiques sociales en façonnent des formes fort variées. Je me trouve obligée de « reconquérir », en quelque sorte, ces objets d'étude, en construisant un savoir linguistique, participant d'une interdisciplinarité, qui semble seule à même de rendre compte de qu'ils représentent.

Les notions d'opinion et de causalité, polythétiques, échappent très clairement à toute approche disciplinaire renfermée sur elle-même. Pour comprendre le concept d'opinion, outre la philosophie, la sociologie et les sciences politiques apportent aujourd'hui des éclairages indispensables et utiles. Le cas de la causalité est plus complexe encore et de ce fait impossible à résumer en quelques phrases. Tour à tour, objet de recherche et outil d'investigation, la causalité intéresse de nombreuses disciplines (physique, biologie, médecine, philosophie, sociologie, psychologie, linguistique, intelligence artificielle...), qui l'approchent et se l'approprient avec la terminologie et les modèles qui leur sont propres. Son étude a donné lieu à une diversité d'approches théoriques fort divergentes entre elles (notion primitive, notion mise en cause dans son bien fondé même, notion définie de manière minimale : « quelque chose qui agit », notion approchée par des typologies : outre les quatre causes d'Aristote, il en existe des centaines...). Si, aujourd'hui, comme le constate Bernard Saint-Sernin (Andler *et al.* 2002, p.937), « la causalité regagne son crédit, mais perd son identité », il est certainement illusoire de vouloir la « fixer » d'un point de vue linguistique.

Classiquement, les linguistes optent pour la démarche sémasiologique dont le but est d'étudier l'unité linguistique à partir des oppositions paradigmatiques de signifiants et des combinaisons entre cette unité et les autres unités de la chaîne. C'est une démarche qui part du signifiant pour

découvrir et aboutir au signifié. On s'intéresse au terme en tant que signe, fonctionnant dans un système linguistique particulier à une société, une culture, une vision du monde.

La spécificité de mes recherches tient au fait qu'elles concernent non pas des phénomènes spécifiquement linguistiques (temps verbaux, prédication...), mais des notions ayant (surtout ou également) une existence extra-linguistique et qui sont exprimées par la langue. Il s'agit donc d'une démarche onomasiologique, qui va d'une idée (d'un concept...) vers ses expressions dans une langue.

Le parcours onomasiologique part du signifié pour découvrir le signifiant. Or, la détermination du signifié ne va pas sans difficultés. Certaines notions fondamentales, comme nous venons de le voir, échappent à la définition. Les catégories sémantiques par lesquelles on tente de les approcher sont parfois définies *a priori*, selon des principes (de symétrie, complémentarité...) propres à satisfaire la rationalité, ou plutôt une certaine logique, de l'analyste humain. Les limites de la démarche onomasiologique tiennent au risque d'imposer à un système linguistique des signifiés qu'ils n'ont pas. Certaines modélisations donnent l'impression que les données linguistiques doivent de gré ou de force s'ajuster au tableau prédéfini. Cet état de fait est très clairement caractéristique des nombreuses études sur la causalité. Les recherches sur les opinions et les sentiments, qui sont aujourd'hui en plein essor, se trouvent confrontées à ce même problème.

2. Ce qu'ils sont : leur nature, leurs fonctions et leurs marques

Dans les différents projets auxquels j'ai participé, nous avons cherché à reconnaître, à partir des traces matérielles du texte, certaines notions, certains objets textuels ou certains types d'information. Mais, précisément, que veut dire « reconnaître dans un texte » une notion telle que causalité ou opinion, identifier un objet grammatical comme le discours rapporté ? Peut-on raisonnablement poser un tel objectif en toute généralité ? Quels sont les écueils d'un tel projet ?

J'ai travaillé sur des objets fonctionnels qui s'avèrent complexes et protéiformes. Les discours rapportés se construisent à partir de plusieurs actions énonciatives (restitution, authentification, interprétation, évaluation, explication) et de ce fait peuvent former des séquences composites de taille variable. Les séries linéaires correspondent à des structures textuelles à la longueur théoriquement illimitée, pouvant aller de la conjonction de deux segments simples jusqu'à des architectures enchâssées à l'empan de plusieurs centaines de mots. Les opinions intègrent dans les différentes configurations possibles la mention de la cible, de la source, de l'attitude, de l'intensité, de l'engagement. La causalité recouvre des situations plus ou moins complexes, s'exprimant parfois à l'échelle de plusieurs phrases (conjonction de plusieurs causes, énumération des effets, chaînes causales, pluralité de faits corrélés...).

Ces notions ou objets textuels n'ont pas les contours clairs. Notamment, du fait de la proximité avec des notions apparentées. Causalité et action sont intimement liées, qualification et évaluation, pas toujours distinguables. Les discours rapportés recouvrent, comme le montre (Rosier 2008), un vaste continuum allant des formes canoniques, via les formes libres et hybrides, aux formes aux confins. Et leurs frontières avec les discours directs, pouvant cacher des discours autres assimilés, sont parfois très ténues. Les séries binaires peuvent se confondre avec certaines configurations reliant (opposant...) deux éléments textuels.

Des rapports contextuels, plus ou moins motivés, peuvent les unir à d'autres faits ou objets linguistiques. L'expression de la causalité dans certains types de textes est intimement liée à la notion de prise en charge. La qualification et la catégorisation vont souvent ensemble, car la saillance des qualités semble dépendre de l'appartenance catégorielle des cibles. La sériation qui supporte souvent des séquences argumentatives peut se trouver balisée par les marques de celles-ci.

Enfin, au sein des énoncés, les notions étudiées peuvent s'intégrer à l'argumentation de l'auteur de diverses manières. Concernant la causalité, les distinctions élaborées touchent au contexte d'expression et à la visée pragmatique de celui qui élabore ou exploite une relation causale. L'énonciateur peut poser l'existence d'une relation causale en prenant la responsabilité de ce qu'il établit. Mais, il peut aussi exploiter une telle relation, en la reprenant à son compte, pour argumenter ou étayer en faveur d'un contenu autre. De la même façon, au sein des jugements d'évaluation, on constate que les moyens linguistiques permettant d'attribuer une qualité à un individu et ceux qui ne font que la convoquer à des fins autres (d'explication, d'élaboration...)

différent nettement. La question de savoir comment ces différents contenus peuvent s'articuler avec le co-texte ne peut donc être évacuée des analyses menées.

Tous ces aspects sont bien évidemment cumulatifs. La manière dont Laurence Rosier, spécialiste du discours rapporté, choisit de caractériser le discours rapporté montre clairement les différents facteurs avec lesquels on est obligé de composer.

« Avant d'être un phénomène grammatical, le discours rapporté est une pratique, sociale et banale, aux formes et aux enjeux multiples. Sa maîtrise relève donc non seulement d'une bonne connaissance de la langue (concordance temporelle et personnelle, transfert d'un discours dans un autre contexte, passage de l'oral à l'écrit ou vice-versa, etc.), mais aussi de la prise en compte des situations de communication et des contraintes des univers de discours » (Rosier 2008, 4^e de couverture).

Reconnaître une notion dans un texte amène à considérer un enchevêtrement de phénomènes linguistiques, se manifestant par des configurations variables d'indices. En règle générale, mon travail a consisté à démêler les facteurs en cause, à produire des descriptions qui portent sur des corrélations de phénomènes sémantiques, énonciatifs et discursifs, à recueillir sur corpus leurs marques effectives.

3. Entre l'empirique et le théorique

Les approches sur corpus sont souvent plus méthodologiques et empiriques que théoriques. Dans mes différents travaux, j'ai privilégié l'observation minutieuse des textes, en prenant, chaque fois que cela était possible, un appui sur des réflexions théoriques.

Travailler sur les données verbales attestées (« brutes » et non épurées), issues des usages effectifs que font de la langue les locuteurs, m'amène à formuler plusieurs observations. Entre la stabilité et la variation, l'explicite et l'implicite, ce qui est fréquent et ce qui est rare, ce qui est simple, simplifiable et ce qui demeure intrinsèquement complexe, ce qui appartient à la langue « générale » et ce qui ne fonctionne que dans des registres particuliers, la tâche de l'analyste s'apparente parfois à une forme d'artisanat, qui répond difficilement aux règles d'une méthodologie scientifique.

Prendre en compte les phénomènes des discours, c'est nécessairement se poser la question du spécifique et du généralisable, du contingent et de l'invariant. Or il n'est pas toujours possible d'y accorder toute l'attention nécessaire. J'ai connu l'inévitable tiraillement entre l'observation d'un sous-ensemble des possibles issus d'un corpus et la généralisation attendue. Le corpus est une manifestation des possibilités de la langue ou d'un discours, mais n'est pas en tant que tel « la langue ». De plus, si le corpus constitue un échantillon de la langue, il est également spécifique des registres et des situations de productions spécifiques, ce que le linguiste se doit de garder en tête.

Quelle est la fréquence d'un phénomène linguistique dans un discours ? Quel est le taux d'ambiguïté d'un terme, d'une construction, d'une structure (constructions...) que l'on tente d'étudier ? Si je décide de travailler sur les segments axiologiques détachés, comment estimer la probabilité de les trouver dans un corpus ? Pour parler plus franchement, est-il « opportun » d'ouvrir un tel chantier ?

Analyser les formes d'expressions associées à un objet ou une notion fait perdre le confort et la sécurité d'un travail qui porte sur un système de formes choisies à l'avance. Décrire un couple d'adverbiaux corrélatifs, comme par exemple *d'une part* et *d'autre part*, même si l'on s'appuie sur les observables attestés n'est pas la même tâche que celle qui mène à identifier toutes les possibilités formelles de balisage d'une série et toutes les formes effectivement employées.

J'ai constaté que la confrontation avec les fonctionnements attestés produit souvent le sentiment que la description est plus complexe qu'on ne le pensait. Cela oblige parfois à modifier ses conceptualisations, si elles ne correspondent pas aux données langagières que l'on souhaite décrire. Cette problématique de la modélisation m'amène aux questions qui touchent à la nature des catégories sémantiques pertinentes pour la description des notions analysées.

La construction de telles catégorisations pose au moins deux problèmes bien connus : (i) celui du « bon » niveau de modélisation (de généralité), (ii) celui de l'appellation des concepts et de leur définition. On constate fréquemment en linguistique que les théories concernées par la même problématique ne découpent pas leurs objets et ne les nomment pas de la même manière. Cela est

particulièrement net dans le cas de la causalité, comme je l'ai montré dans ma thèse (Jackiewicz 1998). Mais, des objets intrinsèquement linguistiques, comme par exemple les relations rhétoriques⁵, n'échappent pas, non plus, à l'inflation terminologique et à un certain arbitraire dans les architectures notionnelles proposées⁶.

Il n'est pas inutile parfois d'observer les divergences de conceptualisation qui découlent des pratiques disciplinaires différentes. Prenons le cas des jugements évaluatifs. La théorie *Appraisal* propose pour sa catégorie « Jugement » un ensemble d'attitudes qui peuvent être valorisées ou dévalorisées socialement, à savoir la capacité, la ténacité, la normalité, la véracité et la décence. Tant le choix de ces attitudes et que leur organisation hiérarchique (regroupement selon l'estime sociale et la sanction sociale) sont le fait des concepteurs du modèle, qui sont des linguistes. La description des compétences sociales que l'on trouve sous la plume d'un sociologue, Michael Argyle dans (Moscovici 1994, pp. 87-114), propose un tout autre regard sur cette même « réalité ». Elle distingue comme principales composantes : (i) l'assertivité, (ii) la gratification et le soutien, (iii) la communication non verbale, (iv) la communication verbale, (v) l'empathie, la coopération et l'attention portée aux autres, (vi) cognition et résolution de problèmes, (vii) la présentation de soi. Ces deux conceptualisations, tout aussi légitimes dans leurs cadres respectifs, s'avèrent somme toute incomparables.

L'examen approfondi de l'architecture notionnelle complète du modèle *Appraisal*, réalisé dans le cadre du projet ANR OntOpiTex relativement à la tâche de caractérisation sémantique de segments textuels porteurs d'opinions, a permis de mettre en évidence un certain nombre de difficultés⁷. Celles-ci sont liées notamment à la couverture partielle des notions représentées et à l'absence de critères clairs et opératoires présidant à leur définition et leur articulation. Ce bilan ne conduit pas à rejeter le modèle dans sa globalité, mais à l'amender de manière pragmatique pour qu'il s'adapte à la nature des textes à traiter, suivant les tâches que l'on cherche à entreprendre.

En TAL, la disponibilité de gros corpus et la forte opérationnalisation du domaine permettent aujourd'hui de reposer la question de la représentation du sens, longtemps considérée comme hors de portée pour l'ordinateur. Mais, c'est très clairement une sémantique fondée sur l'usage, circonscrite à des domaines particuliers et des corpus restreints que l'on cherche à développer. La visée de cette dernière est donc plus modeste, car la modélisation ne vise aucunement l'universalité et les critères de validation reposent en grande partie sur la satisfaction des besoins. La situation la plus radicale étant sans doute celle qui est pointée par Anne Condamines :

« Pour le TAL ou l'ingénierie des connaissances, l'objectif applicatif constitue d'emblée un mode de validation : lorsque la satisfaction d'un besoin identifié initialement a été atteinte, la validation de la pertinence des méthodes et des outils est acquise. » (Condamines, 2005, p.35.)

4. Corpus, normes et usages

Mon matériau d'étude est constitué d'énoncés authentiques, écrits, provenant essentiellement de textes journalistiques (*Le Monde, le Monde diplomatique, La Recherche...*), représentatifs de l'usage de la langue française contemporaine dans les médias.

⁵ On peut consulter le document de Eduard Hovy et Elisabeth Maier (1995) : « Parsimonious or profligate : how many and which discourse structure relations », Unpublished ms., mis en ligne sur <http://www.isi.edu/natural-language/people/hovy/papers/93discproc.pdf>.

⁶ Le danger de l'impressionnisme qui guette les architectures sémantiques peut être illustré de manière anecdotique par un exemple issu de la littérature. Dans une nouvelle de Borges, « *La langue analytique de John Wilkins* », il est question d'une encyclopédie chinoise apocryphe, « *Le Marché céleste des connaissances bénévoles* », dans laquelle les animaux sont classés selon une nomenclature absolument fantaisiste. « Dans les pages lointaines de ce livre, il est écrit que les animaux se divisent en a) appartenant à l'Empereur, b) embaumés, c) apprivoisés, d) cochons de lait, e) sirènes, f) fabuleux, g) chiens en liberté, h) inclus dans la présente classification, i) qui s'agitent comme des fous, j) innombrables, k) dessinés avec un pinceau très fin en poils de chameau, l) *et cætera*, m) qui viennent de casser la cruche, n) qui de loin semblent des mouches. »

⁷ Le bilan détaillé de cette opération est présenté dans l'ouvrage nouveau (voir le paragraphe 2.3.3.2 « Bilan sur le modèle *Appraisal* »).

Travailler sur des données attestées amène à prendre acte des pratiques sociales et des normes sous-jacentes. Idéalement, les corpus dont proviennent les données devraient être clairement caractérisés du point de vue des normes régissant l'écriture des textes qui les composent. Mais, il s'agit de toute évidence de phénomènes socio-linguistiques complexes. Des formes de standardisation culturelle et institutionnelle sont visibles dans les médias, comme dans toutes les sphères de la vie socio-professionnelle. Elles correspondent notamment à une sorte de marché linguistique dans lequel certaines manières de d'interagir (parler) sont valorisées, d'autres dévalorisées. Pierre Bourdieu parle à ce propos d'une forme de capital linguistique et culturel. Mais, on note également des phénomènes de mode, lesquels pèsent parfois très nettement sur les choix d'expression des analystes.

« Quand nous lisons les commentaires qui chantent les louanges d'un propos percutant, de l'œuvre d'un philosophe, de la pièce d'un auteur dramatique, du documentaire persiflant d'un cinéaste ou du livre d'un sociologue, les qualités qu'on leur attribue ne sont pas celles que nous venons d'énumérer, mais bien celles qui sont à la mode. Nous rencontrons alors les termes élogieux comme 'dérangeant', 'radical', 'surprenant', 'subversif', 'prophétique', 'visionnaire', 'décoiffant'... » (Rezsóhazy 2006, p.162)

En règle générale, le linguiste qui exploite un corpus textuel n'est pas averti des normes éditoriales et des règles qui régissent les pratiques d'écriture dans les univers professionnels des auteurs⁸. De toute évidence, les normes qui gouvernent les écrits des médias sont de plusieurs ordres. Elles régissent en priorité le choix et la gestion des contenus. Elles commandent de manière plus technique le traitement des supports. Enfin, de manière plus ou moins explicite, elles concernent les interactions sociales qui s'opèrent au moyen de ces écrits.

La question des contenus touche à un grand nombre de facteurs. Les articles de type biographique, par exemple, en tant que vecteur « performatif » d'identité personnelle et sociale des individus, plongent au cœur des rapports entre le privé et le public, le personnel et le politique. Une telle pratique soulève des enjeux éthiques et déontologiques autour de la protection de l'intime, du privé, de la liberté d'expression, des compétences des journalistes et des analystes. Pour ce qui concerne la presse telle que *le Monde* ou *les Echos*, on note le caractère courtois, relativement peu contestataire, raisonnablement critique des jugements qui y exprimés. Les journalistes forment des impressions mais restent attentifs à ce qu'ils n'aient pas à pâtir de leurs propos. Cela distingue très nettement ce type d'observables des données que l'on peut trouver par exemple dans des blogs ou forums, ou tout simplement dans la littérature⁹ (romans...).

La gestion des supports concerne les questions de la place allouée (taille des unités), de la structuration et de la hiérarchisation des propos, de la redondance de l'information et de sa mise en valeur. Dans les médias actuels, on s'adresse à des lecteurs dont le temps est précieux. Les moyens permettant de les guider et assister (pour soutenir leur mémoire notamment) sont multiples : (i) structuration en sections avec usage de titres et d'accroches, (ii) multiplication de repères visuels, (iii) ajout du paratexte (chapô d'ouverture, clés, trois questions à...), (iv) écriture selon le principe de la « pyramide inversée » où l'essentiel est dit dans les premières lignes et développé par la suite, (v) mise en valeur de ce qui est important, (vi) écriture fortement segmentée (détachement, formes averbales...). L'influence des usages sur la matérialité des textes peut donc être conséquente. De plus, à tous les niveaux, les formes de signalisation choisies sont loin d'être anodines. On constate parfois que le résultat obtenu n'est pas entièrement satisfaisant (indices trop faibles ou plurivoques, balisage insuffisant...), ce qui augmente le coût cognitif de la lecture. Pour éclairer ce dernier point, considérons un cas attesté.

La « machine éditoriale » façonne le texte à publier, ce qui conditionne la manière dont ce texte sera lu. Dès lors, que deviennent les instructions adressées originellement au lecteur par l'auteur lui-même (expert d'un domaine, par exemple) ? Des conflits entre l'organisation interne (discursive) du texte et son organisation typo-dispositionnelle (répondant notamment aux normes de régularité visuelle de la masse textuelle qui règlent le découpage physique en sections) peuvent se produire,

⁸ Pour ma part, j'ai eu la possibilité de consulter le guide rédactionnel à l'usage des journalistes « *Le style du Monde* » (Le Monde 2002). Ce document décrit les différents types d'articles et les caractéristiques qui leurs sont associées, notamment en ce qui concerne le rapport à la subjectivité et à l'expression des opinions. Il fournit entre autres des recommandations en matière des citations.

⁹ Voir l'ouvrage nouveau, Etude 3, section 4.1.4.

comme c'est le cas dans l'extrait issu du journal *la Recherche* (n°369, pp.36-37), schématisé dans la figure 2.

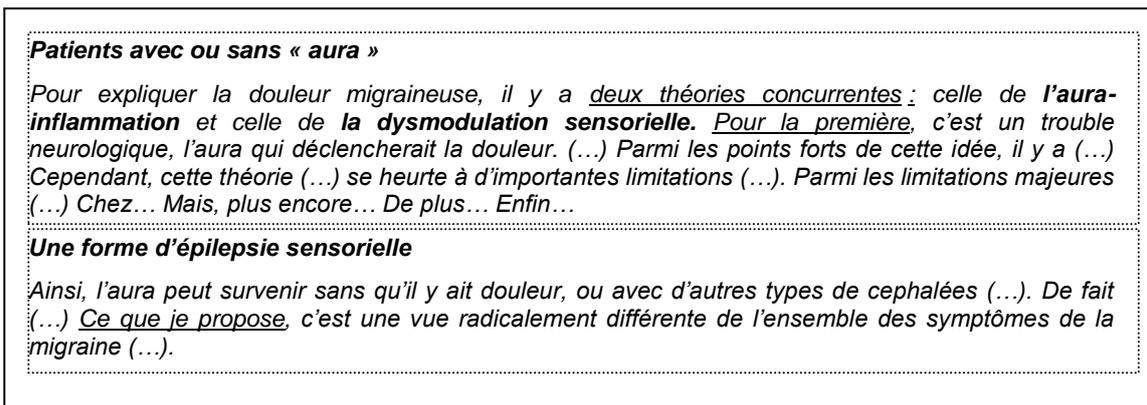


Figure 2 : Représentation schématisée d'un conflit entre la structuration typo-dispositionnelle renforcée par les accroches ajoutées par le journaliste (le cadrage) et la structuration discursive originelle opérée par le scientifique, auteur du texte (le soulignement)¹⁰.

La question des interactions sociales qui se jouent au travers des écrits de la presse est plus complexe encore. Si nombre de ces aspects échappent entièrement au linguiste, certains peuvent être pris en considération pour éclairer sa démarche. Prenons le cas de la citation. Les différentes pratiques impliquant la transmission, l'emprunt, l'appropriation, le travestissement du discours d'autrui doivent obéir à des règles précises. Ces règles ou recommandations commandent des choix d'expression. Par exemple, la consigne stipulant que « le style de la personne est respecté et restitué au plus près » (*Le Monde* 2002, p. 8) peut inciter les journalistes à diversifier notablement les verbes de parole choisis pour introduire les discours cités.

Tous les aspects que je viens de signaler, liés aux normes et aux usages, font partie intégrante de « la réalité » des textes que l'on cherche à analyser et à traiter.

5. Des notions aux marques, des marques aux ressources

Comment passe-t-on des notions aux marques linguistiques et des marques aux ressources nécessaires à la machine ? Ces interrogations sont au cœur de la linguistique informatique.

La question du rapport entre marques et notions est l'une des plus épineuses. Comment reconnaît-on une notion ? Comment passe-t-on des marques aux valeurs qu'elles expriment dans des énoncés ? Les marques ne sont pas la notion, elles y sont associées en contexte. Mais dans quel contexte précisément ?

Il est évident que l'on capte les figures les plus « voyantes », les lexèmes porteurs d'une signification la plus affirmée, des marqueurs les plus manifestes. Or, les éléments environnants jouent toujours un rôle. Ils sont parfois déterminants. Identifier ceux dont le rapport à la marque « principale » est significatif relève parfois d'une expertise linguistique qui ne s'acquière pas immédiatement. Par exemple, la marque du degré est un indice fort de la nature subjective du qualificatif. L'élasticité des épithètes, dont l'interprétation est caractérisée par la relativité, est une propriété remarquable.

Comment circonscrire l'unité d'analyse et de traitement ? C'est une des questions les plus difficiles auxquelles j'ai été confrontée tout au long de mes recherches. Quelle est l'unité minimale que l'on manipule, interprète, traite, annote, extrait, propose en résultat, agrège avec d'autres ? Est-ce toujours et forcément la même pour ces différentes opérations ? Comment déterminer le niveau de grain adéquat (ou optimal) pour un problème donné ? L'unité peut correspondre à une marque ou à une construction, ou bien à un espace qui accueille une configuration de marques. La levée de certaines ambiguïtés s'opère souvent à l'intérieur des unités supérieures à la phrase. C'est notamment le cas pour certains prédicats potentiellement causaux (Jackiewicz 2004).

¹⁰ Remarquons que cette structure sérielle, située sur deux sections et balisée de manière irrégulière, ne pourra être identifiée en entier par un programme informatique.

Les marques sont souvent collectées sur corpus. Comment s'assurer alors que la procédure de collecte est appropriée ? Comment valider les données collectées ? La recherche des marques s'opère généralement en plusieurs étapes : observation de corpus, puis extrapolation et enrichissement à partir de dictionnaires, puis vérification sur d'autres corpus. Ce qui pose toutes sortes de difficultés en pratique. La question de la polysémie doit également être posée, et cela n'est pas toujours le cas. Les ensembles lexicaux réunis dans des champs homogènes créent l'illusion d'une stabilité de signification. Or on se rend souvent compte que la même classe peut contenir des éléments dont la palette des emplois est fort variable. C'est clairement le cas des adverbes de manière orientés vers le sujet (voir l'ouvrage original).

Les ressources linguistiques ne se réduisent pas à des lexiques, car les listes des marques seules ne suffisent pas. Il faut pouvoir leur associer des connaissances explicitant le fonctionnement de ces marques au sein de l'énoncé et souvent à l'échelle du texte. Plusieurs types de ressources linguistiques sont possibles et souvent nécessaires : corpus représentatifs (éventuellement annotés...), phrases « illustratives », arborescences notionnelles (catégorisations), constructions, patrons, collocations. Les données statistiques sur les fréquences d'apparition de ces différents éléments formels à l'intérieur des ensembles textuels constitués pour l'étude permettent d'apprécier l'importance potentielle des phénomènes étudiés. Les règles qui associent formes et valeurs, au sein de configurations syntaxiques précises, constituent une ressource plus élaborée, mais pas toujours suffisante pour une reconnaissance fiable. De plus, elles restent spécifiques aux observables qui ont permis leur élaboration. Toute forme de ressource doit donc être fidèlement documentée.

Pour stocker ces ressources hétérogènes, mêlant souvent faits et textes, nous avons besoin de modes / systèmes de représentation simples et robustes. C'est-à-dire transmissibles d'observateur à observateur avec un minimum de pertes, qu'ils permettent la reproduction des phénomènes et des raisonnements, qu'ils soient modifiables. Les supports les plus abordables et efficaces restent les feuilles des tableurs et les tables des bases de données relationnelles (Habert 2009).

6. Pour une capitalisation des données linguistiques

Les récentes mutations internes de la linguistique informatique, liées notamment à la possibilité d'explorer et d'étudier scientifiquement de vastes corpus écrits ou oraux, amènent progressivement les chercheurs à modifier leurs méthodes de travail. Ce nouveau contexte conduit les linguistes à rendre plus lisibles leurs pratiques, dans la mesure où la systématisation des observables, leur codage et analyse impliquent à minima un support externe qui garde en quelque sorte la mémoire du travail effectué. Les analyses ainsi produites et les ressources linguistiques constituées acquièrent ainsi une autonomie certaine vis-à-vis de leurs auteurs. Elles peuvent donc plus facilement être réexaminées à la lumière d'autres théories ou réutilisées dans d'autres applications.

Mais, malgré une évolution notable, la culture de la transmission des données, des analyses et des résultats n'est pas encore suffisamment développée en linguistique. La revue *Discours* a été pensée pour que les auteurs puissent associer aux articles des ressources, des corpus, des grilles de codages et des terminologies. En réalité, très peu de linguistes ont saisi cette opportunité. L'offre des index multiples, permettant une meilleure visibilité des contenus, n'est pas non plus réellement exploitée.

Il convient de nuancer ce constat en observant que tous les travaux linguistiques ne se prêtent pas d'égale manière à la capitalisation des données. Certaines études combinent plusieurs méthodes et s'appuient à différents stades d'analyse sur des observables aux statuts variés (fabriqués, attestés, bruts, minimaux, échantillonnés, exhaustifs...). Cela rend de toute évidence plus difficile la transmission des matériaux et l'explicitation des procédures d'analyse. Les propos de Catherine Schnedecker concernant les exemples qu'elle emploie dans l'étude des marques corrélatives sont symptomatiques de cette situation.

« Les exemples servant à l'argumentation ou à l'illustration ont une double origine. Tantôt ce sont des exemples fabriqués (par d'autres ou par nous) pour les besoins ponctuels d'une démonstration ou de tests visant à éclairer tel ou tel aspect du marqueur et supportant donc le 'minimalisme'. Tantôt ce sont des exemples authentiques, 'tout venant', collectés 'à la main' au gré de lectures diversifiées (presse, littérature, théorie linguistique) ou provenant de bases de données faciles d'accès (CDRom du Monde diplomatique et corpus ABU essentiellement). Il est évident que l'étude de configurations discursives ou des contraintes

imposées par un genre nécessite des supports textuels 'naturels' et de dimension conséquente. Les limites des corpus 'bricolés' sont bien connues (faible fiabilité, faible représentativité, bruit, etc.) Ils ont néanmoins l'avantage de faciliter l'analyse dite qualitative et de concilier les exigences parfois contradictoires d'études 'paroissiales' (pour reprendre une expression de G. Kleiber) et des études à 'gros grain' que constituent les approches discursives ». (Schneidecker 2006, p.11-12)

Discipline empirique, la linguistique a besoin d'ouillage approprié. Je pense que les réflexions des épistémologues, menées notamment sur la notion d'observatoire (Neveu 2004, p.206), devraient approfondir la question des pratiques expérimentales dans le travail linguistique, à la lumière des possibilités qu'offrent aujourd'hui les nouvelles technologies.

7. Vers une « ontologie » de l'énonciation

Les corpus et les différentes ressources linguistiques sont de plus en plus recherchés en linguistique et en informatique linguistique. Ces ressources sont parfois enrichies par des annotations linguistiques, comme des catégories morpho-syntaxiques, syntaxiques, sémantiques ou plus rarement discursives. Un des principaux problèmes reste néanmoins le manque d'interopérabilité entre des annotations effectuées par des personnes ou équipes différentes.

Il serait intéressant de pouvoir disposer de modèles formels permettant de représenter les caractéristiques énonciatives des textes. Il est évident qu'une telle « ontologie », entendue comme un ensemble structuré de concepts clés, ne peut être construite que progressivement et de manière participative. A terme, chaque utilisateur/contributeur concerné par les données énonciatives pourrait y « emprunter » des concepts, se positionner par rapports à ceux qu'il jugerait imparfaitement définis, en proposer de nouveaux, à la lumière de ses propres travaux.

Une telle ontologie pourrait trouver son utilité dans différentes tâches : annotation de textes, indexation de ressources linguistiques, interopérabilité entre descriptions linguistiques, aide à l'enseignement de la linguistique... Par ce moyen, il deviendrait possible par exemple de représenter de manière schématique la dynamique énonciative d'un discours. Cela permettrait également d'annoter différents contenus textuels, en tenant compte de leur inscription subjective.

Quelques exemples, volontairement simples, peuvent illustrer l'idée sous-jacente. Ainsi, dans les extraits (1-3) issus du *Monde*, il serait intéressant de pouvoir mettre en évidence par des annotations adéquates : les actes de paroles réalisés (*déclaration, affirmation, mise en garde, question, démenti...*), les discours rapportés reformulés par le journaliste (*en clair : ...*), les jugements assumés par le journaliste sur les paroles rapportées, en précisant l'objet du jugement et sa polarité ; dans le cas d'un discours rapporté, celui-ci peut porter sur la forme d'expression (*lapalissade, amateur de litote, fausse question...*), le contenu, rapport interpersonnel établi par ce discours rapporté (*mépris fugace des convenances*). Sans oublier l'identité, si elle est exprimée, des différentes instances de parole (auteur de l'article, *Conseil européen, un ministre, Sarkozy...*), la date (*18 juin*), les lieux de leurs interventions, etc.

(1) Une **déclaration** du Conseil européen du 18 juin **stipule** que, « si, à l'issue d'un délai de deux ans à compter de la signature du traité établissant la Constitution, les quatre cinquièmes des Etats membres ont ratifié ledit traité et qu'un ou plusieurs Etats membres ont rencontré des difficultés pour procéder à ladite ratification, le Conseil européen se saisit de la question ». Cette **affirmation** procède à la fois de la **lapalissade** mais aussi de la **mise en garde** : *'l'Europe des 25 continuera à aller de l'avant, quitte à ce que ce nombre soit réduit à une vingtaine.* (Le Monde, 23 juin 2004)

(2) « Il faut examiner la question de l'ouverture des négociations avec la Turquie avec toute la sérénité possible », **résume** un ministre **amateur de litote**. **En clair** : *'l'élargissement de l'Union à vingt-cinq Etats membres va déjà être difficile à faire accepter aux opinions publiques ; le fait d'ajouter la " question turque " serait le plus sûr moyen de précipiter un échec.* (Le Monde, 11 décembre 2002)

(3) La présence, au premier rang, du préfet Philippe Massoni, secrétaire général du Conseil de sécurité intérieure (CSI) et conseiller du chef de l'Etat, lui inspira cette **fausse question en forme de démenti** : « S'il y avait un malaise [avec l'Elysée], pensez-vous qu'il serait à mes côtés ? ». Puis, **poussant l'avantage**, M. Sarkozy **confia, l'air de rien**, son habitude de « demander au CSI, présidé par le président de la République, de prendre des orientations », **au mépris fugace des convenances** protocolaires... (Le Monde, 16 janvier 2004)

L'amorce d'un tel travail a été réalisée en proposant une esquisse d'ontologie linguistique de l'énonciation textuelle OntoDiscours (Jackiewicz et Laublet 2007). Issue d'une démarche originale par rapport à l'état de l'art des ontologies de la linguistique existantes, ce formalisme permet d'annoter des segments textuels conjointement ou non avec une ontologie d'un domaine. Le démonstrateur associé a été construit sur Protégé et Knowtator. Nos travaux actuels sur la description linguistique et la modélisation des jugements d'évaluation, menés au sein du projet OntOpiTex, se situent dans le même courant. Une première expérience consistant à utiliser conjointement une taxonomie des attitudes tirées de la théorie *Appraisal* (Martin and White 2005) et une ontologie du domaine a été présentée dans (Laublet *et al.* 2010).

Rappelons très brièvement que des propositions d'ontologies pour la linguistique existent déjà. Une des plus connues est GOLD (Farrar and Langendoen 2003) qui s'appuie sur l'ontologie générique SUMO. Cette ontologie a été conçue pour l'annotation de ressources linguistiques provenant de langues en voie de disparition. Elle couvre essentiellement la morphosyntaxe et des aspects grammaticaux. Etant donné que des différences sont inévitables dans les catégorisations utilisées par différents linguistes, (Farrar and Lewis 2005) présentent sous le nom de « communauté de pratique », un cadre de travail fondé sur GOLD. Celui-ci permet d'adapter ou plutôt d'enrichir l'ontologie pour différentes communautés qui se consacrent à telle langue ou qui défendent telle ou telle théorie linguistique. GOLD devient un noyau ontologique qui capture les fondamentaux de la linguistique descriptive, tout en restant ouverte aux adaptations et enrichissements. De leur côté, (Goecke *et al.* 2005) et (Bärenfänger *et al.* 2006) ont proposé des ontologies destinées à modéliser des relations entre unités textuelles (anaphoriques, rhétoriques...).

REGARDS SUR LA SUBJECTIVITÉ, AU TRAVERS DES...

Au croisement de la linguistique du discours, des approches de l'énonciation et de la linguistique informatique, quatre études composent un ensemble que j'ai voulu synoptique de la totalité des travaux de recherche que j'ai effectués jusqu'à présent. Chaque chapitre comprend deux parties : (i) un état des lieux qui donne un aperçu du contexte et de la problématique et qui résume mes recherches sur le sujet, (ii) une présentation abrégée d'un travail jugé le plus significatif. L'ordre de présentation correspond à la progression temporelle des travaux réalisés.

CONNAISSANCES CAUSALES - Le langage permet de communiquer des représentations, qui sont toujours énoncées selon un mode de validation particulier, comme vraies, douteuses, possibles, nécessaires, empruntées à d'autres, etc. Je me suis attachée à montrer dans ma thèse que la notion de causalité, intuitive mais hautement polythétique, peut être approchée par le biais de sa dimension subjective. Les différentes modalités de prise en charge que l'on atteste dans des énoncés causaux rendent compte des représentations de la notion de causalité, que l'énonciateur se fait ou construit et choisit de transmettre. Etant à divers degrés liées conceptuellement et formellement à la notion de causalité, ces formes modales constituent des indices précieux pour son identification et son interprétation dans les textes. Les différentes implémentations que j'ai réalisées au LaLIC (durant ma thèse et les premières années qui l'ont suivie, au sein des systèmes SERAPHIN, SAFIR et ContextO) reposent sur l'exploitation conjointe des indices « directs » de causalité et des marqueurs de prise en charge énonciative.

ORGANISATIONS TEXTUELLES - La subjectivité est inhérente à l'exercice même du « langage qui contient toujours les formes linguistiques appropriées à son expression » (Benveniste). Je me suis intéressée à des énoncés qui contiennent des instructions de lecture guidant le lecteur dans l'interprétation des contenus complexes. Dans le cas de la sériation que j'ai pris pour objet d'analyse, la structure du discours épouse les formes d'organisation que l'énonciateur perçoit dans (ou donne à) la réalité dénotée par son discours. L'étude se situe dans la lignée des travaux sur les cadres de discours, mais concerne de fait les recherches sur les modes d'organisation des discours et leur signalisation. Elle a croisé deux approches : (i) un travail sur les cadres organisationnels et leurs introducteurs prototypiques (marqueurs d'intégration linéaire), donnant lieu à une implémentation dans ContextO et (ii) une étude plus générale de la sériation dans le discours. Les différentes formes attestées contribuant au signalement des structures sérielles ont été examinées dans leur aptitude à cumuler en contexte plusieurs fonctions discursives. Car, sur les opérations de nature purement organisationnelle se greffent souvent différentes relations sémantiques ou argumentatives s'établissant entre les contenus mis en série. Les résultats obtenus ont été exploités pour la construction d'un langage formel permettant de représenter les différents modes d'organisation au sein des textes écrits. Il s'agit du langage LangText, développé pour le système ContextO au sein du Lalic.

DISCOURS RAPPORTÉS - C'est avec le discours rapporté et les discours en circulation que l'on prend la mesure de la richesse et de la complexité des faits énonciatifs. Ma contribution dans ce domaine comprend trois volets : (i) analyse des rapports intersubjectifs et des formes de responsabilité et d'engagement propres aux pratiques journalistiques dans la presse d'information et d'opinion, (ii) formalisation ontologique des concepts énonciatifs pertinents pour la modélisation des discours médiatiques, (iii) analyse et repérage des discours politiques à fort impact. Dans un premier temps, je me suis attachée à construire une typologie des discours rapportés, fondée sur une série d'opérations énonciatives (restitution, authentification, interprétation, évaluation, explication), indexées elles-mêmes sur des rapports intersubjectifs entre différentes instances de parole impliquées ou concernées par ces discours. Je me suis intéressée de manière plus approfondie à l'étude des formes de responsabilité et d'engagement propres aux pratiques journalistiques dans la presse d'information et d'opinion. A partir de ces travaux, j'ai participé, dans une optique plus formelle, à l'élaboration d'une esquisse d'ontologie linguistique de l'énonciation

textuelle OntoDiscours. Ses différentes classes de spécialisation permettent d'annoter finement : le niveau énonciatif (direct, rapporté), les actants et les circonstants, les constituants fonctionnels, les constituants modaux, les éléments paratextuels, ainsi que les marques de surface de ces différents objets. Enfin, plus récemment, j'ai travaillé à la mise en place d'une méthode itérative d'analyse et repérage d'énoncés à fort impact. Cette recherche vise la construction d'outils informatiques permettant de répondre à deux besoins clairement identifiés aujourd'hui : (i) identifier des discours en circulation (repris, commentés, détournés...), (ii) évaluer et analyser l'impact d'une citation particulière. L'étude de faisabilité et le prototype associé ont été réalisés avec la plate-forme SemioLabs développée par la société Noopsis.

JUGEMENTS ÉVALUATIFS - Parmi les manifestations de la subjectivité langagière, les plus évidentes et plus clairement avouées sont celles qui sous-tendent des jugements d'opinion ou de valeur. Étonnamment, cette problématique a été peu explorée en linguistique. Elle se trouve aujourd'hui fortement motivée par les multiples enjeux sociétaux de la veille d'opinion. C'est dans ce contexte que se situent mes travaux récents sur la mise en discours de l'acte évaluatif. Ils suivent pour l'essentiel deux axes. Le premier concerne l'étude des constituants détachés à caractère axiologique. Le deuxième se rapporte au travail collectif, mené dans le cadre du projet ANR OntOpiTex, sur la construction d'un modèle applicatif de l'évaluation.

Les recherches sur les segments détachés croisent d'une manière originale plusieurs approches. J'exploite les travaux sur l'apposition, les constructions détachées et les compléments circonstanciels (Neveu, Combettes, Molinier et Levrier...). Au plan sémantique, je m'appuie sur la modélisation des attitudes évaluatives proposée par la théorie *Appraisal* de Martin et White. Du point de vue discursif, j'observe les relations rhétoriques qui peuvent s'ancrer sur un jugement évaluatif. Sur un versant plus formel, je cherche à dégager à partir de corpus des patrons linguistiques sous-jacents aux constructions et aux expressions axiologiques les plus saillantes, ainsi qu'un lexique constitutif de ces expressions. Sur cette base, des expérimentations informatiques sur corpus plus vastes sont réalisées (avec LinguaStream et Nooj), afin de tester et d'amender les patrons, d'acquérir des lexiques et de collecter des observables. Ces ressources sont destinées à guider l'analyse sémantique du fait évaluatif dans les corpus applicatifs et préparent à l'implémentation d'outils d'analyse d'opinions.

L'étude sur les constituants détachés a fait émerger des structures récurrentes, au moyen desquelles il est possible de capter l'expression des qualités associées à des personnes ou des organisations, ce qui renseigne notamment sur leur visibilité (individuelle ou catégorielle), les facettes ou les composantes de leur réputation, les valeurs actualisés, etc. Le système appositif favorise des saisies partielles et multiples des êtres, focalisées sur des attributs saillants, socialement valorisés, parfois vus comme causalement efficaces des individus. Plusieurs patrons linguistiques ont donné lieu à des descriptions approfondies. Un important travail d'analyse qualitative et quantitative sur les adverbes « orientés vers le sujet » a ouvert un vaste champ d'observation, faisant apparaître notamment une importante polysémie des adverbes attestés, leur faible modalisation, des associations privilégiées et des routines d'expression. Ces connaissances donnent des indications précises sur les traitements à mettre en œuvre pour une identification plus fiable des jugements évaluatifs exprimés au moyen de ces adverbes.

I. CONNAISSANCES CAUSALES

A. Bref état des lieux

A l'origine de mes réflexions sur l'énonciation, il y a un premier ensemble de travaux que j'ai conduits sur des relations sémantiques dans la perspective de leur exploitation informatique, pour l'extraction et la modélisation des connaissances. Il s'agit de deux grandes familles de relations « structurantes » : relations parties - tout et relations causales. Identifier ces connaissances dans des textes nécessite au préalable de s'interroger sur la notion relationnelle en question, sur les modes d'expression linguistiques associées à cette notion, sans oublier les visées pragmatiques des scripteurs.

Dans un premier temps, mon travail sur la causalité a pris place dans le cadre d'un projet appliqué intitulé SERAPHIN, visant la construction automatique de résumé par sélection de phrases importantes (Berri *et al.* 1996 ; Jackiewicz 1999a ; Minel 2002). Ce premier système a évolué, en donnant plusieurs versions (SAFIR, CONTEXTO, EXCOM...), toutes construites au sein de l'équipe LaLIC. L'idée commune à ces systèmes (que l'on ne défend plus aujourd'hui) était de construire des procédures générales indépendantes des genres textuels, des domaines traités et des thématiques abordées. Le travail a porté sur l'identification des marques de surface permettant de statuer sur la saillance ou le degré d'importance des énoncés dans l'argumentation de l'auteur. L'analyse des instructions laissées par l'énonciateur dans les énoncés jugés importants a très vite révélé le besoin d'une systématisation de ces indices énonciatifs.

En effet, les discours contiennent des indices référant à la forme que prennent ses différentes unités, à leur mode d'articulation, au rôle que ces unités remplissent dans la structure globale du document. Annonces, commentaires, ajouts, reformulations, auto-corrections, conclusions informent les contenus, pour guider et faciliter leur réception. Il s'agit, pour reprendre les mots de (Authier-Revuz 1995), d'une « hétérogénéité montrée ». Cette voie, consistant en une exploration systématique de la dimension énonciative des textes, n'a pas été retenue dans les travaux appliqués au LaLIC.

La causalité étant une relation très fréquente dans des discours (sous différentes formes d'expression), il a fallu décider sur quels critères sélectionner ce type d'énoncés. Autrement dit, matérialiser par des éléments tangibles, les critères définitoires pour la reconnaissance des contenus causaux et les critères d'importance pour la sélection de certains d'entre eux. En réalité, deux objectifs ont donc été poursuivis conjointement : (i) construire une procédure efficace pour reconnaître des énoncés causaux, (ii) mettre en évidence des critères opératoires permettant de sélectionner les énoncés causaux les plus saillants (Jackiewicz 1996).

Cette approche, résolument opératoire et nécessairement textuelle, de la causalité m'a amenée à considérer, outre les formes d'expressions propres à la causalité, les modalités énonciatives présentes dans des énoncés causaux. Le propos de mon travail a été de questionner les rapports entre les connaissances causales et l'instance qui les énonce. Comment les moyens d'expression langagiers rendent-ils compte des représentations des relations causales qu'un locuteur humain se fait ou construit et choisit de transmettre ? Quelles modalités de prise en charge traduisent ces représentations ? Quelle est l'incidence de chacune d'entre elles sur l'information causale exprimée ?

J'ai constaté que l'analyse de l'incidence de ces modalités sur les connaissances causales exprimées permet de mieux cerner la nature même de la notion de causalité. Les discours que j'ai analysés présentent la causalité comme une relation *construite et assumée* par un agent cognitif (Jackiewicz 1999b). La causalité langagière apparaît comme une projection interprétative de l'esprit sur ce qui est perçu par les sens, et une construction dialogique traduisant différentes formes d'engagement et d'interaction face aux contenus exprimés.

Les différentes modalités de prise en charge énonciative ainsi mises en évidence, étant à divers degrés liées conceptuellement et formellement à la notion de causalité, constituent des indices précieux pour son identification et son interprétation dans les textes. Cette caractéristique des énoncés causaux aide également à traiter la polysémie de nombreux prédicats qui sont employés pour exprimer des liens de causalité.

B. La causalité prise en charge

1. Problématique

1.1 Pluralité de points de vue dans l'étude de la causalité

Les connaissances causales sont présentes dans nos raisonnements quotidiens, comme dans quasiment tous les domaines de connaissance (scientifique, technique, économique...). Leur rôle dans la structuration de l'expérience humaine, permettant à la fois l'intelligibilité et la maîtrise des phénomènes, les rend indispensables et précieuses dans quasiment tous les domaines d'activité. La causalité fait partie (aux côtés des relations temporelles, entre autres) des relations conceptuelles organisatrices des connaissances qui sous-tendent la perception des processus, des événements et des actions humaines. Ces relations sont fondamentales au sens où il n'est pas d'activité qui, d'une manière ou d'une autre, ne les mette en jeu.

Pourtant, qu'elle fasse l'objet d'une simple croyance, d'une connaissance commune ou d'une connaissance scientifique, la causalité est toujours à la fois intuitive et impossible à saisir dans son essence. Tour à tour, objet de recherche et outil d'investigation, la causalité intéresse de nombreuses disciplines (physique, biologie, médecine, philosophie, sociologie, psychologie, linguistique, intelligence artificielle...), qui l'approchent et se l'approprient avec la terminologie et les modèles qui leur sont propres.

La causalité est omniprésente dans les discours. Les schèmes causaux, comme le montrent de nombreuses études de psycholinguistique (Coirier *et al.* 1996), organisent des narrations, sous-tendent des explications, fondent explicitement ou implicitement des enchaînements argumentatifs. Pour autant, l'approche langagière de la causalité n'est pas aisée.

Admettant des acceptions diverses et d'importantes nuances de signification, elle traverse de nombreuses catégories grammaticales, d'où une grande diversité de procédés linguistiques susceptibles de l'exprimer (Gross et Nazarenko 2004). Plusieurs typologies des causes ont été proposées dans la littérature linguistique. Certaines sont issues des études portant sur des marqueurs (tels que les connecteurs *parce que*, *puisque*, *car...*) (Groupe λ -I 1975), d'autres résultent de l'analyse d'une notion grammaticale particulière (l'agentivité, la transitivité sémantique) (François 1988), (Desclés et Guentchéva 1998), d'autres encore s'inscrivent dans une perspective plus cognitive tournée vers la recherche de primitives sémantiques ou d'universaux langagiers (McCawley 1976), (Wierzbicka 1997), (Talmy 1988), (Desclés 1990), (Jackendoff 1990), (Pottier 2001).

Les catégorisations de la causalité construites à partir de l'étude de ses marqueurs sont essentiellement de trois types, selon le composant de la relation pris en considération. Citons en premier lieu les catégorisations qui s'attachent à caractériser l'action causale efficiente, en proposant des distinctions comme « causation directe - causation indirecte », « causation par contact - causation sans contact », « causation discursive - causation non discursive » (Nedjalkov and Silnickij 1969), (Kordi 1988), (Jackendoff 1990), (Pustejovsky 1995). D'autres typologies cherchent essentiellement à différencier et à structurer les différents effets possibles d'une action causale (« causer – gêner - laisser-faire » ou « création – croissance - maintenance – décroissance – arrêt » (Talmy 1988), (Garcia 1998), (Park and Abraham 2001). Enfin, certaines valeurs de causalité peuvent être dégagées en caractérisant les situations mises en relation (état, processus, événement...) (Gross 1996). En pratique, les typologies mêlent souvent des informations de nature différente.

Sur le plan de l'argumentation, comme l'ont montré différents auteurs (Plantin 1990), (Miéville 1992), (Perelman and Olbrechts-Tyteca 1992), (Jackiewicz 1999a, 1999b), la causalité peut être engagée dans plusieurs sortes d'actions discursives dont certaines cherchent à établir l'existence d'une relation causale nouvelle (acte d'explication), d'autres exploitent une relation causale déjà établie à des fins d'étayage (acte de justification) ou d'inférence (relation d'argument à conclusion).

Il est important de noter que la langue différencie très nettement ces différents emplois, en réservant les moyens grammaticaux (par exemple les connecteurs) à l'expression des enchaînements discursifs fondés sur des liens causaux établis, notamment sur des lois causales

communément partagées¹¹. Elle fait appel à des moyens lexicaux (essentiellement verbaux), plus riches en possibilités d'expression notamment sur le plan des modalités, quand il s'agit de construire une telle relation ou d'en parler. Cette distinction permet de comprendre pourquoi, bien que focalisées sur la même notion, les analyses linguistiques menées sur le lexique et celles qui s'intéressent aux connecteurs divergent considérablement tant dans leurs méthodes, que dans les résultats produits.

1.2 Analyser le point de vue de l'énonciateur

Si les multiples points de vue que la linguistique adopte par rapport à la causalité contribuent indéniablement à mieux cerner cette notion complexe, cette diversité n'aide pas (ou au moins pas directement) à élaborer une vision unifiée de la causalité dans la langue. Il est difficile, voire impossible, d'en tirer un appui théorique solide pour les traitements informatiques des où la causalité est présente. Quelle théorie, quelle typologie, quels marqueurs choisir pour pouvoir filtrer, interpréter et exploiter les données causales déposées dans des textes ?

La réflexion livrée ici est directement issue de plusieurs expériences menées au sein de l'équipe LaLIC pour (i) l'élaboration automatique de résumé, (Jackiewicz 1998), (ii) la fouille sémantique de textes (Minel *et al.* 2001, Minel 2002) et (iii) la recherche d'information sur le Web (Laublet *et al.* 2002). Leur point commun est d'adopter une approche textuelle, sur corpus¹². Les différentes composantes des textes complets (lexicale, grammaticale, énonciative, organisations discursive et physique...) sont analysées et exploitées conjointement pour déterminer les catégories de marqueurs linguistiques pertinents pour les problèmes à résoudre.

La spécialisation de fonction parmi les moyens d'expression de la cause que j'ai évoquée plus haut, est manifeste dans les textes où la problématique causale est explicite et centrale dans le propos de l'auteur. C'est le lexique qui est alors déployé, avec toute sa richesse sémantique, pour décrire et expliquer les liens ou les mécanismes causaux survenant dans et entre les phénomènes. L'analyse sémantique de ces marqueurs conduit à élargir la vision habituelle de la causalité, tributaire du schéma canonique de la cause qui produit et explique son effet, et focalisée sur l'analyse des modes de production causale. D'une part, les explications causales que nous trouvons dans les textes ne présentent que rarement les mécanismes réels de l'efficience causale. Les phénomènes décrits et analysés dans les textes sont généralement fort complexes, aussi les liens de dépendance qui les caractérisent, qu'ils soient statistiques, partiels, indirects ou indéterminés contribuent-ils à augmenter l'intelligibilité de ces phénomènes. D'autre part, l'auteur qui communique une relation causale dit, dans la plupart des cas, bien plus que l'existence de cette relation. Les traces de sa présence subjective sont lisibles à travers de nombreuses formes d'expression, notamment quand il s'efforce de présenter une relation causale comme plausible ou comme intelligible, ou quand il affirme établir par lui-même le lien causal qu'il énonce. Cette empreinte de l'énonciateur peut porter plus amplement sur des segments de texte plus vastes ; de véritables cadres énonciatifs enferment alors l'information causale spécifiant l'origine de cette information (*selon X...*) et précisant le degré d'adhésion à son égard de la part des différents locuteurs dont la parole est convoquée dans le texte (*X prétend que...*).

2. Une lecture en contexte

Etudier les moyens d'expression de la causalité dans l'optique d'un traitement automatique amène à aborder la question de la polysémie de ses marqueurs, et, par ce biais, celle des liens conceptuels que la causalité entretient avec les notions avec lesquelles elle « partage » les marqueurs.

¹¹ On peut se demander si certains connecteurs avant d'être des marqueurs de cause ne seraient pas en premier lieu des marqueurs d'argumentation permettant d'exprimer des enchaînements discursifs, fondés, plus ou moins souvent selon le cas, sur des relations causales, ces dernières étant exploitées comme garants d'enchaînement. Même le plus causal des connecteurs *parce que* permet de construire des séquences argumentatives fondées sur d'autres types de relation, notamment sur une relation définitoire.

¹² Les corpus textuels qui ont été utilisés dans ces études pour l'analyse de la causalité sont : (i) 30 textes techniques et scientifiques EDF sur l'énergie nucléaire, (ii) 100 textes de vulgarisation scientifique portant entre autres sur la migraine, (iii) Le Monde diplomatique (1984-1998) sur CD-ROM, (iv) 1,2 Go de textes sur les catastrophes naturelles issus du Web.

En français, plusieurs centaines de verbes dits causatifs (*favoriser, gêner, empêcher...*) peuvent exprimer selon les cas, soit une action agentive (intentionnelle), soit une relation causale entre faits ou situations. Le recours au cotexte est alors nécessaire pour résoudre l'ambiguïté¹³. Parmi les critères qui interviennent dans la caractérisation distinctive de la notion de causalité par rapport à celle d'agentivité, les modalités méritent une attention toute particulière. Il s'agit d'une part des modalités d'action qui spécifient le degré d'implication de l'agent humain dans l'action qu'il effectue, et d'autre part des modalités de prise en charge qui rendent compte de la position que le locuteur entretient avec le contenu énoncé.

Les modalités d'action sont la trace d'une description intentionnelle, normative, globale ou unifiante de l'action humaine (Ricoeur 1990). En effet, il existe dans la langue une multiplicité d'expressions qui modulent la signification de l'action du point de vue : (i) du degré d'intentionnalité (on peut dire qu'on a fait quelque chose *volontiers, intentionnellement, exprès, dans le but de, ...*), (ii) du contrôle (*commencer à, s'appliquer à, interrompre, maintenir, terminer de, ...* revient à dominer à degrés différents son propre « faire »), (iii) de l'investissement intellectuel (dire qu'un acte a été fait *judicieusement, systématiquement, habilement, laborieusement, intelligemment* l'élève au rang d'une action), (iv) de la norme en vigueur (une action peut être *louable, blâmable, permise, interdite, obligatoire, conseillée...*). L'agent et le résultat de son action font partie d'une seule et même situation. En effet, les capacités d'exécution directe, de contrôle et de téléonomie qu'il met en œuvre dans l'action (Desclés et Guentchéva 1998), sont autant de propriétés qui rendent son acte difficilement séparable de l'effet produit.

(a) *Intentionnellement, systématiquement, les architectes blancs de la politique urbaine ont fait de Chicago une ville où les quartiers se succèdent sans se toucher, où la ségrégation raciale et sociale débouche sur une réalité de quasi-apartheid. (Le Monde diplomatique, avril 94, pp.18-19)*

Les modalités de prise en charge, quant à elles, focalisent l'attention sur la plausibilité et sur l'intelligibilité du lien causal établi entre deux situations distinctes, que l'on peut appréhender indépendamment l'une de l'autre. Elles font apparaître un degré d'incertitude et d'indétermination qui caractérisent souvent (voire intrinsèquement) les connaissances causales, et, de manière corollaire, la nécessité de recourir à des éléments fiables (enquêtes, calculs statistiques, expérimentation...) pour fonder leur existence. Ainsi, l'expression de la causalité peut s'accompagner de toutes sortes de modalités présentant le lien causal comme possible, nécessaire, incertain, constaté par observation, récurrent, dépendant du contexte...

(b) *Une analyse effectuée sur les enfants nés entre 1956 et 1960 par Quang Chi Dinh, de l'INSEE, montrait que la mortalité infantile est fonction de la place occupée par le père dans la hiérarchie des catégories socioprofessionnelles. (La Recherche n°261, jan.94)*¹⁴

La prise en charge énonciative de la causalité constitue à nos yeux l'originalité de cette relation et la distingue de plusieurs autres relations langagières structurantes telles que la temporalité ou l'ingrédience (relation partie-tout). Les enchaînements temporels par exemple ne sont pas soumis aux mêmes questionnements liés à l'interprétation de ce qui est perçu et ne sont que rarement modalisés ou insérés dans des cadres de validité restreints. On peut remarquer au passage que les moyens lexicaux permettant d'exprimer des relations temporelles (*précéder, succéder, suivre...*) sont très limités en comparaison de l'éventail des centaines de verbes disponibles pour qualifier un type particulier d'efficience causale. Par exemple, pour exprimer l'idée que « X augmente/ fait augmenter Y », il existe en français plusieurs dizaines de verbes nuancés cette action causale selon notamment la nature de l'objet qui subit ce type de modification (*renforcer, accroître, agrandir, étendre, amplifier, hausser, approfondir, élargir, grossir, allonger, prolonger, intensifier, stimuler, précipiter, redoubler, accélérer, fortifier, alourdir, accentuer, revigorer, durcir, attiser, envenimer, affoler, exciter, exalter, exacerber...*).

Il est à noter par ailleurs que la problématique causale elle-même est souvent annoncée dans des textes à travers une variété d'expressions qui réfèrent dans leur majorité à une construction intellectuelle : *imputation causale, impression de causalité, suspicion de causalité, interprétation causale, assignation causale, assertion du lien causal, attribution causale, jugement de causalité, degré de certitude du lien causal, détection de causalité, établissement du lien causal...* Toutes ces expressions, qui n'ont pas d'équivalent dans la langue relativement à d'autres relations

¹³ Il s'agit dans la majorité des cas de l'ambiguïté pour la machine et non pour l'interprétant humain.

¹⁴ Dans les énoncés cités en exemple, les prédicats exprimant la causalité sont mis en gras, le soulignement indique les indices de prise en charge énonciative.

structurantes (temporelle, partie-tout, appartenance ensembliste, localisation), évoquent clairement la présence de l'énonciateur qui ne se contente pas de « lire » le monde qui l'entoure, mais cherche activement à le rendre intelligible à partir de données et d'outils dont il dispose.

3. Prendre en charge une connaissance causale

3.1 Notion de prise en charge

La prise en charge énonciative est une notion linguistique fondamentale qui représente la dimension modale (*modus*) emboîtant la dimension prédicative (*dictum*) de l'énoncé. Depuis Charles Bally (Bally 1932), plusieurs linguistes, dont Emile Benveniste (Benveniste 1974) (*discours / récit*) et Antoine Culioli (Culioli 1990) (notion de *lexis* soumise aux modulations énonciatives et appréciatives), ont cherché à montrer l'importance que revêt la dimension énonciative dans la construction du sens des énoncés. (Desclés et Guentchéva 2000) proposent d'analyser la prise en charge énonciative en plusieurs opérations élémentaires, paraphrasables comme suit :

- je dis que ce que je dis est vrai (assertion),
- je dis que ce que je dis est vrai selon une certaine modalité (possible, nécessaire...),
- je dis que ce que je dis est dit par un autre, déterminé (énonciation rapportée) ou indéterminé (médiation),
- je dis que ce qui est dit est vrai dans un contexte déterminé,
- je dis que je constate une situation (événement ou état) et alors je dis que c'est vrai,
- je dis que ce que je dis est inféré par abduction à partir d'indices,
- je dis que ce que je dis est attesté par un témoin (vu ou entendu)...

Certaines langues permettent d'explicitement de nombreuses nuances de prise en charge, notamment en grammaticalisant la catégorie du médiatif (Guentchéva 1997). En français, la prise en charge énonciative est peu grammaticalisée. Les indices linguistiques de certaines de ses valeurs se trouvent dissimulés sous diverses formes d'encodage, et souvent dans des contextes relativement larges. De plus, plusieurs opérations langagières peuvent être intriquées et plusieurs modalités différentes peuvent jouer en même temps.

La grille des valeurs qui suit, inspirée de la typologie générale dégagée par (Desclés et Guentchéva 2000), s'applique spécifiquement aux énoncés causaux pris en charge tels qu'ils s'expriment en français. Ainsi, aux côtés de certaines valeurs modales très générales pouvant modifier n'importe quel contenu prédicatif, on trouve celles qui sont propres à la causalité, mais aussi des valeurs qui sans lui être réservées se retrouvent fréquemment dans des énoncés qui expriment cette notion.

3.2. Les assertions causales assumées par l'énonciateur

L'énonciateur produit souvent des assertions de relations causales qui sont pleinement assumées par lui. Cette prise en charge peut prendre plusieurs nuances, selon la manière dont l'énonciateur choisit de légitimer le savoir causal qu'il exprime. L'assertion peut être fondée sur un constat empirique direct, lié à la perception. Elle peut puiser sa légitimité dans le fait d'être assumée collectivement par une communauté déterminée. Elle peut aussi s'appuyer sur le fait que l'énonciateur tient le contenu prédicatif pour explicable ou intelligible. L'adhésion au contenu prédicatif atteint son degré maximal quand l'énonciateur déclare être lui-même à l'origine de la relation prédicative exprimée.

3.2.1 Assertion simple

L'énonciateur prend en charge la relation causale prédiquée : « je dis que ce qui est dit est vrai » ou, d'une manière plus nuancée, « je dis que ce qui est dit est vrai selon moi ». La relation prédicative assertée est en général déjà constituée et l'énonciateur la reprend à son compte. La prise en charge peut être marquée explicitement (1), mais le plus souvent elle reste implicite.

(1) A cette question **je répondrai que** le façonnement précoce des émotions **a imprégné** dans l'enfant un tempérament, un style comportemental qui lui **a permis** lors de l'épreuve **de** puiser dans ses ressources. (B. Cyrulnik, *Les vilains petits canards*, 2001, p.27)

3.2.2 Assertion de la prédication

L'énonciateur asserte explicitement la constitution de la relation prédicative : « je dis que ce qui est dit est établi par moi-même ». Il est à l'origine de la relation prédicative et son acte de prédication

témoigne de l'adhésion totale au contenu asserté. Celui qui parle crée le lien causal et le signe dans le même acte de parole. Ce type d'assertion est propre aux énoncés causaux.

Plusieurs marqueurs expriment cette double opération de prédication et de prise en charge caractérisant la causalité langagière. Tout d'abord, quelques verbes tels que *attribuer* à, *assigner* à, *imputer* à, *accuser de* intégrés dans des structures du type *j'accuse [] de [], j'assigne la responsabilité de [] à []*. Ensuite, de nombreuses expressions telles que *désigner comme cause de ; établir une relation de cause à effet entre [] et [] ; montrer l'existence d'un rapport causal entre [] et [] ; faire un rapprochement entre [] et [] ; relier [] et [] ; mettre en relation [] et [] ; relier dans un rapport de cause à effet [] et []*... précédées par les pronoms *je* ou *nous*. Enfin, notons que l'action de construction de ce rapport causal exprimée par les prédicats que nous venons de citer peut être présentée comme *raisonnable, logique, claire, arbitraire...* (*on peut raisonnablement assigner les causes de [] à []*).

(2) *De telles difficultés m'ont alors semblé révélatrices d'une faiblesse de notre démocratie dont **j'attribue la cause à une politique constante d'éducation nationale qui mériterait d'être corrigée.** (E. Dufourcq, « Ethique des sciences et langage de la décision » <http://en.science-generation.com>)*

3.2.3 Assertion présentant le contenu prédicatif comme explicable ou intelligible

L'énonciateur assume la relation causale en fondant son adhésion sur le caractère explicable ou intelligible du contenu prédicatif : « je dis que ce qui est dit est intelligible ou explicable selon moi ». Cette causalité s'exprime alors en termes de *raison* et d'*explication* et son existence est (ou doit pouvoir être) justifiée d'une manière satisfaisante pour l'esprit. Un glissement sémantique se produit fréquemment entre la proposition « la cause entraîne l'effet » et la proposition « la cause explique l'effet ». Pourtant fournir la *cause* n'est pas donner la *raison*¹⁵. Un lien causal peut être affirmé tout en ignorant la façon dont la cause produit l'effet. Les *causes* ne deviennent des *raisons* qu'une fois décrites et expliquées. En effet, une question introduite par *pour quelle raison* semble exprimer plus fortement le désir de comprendre que quand elle commence par *pourquoi*. L'emploi du terme *raison* (dans *être la raison de, avoir pour raison...*) présuppose l'existence de la cause. S'il s'agit habituellement de *comprendre, d'analyser, d'examiner, de décortiquer* ou de *disséquer* les raisons, on *recherche, on trouve* et on *assigne* les causes. Les raisons, plus souvent que les causes, sont *profondes, véritables, évidentes, compréhensibles*. Les causes, plus souvent que les raisons, sont *probables, plausibles, hypothétiques, vraisemblables*. Ce sont les *raisons* qui fournissent l'*explication (suffisante, satisfaisante, convaincante...)* des phénomènes, elles sont à la fois ce qui les fait être et ce qui permet de les comprendre.

Parmi les marqueurs qui présentent le contenu causal asserté comme intelligible ou explicable, on dispose d'une famille d'expressions telles que *s'expliquer par (raison) ; explication de [] être (que) [] ; explication de [] être recherchée dans [] ; trouver son explication dans ; donner l'intelligence ; rendre raison de ; avoir pour raison ; être raison de ; là réside raison de ; être la raison qui détermine ; rendre intelligible...* accompagnés d'indices contextuels (adjectifs, verbes) tels que cités plus haut.

(3) *J'ai tâché d'**expliquer** ce type d'inceste **par** le jeu des catégories de l'identique et du différent que je pose comme principielles. (F. Héritier, *De l'inceste*, p.10, 1994)*

3.2.4 Assertion causale collective

L'assertion de la causalité s'appuie sur un savoir collectif large ou propre à une communauté donnée : « je dis que ce qui est dit par la communauté (tout le monde) est vrai ». L'énonciateur s'identifie entièrement à la communauté qui légitime ce savoir. Tombent sous cette catégorie les lois causales qui relèvent du sens commun telles que « *la chaleur fait fondre la glace* » ou « *fumer provoque des maladies graves* », ainsi que certaines lois causales établies et partagées dans des communautés plus restreintes ou des domaines spécialisés.

Quelques expressions permettent en français de signifier cette prise en charge collective on (tout le monde, chacun...) sait que ; il est admis aujourd'hui que ; tout le monde s'accorde à le dire (en incise) ; tout le monde s'accorde à reconnaître que ; il n'est nullement besoin de rappeler que...

(4) *Les points de vue concordent apparemment sur les causes du ralentissement des exportations américaines. Au-delà du bouc émissaire que représente la CEE, tout le monde s'accorde à reconnaître*

¹⁵ Le mot *raison* étant polysémique nous précisons que cette discussion concerne uniquement les énoncés causaux (et non les énoncés agentifs).

les effets de la hausse du dollar, tant dans la perte de compétitivité des exportations américaines que dans le ralentissement de la demande globale, et prône des mesures visant à rétablir cette compétitivité. (LMD, nov.89, p.28)

3.2.5 Constat direct d'un « état de choses »

L'adhésion au savoir causal peut s'appuyer sur la perception visuelle directe (en situation) d'une certaine interaction causale « je dis que je constate en ce moment ce qui est dit » ou « je dis que j'ai constaté ce qui est dit ». Selon les cas, plusieurs significations précises peuvent être associées au constat d'une relation de nature causale. Le plus souvent, il s'agit de plusieurs observations d'enchaînements d'événements similaires conduisant à la mise en évidence d'une corrélation (qui n'est pas encore à proprement parler une relation causale). L'observation est une manière de valider un savoir causal, fondée sur une certaine implication personnelle de l'énonciateur dans la situation de causation. Elle est particulièrement intéressante quand elle peut être réitérée.

Les verbes *constater, remarquer, observer, noter, apercevoir*, ainsi que les nominalisations qui en sont issues (*observation, constat...*) fournissent les moyens d'expression de ce type d'assertion.

(5) Les observations que j'ai pu faire et refaire avec mes étudiants nous ont beaucoup appris : les goélands perçoivent de véritables organisations sonores qui **provoquent** des comportements différents : cris d'appels, cris de triomphe, cris d'alarme, cris de parade. (B. Cyrulnik, *La naissance du sens*, 1995, p.29)

3.3 Les assertions causales contextualisées

Le savoir causal est souvent asserté comme dépendant d'un point de vue ou d'un contexte particulier, qui déterminent ainsi son cadre de validité. Ce type d'assertion contextualisée est toujours assumé par l'énonciateur. Celui-ci assume directement non pas la relation causale, mais le fait de la rendre dépendante d'un contexte. Plusieurs cas de figure, relativement proches les uns des autres, peuvent se présenter selon les circonstances dans lesquelles le contenu prédiqué peut être considéré comme valide. Une connaissance causale peut être vraie uniquement à l'intérieur d'un cadre spatial ou temporel, au sein d'une discipline ou une théorie particulière, ou encore peut n'être tenue pour fondée que par certaines personnes ou par certains groupes de personnes (critère énonciatif). Le contexte de validité restreint permet de comprendre pourquoi plusieurs explications causales différentes (voire concurrentes) peuvent être proposées en même temps, chacune relativement à son contexte (relativité explicative). Les rédacteurs ont en général à cœur que leurs lecteurs reconnaissent correctement les limitations de validité du savoir transmis.

3.3.1 Assertion valide dans un cadre temporel ou spatial

Le contexte de validité d'une assertion causale peut être de nature temporelle et/ou spatiale : « je dis que ce qui est dit est vrai dans un contexte spatio-temporel déterminé ». Un savoir causal pouvant évoluer dans le temps, une explication causale donnée peut être valide uniquement à l'intérieur d'un intervalle temporel déterminé. L'efficacité causale d'un fait peut aussi se limiter spatialement. La signalisation de ce type de cadre passe par l'emploi d'expressions scéniques apparaissant en tête de phrase ou dans une incise *en France, en 1901, avant 1940, chez l'homme, chez l'animal...*

(6) L'Institut Rexecode prévoit qu'en France la consommation **sera, en 1993**, la principale **source de croissance** dans la mesure où les revenus réels des ménages devraient s'élever de 2 %, et où les impulsions ne sauraient provenir ni d'un marché international anémié ni d'un investissement déclinant, freiné par le coût élevé du crédit. (LMD, déc.92, p.32)

3.3.2 Assertion à l'intérieur d'un cadre de connaissance

Le cadre de validité d'une assertion causale contextualisée peut être de nature gnoséologique : « je dis que ce qui est dit est vrai pour une discipline ou une théorie donnée ». Le marquage des cadres de connaissance fait appel à des expressions scéniques apparaissant en initiale détachée ou au sein d'une incise, comme *en biologie, dans la théorie de*. Précisant le domaine, le cadre théorique ou le modèle de référence, ces cadres contribuent souvent par la même occasion à légitimer le savoir causal concerné.

(7) Dans la théorie standard, dite néoclassique, le libre fonctionnement du marché **conduit** l'économie dans son ensemble à un état optimal. (LMD, oct.96, p.26)

3.3.3 Assertion selon un point de vue

L'assertion causale selon un point de vue signale l'emprunt de l'information à un tiers « je dis que ce qui est dit est vrai selon X », sans que celui-ci soit nécessairement à l'origine de sa constitution (le cas décrit dans le § 3.3.4). Les marques de ce type d'assertion *selon X, d'après X, pour X, chez les X...* apparaissent le plus souvent en position initiale ou au sein d'une incise (8). L'auteur du texte assume le fait d'attribuer l'énonciation de la relation causale à un tiers, mais ne se prononce pas forcément sur son adhésion propre à cette information.

(8) *Pour Anaxadore, la détermination du sexe vient du père : les garçons **proviennent du** testicule droit, le plus chaud, les filles du gauche. Pour Empédocle, c'est la plus forte chaleur de la matrice, selon l'état du sang menstruel, qui **fait naître** un garçon ou une fille. Comme on le voit, dans les deux cas, c'est une plus forte chaleur qui **fait concevoir** un mâle. Aristote critique ses devanciers sur certains points, car ce n'est pas une mince affaire, dit-il, que de démontrer que la mise en présence du froid **déclenche** la production d'un utérus sur le fœtus, dans la matrice de la mère, bien qu'il conserve la vertu de l'opposition entre chaud et froid. (F. Héritier, Masculin / féminin, 1996, p.192)*

3.3.4 Assertion d'une prédication attribuée à un tiers

Quand l'énonciateur affirme « je dis que ce qui est dit a été établi par X » l'opération de création d'information est attribuée à un tiers. Deux niveaux de prise en charge peuvent être distingués. L'énonciateur assume directement le fait d'attribuer à un tiers la constitution d'une prédication (ici causale), mais ne se prononce pas sur son adhésion propre au contenu prédictif exprimé. L'énonciateur attribue la prédication à un tiers et fait connaître sa position par rapport à la connaissance prédiquée. Cette position peut être précisée entre autres au moyen des expressions telles que *très justement, avec raison, dans une analyse brillante...* pour marquer l'adhésion ou telles que *hâtivement, abusivement, à tort...* pour signifier une distanciation.

Les marqueurs qui expriment ce type d'assertion sont X accuse, X assigne, impute la responsabilité de [] à [], X établit un lien entre [] et [].

(9) *Ce parti, et tout particulièrement son chef, M. Jean-Marie Le Pen, **ont désigné, de manière démagogique,** les travailleurs immigrés **comme la cause principale des** difficultés des Français. (LMD, mars 97, p.1)*

3.4 Les médiations causales

La prise en charge médiatisée introduit un tiers plus ou moins déterminé. L'énonciateur prend de cette façon de la distance par rapport à ce qu'il énonce. Ne pouvant pas être légitimées par une origine reconnue, les relations causales ainsi médiatisées sont généralement incertaines et restent à vérifier.

3.4.1 Prise en charge par une tierce personne indéterminée ou reprise d'un « oui-dire »

Ce type de médiation met en scène un énonciateur indéterminé : « je dis qu'il y a quelqu'un qui a dit que ce qui est dit est vrai » ou « je dis qu'il y a quelqu'un qui a dit ce qui est dit et ce qui est dit est vrai ou pas ».

Pour exprimer un désengagement, l'énonciateur peut recourir au conditionnel, ainsi qu'à des expressions telles que *d'après, selon les rumeurs ; il paraît que, il paraîtrait que ; paraît-il* (en incise) ; *dit-on* (en incise) ; *le bruit court que*. La construction X est attribué à Y peut également rendre compte de la valeur en question.

(10) *Une enquête de l'Assistance publique de Paris révélait en mars 1993 que près de 10 % de malades étaient victimes d'une infection contractée durant leur séjour. D'après diverses sources, cette calamité **serait responsable de** six mille à dix-huit mille morts par an auxquels on peut ajouter les accidents thérapeutiques qui feraient quelque dix mille victimes selon l'Association des victimes d'actes médicaux (AVAM). (LMD, fév.95, p.21)*

3.5 Les énonciations causales modalisées

La présence des éléments modaux dans l'expression de la causalité n'a pas échappé aux linguistes, comme en témoignent notamment les propos de Ferdinand Brunot :

« Il est peu de rapports qu'on ait plus souvent besoin de mettre en relief. Plus le rapport de causalité entre deux faits est douteux, plus on peut avoir besoin d'y insister » (Brunot 1922, p.823)

« Le rapport de cause peut être lui-même dans toutes sortes de modalités ; on peut le présenter comme certain, possible, irréel. Pour assurer, on dira : c'est sûrement,

certainement... parce que. Pour présenter avec réserve au contraire, on se servira des moyens ordinaires : peut-être, sans doute : s'il est si timide, c'est peut-être manque d'habitude » (Brunot 1922, p.827).

(Talmy 1988) et (Kordi 1988) ont interrogé, chacun dans son propre cadre théorique, les liens entre la causalité et la catégorie grammaticale de modalité. Pour Leonard Talmy, les modalités expriment des rapports qui relèvent de la dynamique des forces (catégorie sémantique qui constitue la généralisation de la notion de cause). L'auteur montre que la notion de dynamique des forces s'exprime, autre autres, par les auxiliaires modaux (*John can / may / must / ought / should / would / need / dare / had better not leave the house*), en particulier quand il s'agit de rendre compte des interactions psychosociales entre entités sensibles. Selon Elena Kordi, inversement, il est possible de dégager dans la signification des verbes causatifs des composants modaux. Ainsi, les verbes tels que *contraindre, pousser, ordonner, prier, persuader, permettre, empêcher...* expriment une situation de causation dans laquelle sont engagés agent et patient (dotés de tendances modales propres, de volition, par exemple, susceptibles de changement).

Dans les lignes qui suivent, je discute de l'incidence des modalités caractéristiques des énoncés causaux choisis pour l'étude. Dans toutes ces différentes énonciations modalisées, la modalité porte sur la vérité du contenu prédicatif, ce qui peut être traduit par la formule « je dis que ce qui est dit est vrai selon une certaine modalité ».

3.5.1 Modalité ontique ou aléthique

Le contenu d'une proposition causale peut se trouver modifié par une idée de nécessité, d'impossibilité, de possibilité ou de contingence : « je dis que ce qui est dit est nécessaire ou possible ou contingent ou impossible ». L'idée de nécessité exprime le caractère de ce qui est rigoureusement déterminé, soit en raison de la nature des choses, soit en vertu des lois de la logique. Quant à la possibilité, si elle s'ancre également dans la nature des choses, sa valeur d'éventualité laisse la place aux variations individuelles, ce qui fait que l'on parlera de ce qui est faisable ou de ce qui peut se produire, plutôt que de ce qui se produira inmanquablement. La valeur de possibilité se marie naturellement avec la causalité dont le caractère est de toute évidence contingent. En effet, dans la vie de tous les jours les exemples abondent où une cause (ou un ensemble de causes) produit un effet non pas inéluctablement mais dans un certain pourcentage de cas (*fumer peut causer un cancer des poumons*). La nécessité appliquée à une relation causale semble, par contre, beaucoup plus discutable, car cette notion n'a pas vraiment sa place dans le monde fait d'événements singuliers, qui est un domaine de la probabilité et de l'*a posteriori*. Il semble donc qu'elle exprime davantage une nécessité d'ordre intellectuel qui, en introduisant un minimum de stabilité, aide à nous repérer dans un monde fait de phénomènes contingents et à en prédire de manière sûre certaines séquences.

Un pouvoir causal peut être présenté comme inhérent à un objet, s'assimilant ainsi à une propriété ou à une fonction de cet objet. Cette propriété (fonction) est à divers degrés possible ou nécessaire. Considérons l'énoncé *la tyramine est un facteur biochimique migrainogène*. La propriété causale en question sera considérée comme appartenant à la nature biochimique de la substance si le déclenchement de migraine a lieu avec suffisamment de régularité. Pour être plus près de la réalité, car nous savons que chez certains sujets migraineux, les aliments riches en tyramine ne déclenchent pas de crise, nous dirons *la tyramine susceptible de (peut) provoquer des crises de migraine*. Le rôle des modalités du possible et du nécessaire dans un énoncé causal est d'exprimer une efficience causale latente, plutôt réalisable que réalisée. Les énoncés causaux marqués par ce type de modalité constituent des énoncés génériques portant sur des types d'objets ou de situations. Sur le plan argumentatif, l'emploi des modalités aléthiques permet de renforcer la réalité du lien causal en le présentant comme inhérent aux choses.

Parmi les moyens d'expression linguistique, notons : *nécessairement ; fatalement ; forcément, inmanquablement ; infailliblement ; inévitablement ; inéluctablement ; inexorablement ; potentiellement...* ; *d'une manière potentielle ; en puissance ; par son essence même ; il est possible que ; il est nécessaire que...* ; *être doté de capacité de ; être susceptible de ; être capable de, avoir un pouvoir infectieux (anxiogène, fragilisant, psychogène, myogène, lactogène, pathogène, thermogène...)*...

(11) L'influence des circonstances externes peut, selon Buffon, **déterminer** un changement avantageux pour l'animal (allongement du poil par le froid, etc.); mais, d'une façon générale, le milieu ne commande

pas automatiquement le changement de forme qui favorise la vie dans ce milieu. (J. Rostand, Esquisse d'une histoire de la biologie, p. 62).

3.5.2 Modalité épistémique

Un énoncé causal modifié par une modalité épistémique rend compte d'un état de connaissance ou d'un degré de certitude de l'énonciateur par rapport à la réalité du lien causal prédiqué « je dis que ce qui est dit est certain ou probable ou improbable... ». Les modalités épistémiques marquent la prise en charge ou au contraire le refus de prise en charge de la vérité de ce qui est énoncé, avec tous les degrés possibles de distanciation par rapport à cette vérité *je suis sûr que, je sais que, je ne suis pas certain que, j'ai le sentiment que, je crois que, je doute que, je ne sais pas si...* On peut étendre la notion de modalité épistémique aux cas où il ne s'agit pas directement de l'attitude de l'énonciateur. Les phrases du type « X Verbe_d'opinion que p » (X *pense, est sûr, croit, s'imagine, ignore... que p*) permettant d'informer le destinataire des croyances d'un tiers. L'empreinte épistémique peut amener à interpréter une relation causale comme une construction opérée par l'esprit humain. Elle fournit alors une signature de l'assertion causale exprimée.

Les modalités épistémiques renvoient tant à des situations particulières, où elles participent à l'explication d'un événement unique, qu'à des situations génériques. Sur le plan argumentatif, les modalités épistémiques offrent de nombreux procédés permettant selon le cas d'affaiblir ou de renforcer la validité d'une assertion causale. Directement, par le choix du verbe : dire que « X *s'imagine, se figure que p* » revient à dire que « X *croit à tort que p* ». On peut également mettre en doute/renforcer la validité d'une assertion causale en la présentant comme une croyance (un savoir) de quelqu'un à qui on refuse/on accorde un capital d'autorité.

Une modalité épistémique peut être exprimée par certains adverbess (*assurément ; certainement ; probablement ; évidemment ; incontestablement ; indéniablement ; indiscutablement ; plausiblement ; vraisemblablement...*), plusieurs expressions impersonnelles (*il est vraisemblable que, il est plausible que ...*), un grand nombre de verbes en que (savoir que, penser que, croire que...), le modal pouvoir, etc.

(12) *On peut évidemment supposer que la baisse de mortalité due aux maladies cardiovasculaires **est liée** à une modification des comportements en matière de tabac, dans le domaine alimentaire et dans celui de l'activité physique. Une part, sans doute, doit être aussi **attribuée aux** thérapies médicales et notamment au suivi de l'hypertension. (La Recherche, n°261, janv.94)*

3.5.3 Modalité de fréquence temporelle et d'habitude

L'énonciation modalisée « je dis que ce qui est dit se réalise (est vrai) habituellement, en général, fréquemment... » met en évidence le caractère fréquent ou qui tient de l'habitude des faits prédiqués. Elle suppose un savoir sur des événements déjà accomplis et marque la possibilité d'extrapolation sur du « encore possible ».

Les adverbess qui portent sur des occasions ou des cas *toujours, parfois, quelquefois* ou exprimant l'idée de fréquence *fréquemment, de temps en temps, rarement, souvent* marquent, du point de vue modal, la possibilité de trouver plusieurs occurrences spécifiques d'événements, et cette possibilité, qui repose sur une expérience passée, vaut en principe, pour les événements futurs. En effet, la mention de l'itération et/ou de la fréquence d'apparition d'un phénomène même si elle n'autorise pas à faire de prédictions fiables pour l'avenir, appelle l'interrogation sur la possibilité de sa réapparition. Dans un contexte causal, c'est le jugement de valeur porté sur les effets d'un événement susceptible de se reproduire qui détermine l'intérêt d'une projection dans l'avenir. En disant qu'un fait est habituel, on affirme davantage que son caractère simplement répétitif. On pose que la répétition n'est pas accidentelle et qu'elle manifeste une disposition, une tendance. C'est donc la possibilité de généralisation permettant la prédiction qui constitue l'essentiel de ce type de modalité. Les modalités d'habitude donnent à une relation causale le statut d'une tendance ou d'une loi générale, ce qui revient à la présenter comme établie et acceptée. Les énoncés causaux ainsi modalisés sont susceptibles d'une lecture non événementielle. Les modalités d'habitude sont exprimées par les adverbess tels que *en général, d'habitude, généralement, habituellement, invariablement, ordinairement, traditionnellement, usuellement...*

Les modalités de quantification temporelle présentent souvent la relation causale comme le résultat d'une observation. L'approche temporelle de cette relation, fondée uniquement sur la succession régulière des phénomènes, sert à pallier la méconnaissance du mécanisme de l'action causale. De plus, elles s'appliquent à des phénomènes causaux dont on peut observer plusieurs occurrences

dans un temps relativement court. Elles ne concernent donc pas les événements uniques (comme le déclenchement de la Première Guerre Mondiale, par exemple). Les modalités temporelles et d'habitude présentes dans un énoncé causal traduisent dans certains cas la possibilité d'avoir pour un fait donné plusieurs causes possibles ou plusieurs conséquences possibles. Le phénomène d'échec scolaire, par exemple, est susceptible de recevoir plusieurs explications différentes.

(13) *Rappelons seulement que, sauf le cas assez rare d'une insuffisance intrinsèque de l'intelligence, l'insuccès scolaire **tient quelquefois dans** l'adaptation anxieuse à l'école. (L'anxiété et l'angoisse, Que sais-je ? n°1661)*

4. Double dimension de la causalité « langagière »

Les énoncés causaux que nous trouvons dans des textes ne constituent pas un simple calque des interactions entre des phénomènes du monde extra-linguistique. L'efficience causale n'étant généralement ni accessible à nos sens, ni directement intelligible, les rapports causaux sont souvent établis avec plus ou moins d'incertitude, à partir d'une investigation impliquant des observations, des calculs, des efforts de théorisation. L'attitude à l'égard des connaissances causales transmises, tout comme leur dépendance par rapport à un contexte ou à un point de vue particulier, ou encore la mention de la démarche qui a permis de les établir, font partie des connaissances qui sont souvent communiquées, car les relations causales dépendent étroitement des connaissances et des besoins en action de celui qui propose une explication. Il a été montré que des investigateurs inspirés par des motifs intellectuels ou praxéologiques distincts, isolent dans le même tissu factuel des causes distinctes (relativité explicative ou contraste explicatif) causale (Garfinkel 1981), (Largeault 1985), (Puech 1990), (Boudon 1990), (Plantin 1996), (Jacob 1997), (Nazarenko 2000).

La présence significative des phénomènes de prise en charge énonciative dans l'expression des connaissances causales m'a amenée à défendre l'idée que les relations causales exprimées dans et par la langue pouvaient être vues comme l'intégration de deux relations constitutives : une relation *efficiente* entre situations et une relation *subjective* entre cette relation efficiente et le sujet parlant qui la prend en charge (Jackiewicz 1999b ; Desclés et Jackiewicz 2006). De nombreux marqueurs témoignent de cette double dimension de la causalité dans la langue. Certains encodent directement la relation complexe, où la relation efficiente et la relation subjective sont intriquées. C'est le cas notamment des marques qui expriment l'assertion de la relation prédicative (14). D'autres, constitués généralement par des ensembles de marques de niveaux différents (typiquement, un prédicat causal et un marqueur de modalité : adverbe, verbe en *que...*), rendent la co-présence de ces deux relations plus transparente (15-16).

(14) *Arrhénius **lie** définitivement la dérive de l'effet de serre **et** l'utilisation des combustibles fossiles.*

(15) *L'ingestion de vin **aurait un effet** protecteur **sur** l'organisme.*

(16) *Selon Freud, l'oubli **est le fruit** d'un refoulement.*

La relation *subjective* et la relation *efficiente* sont étroitement liées entre elles. Tout d'abord, la relation subjective peut modifier sensiblement le contenu de la relation efficiente, ensuite ces deux relations sont souvent indissociables sur le plan de l'expression, enfin, la relation subjective guide l'interprétation de la relation efficiente quand cette dernière est exprimée par des marques polysémiques. Ces trois raisons expliquent pourquoi le traitement des connaissances causales inscrites dans les textes se trouve nécessairement confronté à la question des modalités énonciatives.

Qu'il s'agisse d'une démarche de diagnostic (en médecine...), des analyses de déterminations causales de phénomènes sociaux ou plus généralement de la veille sur les signaux annonceurs de changements ou les relations de dépendance inédites à l'intérieur des univers surveillés, les connaissances causales ne peuvent être isolées de leur contexte énonciatif. Leur intelligibilité et leur efficacité dépendent d'une expression claire et explicite des limites de leur validité (temporelle, spatiale, gnoseologique...) et du degré de certitude qui les caractérise. On peut se demander si ce n'est parce que les connaissances causales se construisent souvent progressivement et dans l'interaction qu'elles sont si finement « étiquetées ».

Inversement, quel que soit le domaine applicatif concerné, pouvoir identifier les différentes voix énonciatives qui s'expriment dans un (ou plusieurs) texte(s) relativement à une connaissance causale, permet de savoir qui s'est exprimé à son sujet, qui fait autorité en la matière, comment

cette connaissance circule (qui cite qui ? est-elle partagée ou au contraire s'agit-il d'un point de vue singulier ?). Pouvoir identifier quelles positions sont formulées à l'égard de cette connaissance par les différentes voix en présence, c'est pouvoir jauger le degré d'adhésion (forte, réticente, nuancée) de chacune d'entre elles, et savoir, pour l'ensemble des voix qui s'expriment, si les points de vue convergent ou divergent, s'inscrivent dans une polémique ou s'ignorent mutuellement. En somme, c'est la dimension dialogique du langage causal qui est révélée par ces différents cas de figure.

5. Conclusion

Le sujet parlant inscrit continuellement sa présence dans son discours. Cette présence peut être plus ou moins visible. Dans le cas des énoncés causaux, elle apparaît comme particulièrement prégnante. La causalité, manipulée dans et par la langue, permet d'organiser des faits d'expérience extra-linguistique. Elle est en même temps empreinte de la subjectivité de celui qui l'énonce et engagée (directement ou indirectement, implicitement ou explicitement) dans des interactions avec d'autres instances de parole. La causalité porte souvent la signature de celui qui l'établit ou la légitime. Le *dictum* est alors indissociable de son *modus*. Elle peut se voir associer également d'autres informations, telles que le degré d'incertitude ou la dépendance par rapport à un point de vue particulier ou à un contexte de validité restreint. Toutes ces données font partie de la « réalité » de la cause. Elles sont nécessaires pour rendre les connaissances causales intelligibles et opératoires, c'est-à-dire exploitables dans l'action.

II. ORGANISATIONS TEXTUELLES

A. Etat des lieux

Participer au travail collectif sur le résumé automatique et la fouille sémantique de textes par points de vue (Minel *et al.* 2001 ; Crispino *et al.* 2004) m'a amenée à aborder la problématique des organisations textuelles.

Pouvoir identifier et annoter sémantiquement des objets textuels notionnels ou fonctionnels (tels que définition, citation, séquence causale...), nécessite au préalable une étude de la manière dont les frontières de ces objets et leurs modes d'organisation (succession, enchâssement...) sont signalés dans et par le discours. Lors du filtrage sémantique des textes, un fragment identifié comme répondant au critère d'extraction peut parfaitement être intégré dans une configuration de segments discursifs fortement corrélés entre eux. S'il n'est pas toujours pertinent de présenter à l'utilisateur l'ensemble de la structure qui comprend le segment ciblé, il est sans doute intéressant de lui offrir la possibilité d'y accéder dynamiquement et de l'assister dans le parcours des éléments liés au segment en question. De manière analogue, la lisibilité des synthèses obtenues par extraction de phrases dépend très clairement de la gestion des liens de cohésion établis avec le co-texte de ces phrases. Par exemple, gérer les liens de cohésion dans un texte tel que (1) demande a minima de pouvoir : (i) identifier les référents des pronoms anaphoriques *elle*, (ii) reconnaître l'enchaînement argumentatif (*car*), (iii) délimiter la séquence énumérative qui réunit les arguments (*primo, deuzio, tertio*) avancés en faveur de la thèse exprimée dans l'amorce (*cette technologie serait en passe de...*).

(1) *Cette technologie serait en passe de sortir du marché pour initiés où elle était jusqu'à présent confinée. Car, primo, elle répond à de vrais besoins de mobilité jusqu'à présent non satisfaits chez les usagers. Deuzio, elle arrive à point nommé pour donner une bouffée d'oxygène aux entreprises de télécoms, et indirectement à l'ensemble du secteur des technologies de l'information et des télécommunications. Et, tertio, l'évolution de la réglementation, en France, fait que n'importe qui peut, depuis novembre 2002, l'exploiter sur tout le territoire.* (Le Monde 13 mai 2003, p.5)

Plus généralement, l'une des difficultés que connaît le traitement automatique des textes tient au fait que l'on ne sait pas (ou pas toujours de manière satisfaisante) comment les découper en unités de traitement. La prise en considération des indicateurs graphiques et typographiques constitue un préalable, mais les phrases se révèlent bien souvent des unités trop étroites et les paragraphes ou les sections, des unités trop vastes. Le fait de pouvoir segmenter les textes en prenant appui sur l'organisation des contenus réalisée par l'auteur lui-même représente un apport significatif, quelle que soit la visée ultérieure du traitement. Cela peut servir à différentes fins : (i) parcourir le texte en suivant ces objets (navigation intra-documentaire), (ii) en extraire certains, conformément à un critère sémantique ou à un jugement d'importance (synthèse sélective), (iii) découper le texte en portions thématiquement homogènes, (iv) écarter du traitement certaines parties du texte, etc.

En TAL, ces besoins ont impulsé un développement significatif d'applications directement intéressées par certains aspects de la textualité, tels que la cohésion thématique, référentielle, anaphorique ou relationnelle pour la segmentation en unités de traitement et la navigation textuelle (Bilhaut *et al.* 2003), (Couto *et al.* 2004), (Hernandez et Grau 2005). Les travaux sur l'identification de certains « objets » textuels comme les titres, les annonces thématiques, les annonces de plan, les énoncés conclusifs, les références bibliographiques... (Minel *et al.* 2001), (Laublet *et al.* 2002), (Abdalla and Teufel 2006), (Jacques et Rebeyrolle 2006), (Teufel 2006) se sont poursuivis, en bénéficiant d'avancées conjointes en linguistique textuelle et en traitement automatique.

En linguistique textuelle, un important courant de recherches s'est constitué sur les questions de cohésion et de cohérence, visant l'exploration systématique des modes d'articulation des discours et leur signalisation dans des textes. En généralisant, on peut dire que les théories ou modèles proposés dans ce cadre s'attachent à expliquer comment les régularités que l'on peut observer dans l'usage de certains marqueurs à l'échelle de la phrase et du discours peuvent conduire à des contraintes d'emploi et devenir des faits de langue. Sont également abordées les questions concernant le découpage des documents en unités de composition et de traitement, l'identification

et la configuration de ces unités, les valeurs sémantiques que l'on peut leur assigner et les liens qu'elles entretiennent ou instituent. Les études portent sur des expressions ou constructions de toute nature pour autant qu'elles s'attachent à montrer comment les formes analysées s'inscrivent dans les énoncés où elles apparaissent, comment elles s'articulent avec les énoncés ou segments précédents et comment elles pèsent sur la forme et le contenu des énoncés ou segments ultérieurs. L'ouvrage collectif dirigé par Frédéric Calais (Calais 2006) donne un aperçu de la diversité des recherches menées en linguistique textuelle et en analyse de discours sur les questions de cohérence et de cohésion.

Dans ce contexte, un courant de recherches consacré à l'encadrement du discours s'est constitué, suite notamment aux travaux de Michel Charolles (Charolles 1995, 1997, 2002). La notion générale de cadres y est employée pour désigner les circonstances dans lesquelles il faut envisager un certain état ou une série d'événements. Les cadres contribuent à partitionner l'information dans des rubriques homogènes. Ils instaurent un lien de cohésion textuelle que le lecteur reconstruit à partir de nombreux indices et en particulier en s'appuyant sur les introducteurs de cadres. La portée de ces marques, syntaxiquement non intégrées à l'énoncé où elles figurent matériellement, généralement en position initiale, peut s'étendre sur plusieurs phrases créant ainsi une véritable unité textuelle, homogène thématiquement et relativement autonome par rapport au contexte. On trouvera dans (Charolles et Péry-Woodley 2005) et (Vigier et Terran 2005) deux livraisons consacrées aux différents types de cadres de discours (médiatifs, temporels, spatiaux, organisationnels, thématiques...). Les auteurs y interrogent notamment la notion de portée, discutent la capacité d'indexation de plusieurs familles d'adverbiaux, traitent des interactions entre cadres et relations de discours.

Notons que plusieurs travaux sur l'encadrement du discours ont donné lieu à des collaborations étroites entre linguistes et informaticiens, notamment sur l'identification et la délimitation automatique des cadres thématiques, des cadres médiatifs et des cadres organisationnels, dont (Ferret *et al.* 2001) et (Couto *et al.* 2004). Plus récemment, le projet ANR Annodis¹⁶ (2007-2010), mené par les laboratoires CLLE-ERSS, IRIT (LILAC) et GREYC, s'est donné pour objectif la construction d'un corpus annoté et outillé, pour l'étude de l'organisation du discours en français (Péry-Woodley *et al.* 2009). Les travaux menés dans ce cadre ont porté sur l'annotation des structures discursives de haut niveau, dont les structures énumératives (Ho-Dac *et al.* 2010).

Ma participation aux différents travaux de cette communauté (composée notamment de plusieurs membres des laboratoires LATTICE, LaLIC(C), ERSS, GREYC) concerne essentiellement l'étude des expressions introductrices de cadres organisationnels. Il s'agit des adverbiaux corrélatifs comme *premièrement/ deuxièmement, d'une part/d'autre part, d'un côté/de l'autre, en premier lieu/en second lieu...* qui inscrivent les contenus dans des séries linéaires et les localisent métadiscursivement dans le développement du discours en cours d'élaboration.

Sur le plan théorique, cette recherche concerne en réalité deux sortes d'objets distincts, bien qu'intimement liés entre eux : objet discursif de nature énonciative « cadre organisationnel » et objet abstrait « série ».

J'ai procédé en trois temps. J'ai réalisée une première étude, décrite dans (Jackiewicz 2002) et (Jackiewicz et Minel 2003), sur (i) le fonctionnement sémantique et discursif des marqueurs d'intégration linéaire, introducteurs prototypiques des cadres organisationnels, ainsi que sur (ii) les caractéristiques morphologiques des séries balisées par ces marques. Ce travail a donné lieu à une implémentation dans le système ContextO. Le programme correspondant identifie les ouvertures des cadres organisationnels et délimite la structure en série sur la base d'un ensemble d'heuristiques issues des observations de corpus.

Ensuite, sur la base de cette première expérience, j'ai tenté de déplacer mon regard des marques vers l'objet lui-même : la structure en série et ses fonctions. Je me suis intéressée à la sériation en tant que résultat d'une opération complexe, faisant appel à la fois au balisage des unités textuelles, à leur groupement et à leur mise en ordre (Jackiewicz 2005). Sur ces opérations de nature purement organisationnelle peuvent se greffer différentes relations sémantiques ou argumentatives s'établissant entre différents composants de la série. Cette approche, à la fois plus large, plus systématique et plus discursive, aide à dominer la diversité des moyens que les auteurs emploient effectivement pour signaler des contenus sériés. Les corrélations des marques attestées, souvent

¹⁶ <http://w3.erss.univ-tlse2.fr/annodis>.

fort hétérogènes entre elles, témoignent très clairement des différentes configurations d'opérations énonciatives, sémantiques et argumentatives. La section B permet de prendre connaissance de ce travail d'une manière plus approfondie.

Enfin, en collaboration avec Gustavo Crispino et Jean-Luc Minel du Lalic, à partir des études menées sur les cadres organisationnels et les cadres thématiques, j'ai travaillé à la construction d'un langage de représentation des connaissances linguistiques LangTex. La vocation de ce langage formel était de spécifier d'une manière unifiée les descriptions linguistiques nécessaires au repérage d'objets textuels qui organisent les textes écrits (Crispino *et al.* 2004). Ce travail a été mis en œuvre dans le système ContextO.

Par ailleurs, courant les années 2007 et 2008, j'ai participé en tant que membre associé aux travaux sur les adverbiaux cadratifs conduits dans le cadre du projet ANR SFA « *Spatial Framing Adverbials : linguistic and psycholinguistic approaches* », (Lattice, ENS, Paris 3). En est issue une étude spécifique sur les expressions détachées en « en », construites sur le schéma « en Adj N » (Jackiewicz *et al.* 2009b).

B. Les séries linéaires dans le discours

1. Introduction

Tout discours résulte d'un processus d'énonciation lequel, outre instancier les places des actants et informer les coordonnées spatio-temporelles, délimite et articule les contenus. Ces contenus peuvent être liés, hiérarchisés, divisés, enchâssés... suivant le contenu propositionnel dénoté, les schémas argumentatifs et les visées pragmatiques de l'énonciateur. Un énonciateur coopératif peut augmenter très nettement la lisibilité de son discours en rendant saillantes certaines de ses caractéristiques formelles ou fonctionnelles. Il dispose à cet effet d'une large palette de moyens typo-dispositionnels et discursifs (découpage en paragraphes, diverses « balises », commentaires autonomes, mots de liaison...). Ces indices donnent au co-énonciateur, dans une visée résolument dialogique, des instructions pour la gestion de l'activité de lecture et de la compréhension.

La *sériation*, qui est un procédé fondamental d'articulation textuelle permettant à la fois d'ordonner et de situer sur un axe d'équivalence paradigmatique des portions de texte, s'est imposée rapidement comme l'objet de traitements transversaux à plusieurs applications informatiques. De plus, l'intérêt théorique de cette opération sur le plan de l'analyse du discours est évident. La sériation dans le discours est un procédé énonciatif par lequel l'énonciateur informe le co-énonciateur qu'il donne à son propos la forme d'une suite ordonnée d'unités équivalentes par rapport à un critère, ce qui peut être paraphrasé par la formule : « je dis que ce qui est dit forme une succession ordonnée d'éléments de même nature ». Ainsi dans la séquence (2), les adverbiaux *Primo* et *Deuzio* fonctionnent de concert pour introduire respectivement le premier et le deuxième item, au sein d'une série de deux éléments. La longueur de la série et le principe fédérateur qui unit ses composants sont explicitement annoncés dans l'énoncé « amorce » : *Deux éléments fondent...*

(2) *Deux éléments fondent le plan de financement choisi par le gouvernement pour réussir sa réforme des retraites. **Primo**, a encore rappelé François Fillon, ministre des affaires sociales, devant les parlementaires de l'UMP le 5 mai, « il n'existe pas de trésor caché » qui permettrait de raser gratis. **Deuzio**, il est hors de question de revenir sur la promesse de Jacques Chirac de réduire, ou tout au moins de stabiliser, le niveau des prélèvements obligatoires. (13 mai 2003, p. 2, LE MONDE)¹⁷*

L'opération de sériation constitue également un puissant facteur de cohésion textuelle, combinant les relations de connexion entre items de la série et les relations d'indexation, créatrices d'items en tant que blocs homogènes et distincts. Dans cette optique, il est intéressant d'aborder l'étude des marques impliquées dans l'expression des séries du point de vue de leur capacité simultanée à connecter des segments textuels et à les indexer (ou cadrer).

Dans la présente section, je me propose de montrer à partir d'une expérience complète, allant d'une analyse linguistique à l'implantation, menée sur l'objet textuel « série linéaire » (Jackiewicz 2002 ; Jackiewicz et Minel 2003 ; Couto *et al.* 2004), comment l'objectif du repérage automatique d'objets et de structures de discours (i) oriente le travail linguistique de collecte et d'analyse de marques d'organisation textuelle, (ii) questionne le cadre méthodologique, faisant notamment apparaître la nécessité de représenter formellement et de traiter informatiquement des faisceaux d'indices contextuels, relatifs à plusieurs opérations discursives en présence, (iii) amène à considérer la question des normes et des usages à l'œuvre dans les corpus textuels exploités.

Les phases de recueil et d'analyse des marques et des connaissances linguistiques impliquées dans l'expression des séries sont aussi complexes qu'essentielles. Elles font apparaître que ce travail nécessite conjointement (i) le recours à des corpus textuels, car les manuels de grammaire ne disent pas comment construire et signaler une série ; (ii) des analyses linguistiques « classiques », car les exemples attestés ne sont pas suffisants, aussi est-il nécessaire de procéder à des manipulations paradigmatiques pour faire apparaître des contraintes et des régularités d'expression ; (iii) la prise en considération des propriétés de nature logique et cognitive de la sériation, ce qui aide à cerner la nature de cette opération dans le discours et, par conséquent, à dominer la diversité des marques linguistiques impliquées dans la sériation discursive.

¹⁷ Le gras marque les introducteurs d'item, le souligné les éléments pertinents des amorces et des rétro-évaluations.

Ce travail a été mené dans la lignée des recherches sur l'encadrement du discours (Charolles, 1997), bénéficiant en partie du contexte du projet REGAL. Il s'est appuyé sur les travaux portant sur les marques d'intégration linéaire (Turco et Coltier 1988)¹⁸, les énumérations (Péry-Woodley 2000) et (Luc et Virbel 2001). L'objet « énumération » que ces travaux ont étudié dans sa réalisation « typo-dispositionnelle » englobe la structure discursive linéaire formée d'une suite de cadres organisationnels.

« L'énumération constitue un cas remarquable d'exploitation des possibilités liées à l'écriture. D'une part, la forme écrite autorise le développement d'énumérations aussi longues et enchâssées que nécessaire et d'autre part, ces énumérations constituent un cas particulièrement clair de la correspondance entre les formes discursives développées à base d'expressions adverbiales (premièrement, deuxièmement, puis, enfin,...) et des formes syntaxiquement réduites comportant des traces typo-dispositionnelles de ces réductions (usage des numératifs, des diacritiques, etc.). Aussi, les énumérations participent à la fois de structures entièrement visuelles qui n'ont pas de sources dans l'oral et de structures entièrement discursives. » (Luc et Virbel 2001).

Mes résultats permettent de questionner correspondance posée entre les organisateurs discursifs et les organisateurs « visuels », dans la mesure où ils apportent des données précises, entre autres sur la longueur et la structure des séries balisées discursivement. Comme on le verra dans la suite du texte, généralement plus courtes et moins complexes dans leur structure que leurs « homologues visuelles », les séries balisées par les marqueurs d'intégration linéaire sont aussi plus difficiles à identifier à cause, entre autres, de la polysémie des formes qui peuvent jouer le rôle des organisateurs discursifs.

2. Les séries linéaires : observation en corpus

Que peut-on apprendre sur corpus en matière de la formation des séries linéaires ? Quelles sont leurs caractéristiques morphologiques ? Par quelles formes d'expressions sont-elles balisées ? Sur un fond textuel composé des articles du *Monde* et du *Monde diplomatique*¹⁹, j'ai collecté manuellement 311 séries textuelles différentes, ce qui représente, en matière de balises d'items, autant de suites d'introducteurs discursifs différentes {(1) *d'une part, de l'autre* ; (2) *en premier lieu, en second* ;... ; (311) *primo, deusio*}. Cette manière de prélever les observables cherche à mettre en évidence les configurations possibles, et non leurs fréquences d'emploi. Une série est composée d'au moins deux items. Elle est identifiable grâce à l'amorce et/ou aux introducteurs de ses items, dont au moins un doit être une marque d'intégration linéaire « classique » telle que {*d'une part, d'autre part, de l'autre* ; *d'un côté, d'un autre côté* ; *en premier lieu, en second lieu,...* ; *en dernier lieu* ; *d'abord, ensuite, enfin* ; *dans un premier temps, dans un second temps...* ; *premièrement, deuxièmement...* ; *primo, secundo, tertio...*}.

2.1 Une série « typique » et ses variantes

Il ressort de cette analyse que les séries textuelles balisées par des marques discursives sont (i) typiquement constituées de 2 ou de 3 items (avec une longueur maximale attestée de 10 items, mais théoriquement illimitée pour certaines catégories d'introducteurs), dont la portée peut aller d'un syntagme à une suite de paragraphes ; (ii) structurées à un niveau, plus rarement à deux niveaux (emboîtement) ; (iii) précédées par une amorce explicite ou reconstruite *a posteriori* ; (iv) parfois clôturées par une rétro-évaluation portant sur l'ensemble de la série.

Sur le plan de la signalisation discursive, (i) les items sont généralement introduits par des marqueurs d'intégration linéaire (MIL) (*en premier lieu, premièrement, d'abord...*), dont les configurations composent assez librement des introducteurs issus de la numération et des paradigmes temporels et spatiaux, mais qui peuvent également être incomplètes ou partiellement relayées par d'autres marques (*mais, parce que...*) ; (ii) l'amorce explicite contient généralement un classificateur (*élément, étape, composante...*) et un quantificateur (*deux, double, triple...*), qui assurent d'une manière complémentaire aux MIL le signalement des items (leur paradigme

¹⁸ Signalons deux études récentes de grande ampleur concernant les marqueurs corrélatifs (Schneidecker 2006) et (Svensson 2010).

¹⁹ Le *Monde diplomatique* (LMD) sur un CD-ROM contient 12 520 documents parus entre 1984 et 1998. Le CD-ROM du *Monde* contient l'édition complète des années 2002 et 2003. Ce corpus a été choisi en raison du faible degré de structuration visuelle (dispositionnelle et typographique) des textes qui le composent.

commun, leur nombre) ; (iii) la rétro-évaluation, lorsqu'elle est présente, est introduite par des marques de reformulation (*en bref, autrement dit, en définitive...*) ou de quantification totalisante (*tous, au total, les deux, l'ensemble...*).

La nature proprement discursive des introducteurs exerce une incidence sur les caractéristiques morphologiques de la série. Généralement plus courtes et moins complexes dans leur structure que leurs homologues « visuelles », les séries balisées par les marqueurs d'intégration linéaire sont aussi plus difficiles à identifier, en raison de la variabilité et de la polysémie des formes qui peuvent jouer le rôle des organisateurs discursifs. On constate toutefois que plus une série est longue et/ou plus ses items sont de taille importante, plus le signalement discursif est saillant (y compris visuellement) et régulier.

Ce résumé des propriétés caractéristiques des séries linéaires, qui donne l'impression d'une variabilité difficile à dominer, réfère à une somme de données soumises à l'influence d'au moins trois facteurs. Premièrement, la morphologie d'une série est susceptible de variations. Deuxièmement, pour une série donnée, il existe plusieurs possibilités logiquement équivalentes entre elles, mais sémantiquement et stylistiquement différentes, permettant de la signaler. Enfin, sur ces deux plans (de la morphologie et de l'expression), le respect ou non de certains critères de lisibilité (qu'ils soient « officiels » et imposés par des normes éditoriales, ou subjectifs dictés par des choix personnels de l'auteur), a une incidence sur l'effort de lecture nécessaire pour délimiter et interpréter correctement une série.

2.2. Signalement des séries : règles prescriptives et faits attestés

La question des normes d'écriture qui s'imposeraient *a priori* aux scripteurs a été posée, et cela avant toute analyse des observables. Le guide rédactionnel du Monde (*Le style du Monde*, 2002) ne fournit pas de recommandations sur la structuration sérielle des textes. Il existe toutefois des règles, non consignées dans ce document, que certains journalistes²⁰ affirment avoir apprises lors de leur formation et appliquer systématiquement dans leurs écrits. Il s'agit, pour ce qui concerne la sériation, de la règle stipulant que toute énumération doit être introduite d'une manière régulière (« si j'écris *d'une part*, je dois poursuivre par *d'autre part* ou par *de l'autre* »), et de celle qui régit l'emploi de l'adjectif *second*, en le réservant aux énumérations composées de deux éléments.

Sur un échantillon de 200 séries tirées du *Monde* : on constate que (i) seulement 18 % des séries possèdent des balises régulières, (ii) dans 40 séries sur 160 séries de longueur supérieure ou égale à trois items, l'introducteur du deuxième item comporte l'adjectif *second*. Voici quelques exemples de séries d'introducteurs qui échappent à l'une ou à l'autre règle : *Premier commandement, Deuxio, Tertio ; En premier lieu, En second, En troisième lieu, En quatrième lieu, Ensuite, Enfin ; Primo, En deuxième lieu, Troisième grief ; D'abord, En second lieu, Troisièmement, Enfin ; Ø, Ensuite, Troisièmement, Enfin ; Primo, Ø, En troisième lieu ; Ø, Ø, Tertio, La quatrième conséquence*. Le symbole Ø représente une balise manquante. Par ailleurs, on note la présence parmi ces introducteurs de formes qualifiées par (Grevisse 1993) de « barbares » : *deusio, deuzio, deuxio*. Les formes irrégulières figurent tant dans des articles rédigés par des journalistes « maison », que dans des écrits signés par des correspondants ou des experts extérieurs.

3. Les séries linéaires à la lumière des analyses linguistiques

Les données collectées sur corpus révèlent les procédés morphologiques et sémantiques de constitution de balises (introducteurs d'items). Elles donnent également des indications sur d'éventuelles contraintes qui peuvent régir l'association des balises au sein d'une série particulière. Sur cette base, il devient possible de prédire des séries non attestées dans le corpus, mais réalisables en pratique. Les séries attestées livrent par ailleurs un savoir précieux sur les conditions contextuelles dans lesquelles une forme *a priori* polysémique ou polyfonctionnelle telle que *d'une part, premièrement* ou *enfin* pourra être interprétée comme un introducteur potentiel d'item au sein d'une série.

3.1 Création de paradigmes d'introducteurs

Les introducteurs attestés au sein des différentes séries prélevées dans notre corpus ne représentent pas toutes les balises possibles. Il est donc nécessaire d'enrichir les paradigmes

²⁰ Il s'agit d'un journaliste du *Monde* que j'ai interrogé personnellement.

amorçés, regroupant successivement tous les introducteurs possibles pour le premier, le deuxième... le n-ième item de la série. Chaque paradigme enferme les marques logiquement équivalentes, car elles réfèrent au même rang dans la série, mais différentes sur les plans morphologique et sémantique.

Plusieurs procédés permettent de compléter les paradigmes constitués. Tout d'abord, la plupart des séries régulières étant potentiellement infinies, on est amené à ajouter aux formes attestées pour les k premiers items d'une série, les formes existantes qui vont au-delà du k-ième item. Ainsi, on « prolongera » la série {*primo, secundo, tertio*} par les termes {*quarto, quinto, sexto, septimo...*}. En deuxième lieu, il convient de systématiser pour certains introducteurs les formes sensibles au genre, ainsi que les déterminants qui peuvent leur être associés. De cette façon, si l'on atteste l'existence des syntagmes *ce premier / le premier / un premier*, on peut compléter le paradigme du rang 1 par les formes *cette première, la première / une première*. En troisième lieu, il est nécessaire de vérifier pour certains introducteurs composés la possibilité d'y insérer des éléments adverbiaux ou adjectivaux (*en premier lieu / en tout premier lieu...*) en surveillant les conséquences sur le plan sémantique que cela entraîne. Notons enfin que les quelques rares occurrences d'introducteurs verbaux attestées (*commencer par, se terminer par...*) soulèvent la question de savoir s'il est opportun d'accepter comme balises potentielles leurs synonymes les plus courants.

A l'issue de ce travail, il devient possible de ré-examiner les ressources attestées à la lumière de celles qui sont attestables. On constate qu'il existe en moyenne 20 possibilités différentes pour introduire un item. Ce nombre est sensiblement plus important pour le deuxième item (49 formes différentes, toutes attestées), ce qui traduit un besoin fréquent d'exprimer une opposition ou plus simplement une nette différenciation entre deux unités. Les séquences formées de deux éléments se caractérisent donc sur le plan stylistique par une grande diversité d'introducteurs possibles, tous effectivement employés. Le deuxième fait remarquable est que dans le cas des séries longues (quatre items ou plus), peu de balises disponibles sont effectivement exploitées par les auteurs. Les séries longues tendent à être régulières (certainement pour mieux soutenir la mémoire du lecteur), ce qui se manifeste entre autres par l'emploi préférentiel des introducteurs appartenant au même paradigme. Deux « familles » d'introducteurs sont utilisées de préférence pour introduire de longues énumérations d'items discursifs : {*premièrement, deuxièmement, troisièmement...*} et {*le premier, le deuxième, le troisième...*} avec, pour cette dernière, plusieurs variantes possibles liées au genre et à la détermination.

3.2 Typage des introducteurs

Une fois les paradigmes d'introducteurs constitués, il est possible de les décrire sur les plans morpho-syntaxique, sémantique et fonctionnel. Plusieurs critères, intervenant conjointement ou de manière indépendante, permettent de distinguer :

- les balises d'ouverture (*d'abord, en premier lieu...*), de relais (*ensuite, en second lieu...*) et de clôture (*en dernier lieu, enfin...*) ; on note que pour signaler le dernier item de la série, les auteurs n'hésitent pas à composer des introducteurs « redondants » à partir de plusieurs indices de clôture (*enfin le dernier élément...*) ;
- les balises indépendantes (*en premier lieu, premièrement...*) et les balises liées (*le premier objectif, troisième composante...*), la liaison s'établissant entre l'item et l'amorce par l'intermédiaire d'un classificateur (*objectif, composante...*) ;
- les balises simples (*troisièmement, enfin...*) et les balises composées (*enfin le troisième élément*) ;
- les balises neutres (rôle uniquement organisationnel : les marques d'intégration linéaire) et les balises bi-fonctionnelles (rôle organisationnel et argumentatif : *mais, parce que...*) ;
- les balises discursives, les balises typographiques (tiret, puce, numéro), les balises typo-discursives (*- premièrement... ; - deuxièmement...*).

Ces distinctions sont importantes, car elles permettent d'adapter finement les heuristiques de repérage aux différentes catégories d'introducteurs. Par exemple, une balise bi-fonctionnelle est recherchée quand on ne trouve pas à l'intérieur d'un espace de recherche correspondant à un item donné de marque d'intégration linéaire susceptible de l'introduire. Comme on le voit dans (3), la deuxième occurrence du connecteur *parce que* introduit (comme la première) un argument étayant une position, mais elle assure également une fonction de balise.

(3) *Et pourtant, de plus en plus nombreux sont ceux qui dénoncent ce système comme une imposture. En premier lieu parce que, en Europe, il a toléré l'apparition de 20 millions de sans-emploi et de 50 millions de pauvres... Parce que certains Etats en viennent à accepter une sorte de tiers-mondisation de leurs sociétés. (LMD octobre 96, p.1)*

3.3 Propriétés contextuelles des introducteurs

Les marques discursives qui introduisent les items des séries linéaires sont dans leur majorité polysémiques et polyfonctionnelles (*d'abord, d'une part, premier, d'un côté...*). En tant que candidat introducteur d'item, une occurrence de forme doit (i) vérifier les conditions (locales) dans lesquelles elle devient un introducteur de cadre de discours (cadre organisationnel, en l'occurrence), (ii) être mise en relation avec d'autres marques qui charpentent la structure en série (marques potentiellement introductrices d'items, équivalentes et complémentaires ; classificateurs au sein d'une amorce...), (nous explicitons les différents cas de figure possibles dans le § 5).

En qualité d'introducteur de cadre, le marqueur ne doit pas être syntaxiquement intégré à l'énoncé où il figure matériellement, ce qui contraint l'ensemble de ses positions possibles. En général, il figure en initiale détachée, souvent absolue, ou dans une incise. Un certain nombre de marques significatives peuvent se trouver dans son contexte proche (gauche, plus rarement droit), comme on le voit dans (4). Dans une architecture en série, il s'agit le plus souvent (i) des signes de ponctuation tels que deux-points ou point-virgule, (ii) de certains connecteurs (*mais, et, surtout, parce que...*) ou marqueurs de clôture (*enfin, finalement...*), (iii) de divers marqueurs tels que *si, c'est, comme...*, (iv) des marques graphiques d'énumération telles que tirets, puces ou numéros.

(4) *Les rares compagnies qui le font, et qui modulent leurs tarifs en conséquence, sont unanimes: le risque ne dépend absolument pas de la cylindrée mais, **d'une part**, de "l'image" du modèle concerné (les "sportives" ont plus d'accidents que les routières) **et, d'autre part**, et surtout, de l'expérience du pilote. (LMD, oct. 93)*

3.4 Contraintes sur la co-occurrence des introducteurs

Les séries attestées témoignent d'une grande « souplesse » dans le signalement discursif de leurs composants. Pour autant, cela n'entraîne pas une combinatoire totalement libre entre introducteurs. On constate que le phénomène de répétition (possible pour les introducteurs relais qui ne marquent pas la position de l'item : *autre, ensuite, puis...*), reste réellement marginal et qu'il existe une tendance forte à différencier entre eux les items d'une série, ce qui passe par le choix des introducteurs devant être, eux aussi, différents. Dans le cas des séries longues, le signalement tend à être régulier et homogène (une fois ouverte par *premièrement*, la série se poursuit avec *deuxièmement, troisièmement, quatrièmement*, etc.). L'existence possible d'emboîtements autorise la succession de deux balises d'ouverture ou de deux balises de clôture. On note alors que les auteurs prennent généralement soin de les différencier nettement entre elles. Dans l'extrait (5), on trouve ainsi l'introducteur *Le premier* suivi par *Premièrement*.

(5) *Si l'on tente de formaliser quelques évolutions possibles du continent, ce que l'on pourrait nommer deux "types idéaux" s'imposent. **Le premier** est celui de la segmentation généralisée. Elle pourrait être précipitée par trois éléments. **Premièrement** un échec politique et institutionnel à l'Ouest (...) et à l'Est (...). Cette double évolution signifierait (...) **Deuxième élément**: le maintien d'importantes quantités d'armes de toutes natures sur le continent, alors que s'affaiblissent les pouvoirs ayant vocation à les contrôler. (...) **Troisième élément**, un développement économique anarchique couplé à une démocratisation chaotique, dessinant en Europe des zones très différenciées: trous noirs, zones d'ombres, monde riche, etc. (...) Dans un continent ainsi haché, (...) **L'autre "type idéal"** d'évolution de la situation européenne est celui de la recomposition (...). (LMD, décembre 92, pp.10-11)*

4. Nature cognitive et logique de l'objet « série »

La sériation est l'un des procédés fondamentaux permettant d'articuler des contenus textuels (Vandendorpe 1999). Les listes (séries, énumérations) s'avèrent être des objets aussi anciens que l'écriture (Goody 1979). C'est une opération de nature cognitive qui intervient dans plusieurs sphères d'activité humaine. Elle est de toute évidence impliquée dans le processus général de planification d'action. Selon Piaget, la sériation est nécessaire à l'acquisition du concept de nombre (Piaget 1974). Des études plus formelles menées en rapport avec la sériation, notamment par Grize (Grize 1974), mettent en évidence trois propriétés de base que vérifient les éléments d'une série : être disjoints, être équivalents et être ordonnés.

4.1 Découper et baliser

Selon la première propriété, les listes (séries) disjoignent les éléments. Dans le discours, une série implique discontinuité et non continuité. Elle trace dans le matériau textuel se déroulant linéairement les frontières des unités, permettant ainsi leur identification et leur différenciation. Les limites de ces segments doivent être clairement signalées. Les relations syntaxiques et la ponctuation, qui mettent en évidence des unités infra-phrastiques, au-delà de la phrase se voient relayées par la présence de marques discursives, sorte de « balises », dotées de capacités d'indexation et d'articulation. Ces balises ont pour effet de séparer les éléments du discours les uns par rapport aux autres et de les séparer du contexte plus large, dont lequel s'inscrit l'ensemble de la série. Cette séparation, à son tour, aide à interpréter correctement des contenus présentant un risque d'ambiguïté. En effet, c'est quand le découpage du contenu en unités de sens ne va pas de soi, que la présence de balises s'impose. C'est le cas dans (6), où le journaliste qui rapporte un extrait d'un livre (dans lequel, il a d'ailleurs opéré des coupures) est obligé de rajouter des balises (qu'il place entre crochets) pour que le lecteur puisse identifier sans ambiguïté les fragments correspondant à chacune des six erreurs dont traite le texte.

(6) *Jean-Marie Messier livre sa version de la crise de Vivendi Universal – VERBATIM « Claude Bébéar est là, en véritable coprésident » EXTRAITS de Mon vrai journal de Jean-Marie Messier (éd Balland). Les six erreurs de Jean-Marie Messier. Ma première erreur est certainement de ne pas avoir opéré plus vite et plus radicalement la séparation définitive entre Vivendi Environnement [les activités historiques de la Générale des eaux] et Vivendi Universal (...). La deuxième erreur a sans doute été de ne pas faire évoluer le conseil d'administration (...) [Troisième erreur] j'ai trop tardé avant de changer l'équipe Canal+ et relever Pierre Lescure de sa responsabilité (...). J'ai eu tort de me laisser séduire par l'homme et intimider par l'icône (...). Son incapacité à organiser Canal+ était patente (...). [Quatrième erreur] Nous en avons peut-être fait un peu trop, trop vite (...). Les acquisitions de Maroc Telecom ou celle de mp3.com étaient certainement moins prioritaires (...). [Cinquième erreur], la relation avec les politiques français. [Sixième erreur] Ma surcommunication. Je n'ai pas écouté les mises en garde répétées de mes collaborateurs, de mes amis, d'Antoinette [sa femme]. (Le Monde 12 nov. 2002, page 12)*

Les exemples (7) et (5) illustrent deux autres cas où l'introduction de balises s'impose. Dans l'extrait (7), il s'agit de partitionner en deux un ensemble de plus de 2 éléments. Dans l'extrait (5) (§3.4), on note des énumérations enchâssées. Le premier élément d'une série de deux contient à son tour une série de trois items. L'ensemble étant linéaire, il faut des balises explicites pour que le lecteur puisse reconstituer cette structure enchâssée.

(7) *Il est vrai que, sur l'échelle du social-libéralisme, l'Allemagne se situe quelque part au milieu, entre la France de Lionel Jospin d'une part, l'Espagne de José María Aznar et l'Italie de Silvio Berlusconi, d'autre part. (Le Monde, 11 04 02, p.7)*

4.2 Grouper

La deuxième propriété des listes met en lumière l'équivalence de leurs éléments, lesquels, de ce fait, sont considérés comme étant de même nature ou de même importance. Les items d'une liste sont équivalents, car (i) ils sont regroupés sur la base d'une propriété ou d'un ensemble de propriétés qu'ils partagent tous, (ii) leurs caractéristiques propres ne sont pas considérées. En effet, l'abstraction des qualités qui dégage le principe fédérateur conduit à écarter les qualités différentielles. Dans le discours, c'est en général l'amorce qui exprime le principe fédérant les items, voir (7). Un certain nombre de classificateurs (*cause, élément, cas, axe, fonction, domaine, point, piste...*), dont le trait commun est leur capacité à exprimer des relations d'appartenance (ensembliste ou méréologique), peuvent participer à l'expression de ce principe²¹.

Mais, il n'existe pas toujours d'amorce explicite. La séquence textuelle qui précède la série peut être interprétée *a posteriori* comme une amorce, si les items de la série entretiennent tous une relation argumentative avec cette séquence. Dans (8), une relation d'étayage est instaurée entre une thèse et les différents arguments chargés de l'étayer. L'indice fort d'une telle série « argumentative » est la présence de connecteurs (*parce que, en raison de...*) qui doublent ou remplacent les MIL. L'information sur ce qui unit la série peut également être exprimée par une

²¹ Notons que les classificateurs « temporels » (*temps, phase, étape, épisode...*) sont un indice fort en faveur d'une organisation temporelle des contenus « sériés » et non uniquement énonciative, comme c'est le cas pour la plupart des séries introduites par des MIL. Le cas des séries temporelles mérite de toute évidence une étude en soi.

rétro-évaluation, comme le montre l'exemple (10). Ce segment, placé après les items, conclut en quelque sorte l'ensemble, ou affirme quelque chose de général à propos de la totalité des items.

(8) *Une fois passées les élections françaises et allemandes, à l'automne, l'Europe va devoir faire six ensembles de choix des plus importants. **Premièrement**, ce sera, au mois de décembre à Copenhague, l'élargissement à vingt-cinq pays. **Deuxièmement**, dans les six premiers mois de 2003, la présentation par la Convention, dirigée par Valéry Giscard d'Estaing, d'une nouvelle Constitution. **Troisièmement**, (...). **Quatrièmement**, (...). **Cinquièmement**, (...). **Enfin**, (...).* (Le Monde, 30 juillet 2002, p.12)

(9) *Ayant, à plusieurs reprises, exprimé son admiration pour le général putschiste chilien Augusto Pinochet, le général Lebed serait-il tenté par le coup d'Etat? Contrairement à ce qu'a prétendu, le 16 octobre dernier, le ministre de l'intérieur, M. Anatoli Koulikov, c'est fort improbable. **En premier lieu**, parce qu'Alexandre Lebed connaît l'état de profonde désorganisation des forces armées; **ensuite**, parce qu'une tentative de cette nature entraînerait la dislocation de la Fédération de Russie de la même manière que la tentative de putsch de 1991 entraîna la dislocation de l'Union soviétique; **enfin**, parce que le général, qui est un "national-populiste pur", perdrait toute sa popularité.* (LMD, novembre 1996, p.1)

(10) *Imaginons le téléspectateur de base, si cette catégorie, non répertoriée par l'Insee, existe encore. Qu'est-ce qu'il en déduit le téléspectateur de base ? **Premièrement**, que M. Sarkozy n'habite pas Saint-Germain-des-Prés. Exact, il est domicilié à Neuilly. **Deuxio**, qu'il n'est pas content qu'on montre à la télé, reportages à l'appui, que rien n'a changé sur le terrain depuis qu'il est là, et que la situation a même empiré. **Troisièmement**, qu'Elise Lucet l'énerve un peu, beaucoup, passionnément. **Bref**, on avait droit, jeudi soir sur France 3, à une de ces empoignades entre ministre et journaliste comme on en voit peu sur le PAF.* (Le Monde, 19 octobre 2002 p.34)

4.3 Ordonner

La troisième propriété d'une liste (série), intimement liée aux deux précédentes, veut que ses éléments soient ordonnés. Etant donné que tous les éléments d'une série sont équivalents, la seule façon de les distinguer est de les ordonner et de leur associer une position. L'ordre d'énumération est donc la seule propriété compatible avec la propriété d'équivalence. Dans le discours, les items d'une liste sont forcément ordonnés, en suivant le flux de la chaîne verbale. Mais le fait de donner une indication supplémentaire, qu'elle soit relative (*l'autre, d'un autre côté, le dernier...*) ou absolue (*le deuxième, secondement, dans un deuxième temps...*), sur la position d'un item, permet de l'isoler plus facilement de l'ensemble, de l'identifier parmi les autres items pour y faire référence. Dans les séries longues (c'est-à-dire, celles dont la longueur est égale ou supérieure à quatre items), les introducteurs précisent en règle générale le rang de l'item (*en troisième lieu, quatrièmement...*). Dans notre corpus, une seule série (sur 91) échappe à cette règle : {*D'abord, puis, ensuite, enfin* }.

5. Opérations, structures et configurations de marques

Les observations que je viens de livrer montrent que la sériation dans le discours est bien une opération complexe faisant appel à la fois au balisage des unités textuelles, à leur groupement et à leur mise en ordre. Mais, la nature discursive des items (et de leurs balises) intègre également une autre dimension qui, elle, est proprement langagière. En effet, sur les opérations de caractère purement organisationnel, peuvent se greffer des relations argumentatives d'étayage s'établissant entre le segment amorce et les items de la série, ou, dans le cas des séries binaires, des rapports sémantiques d'opposition, de concession, de hiérarchisation ou conséquence entre leurs items. Les marques de ces relations sémantiques ou argumentatives peuvent alors endosser le rôle d'organisateur discursifs.

La mise en évidence de ces différentes opérations discursives constitutives de la sériation (ou compatibles avec cette structure) permet de dégager plusieurs structures apparentées entre elles, mais sémantiquement et morphologiquement différentes. A chaque type de structure correspond un ensemble différent d'opérations, ce qui se traduit par une configuration de marques différente. C'est seulement à l'intérieur d'une configuration de marques propre à une structure particulière, qu'il est possible de chercher à distinguer des indices déclencheurs et des indices complémentaires, comme le préconise la méthode de l'exploration contextuelle que nous avons adoptée pour formaliser nos données. Rappelons que cette méthode consiste à identifier dans un texte, en premier lieu, les unités linguistiques appelées *indices pertinents* ou *indicateurs* qui sont significatifs pour la caractérisation (étiquetage) des segments en fonction d'une tâche donnée, et ensuite de rechercher, avec des règles spécifiques, des *indices complémentaires* dans le contexte des indices pertinents, indices complémentaires qui vont permettre la prise de la décision adéquate.

Quatre types de structure en série ont été distingués. La première, prototypique, fondée sur la composition des opérations de balisage, de groupement et d'ordre, présente sur le plan de la signalisation linguistique deux variantes. Dans le cas où la série contient une amorce explicite, c'est précisément cet élément qui constitue l'indice déclencheur, les MIL apparaissant comme des indices complémentaires, non tous nécessaires. Dans le cas où il n'existe pas d'amorce, c'est l'occurrence du MIL de rang un ou deux qui constitue le déclencheur, les autres MIL et les marques de rétro-évaluation ayant le statut d'indices complémentaires.

La deuxième structure, qualifiée d'argumentative, a ceci de particulier qu'elle ne contient pas d'amorce et que chaque item est précédé (en plus d'un MIL ou à sa place) d'une marque argumentative (*parce que, en raison de*). Le groupement n'est opéré qu'*a posteriori*, car ce qui réunit les items entre eux c'est le fait de former plusieurs arguments distincts en faveur d'une même thèse.

Le troisième type de structure en série contient l'amorce et un seul item jugé le plus important ou le plus représentatif. La notion d'ordre entre items n'a pas de raison d'être. L'indice déclencheur est constitué par l'introducteur du premier item qui le plus souvent est *en tout premier lieu*. L'amorce est l'indice complémentaire important, car c'est elle qui informe de l'existence de plusieurs éléments qui forment effectivement la série. Le deuxième indice complémentaire est l'absence de MIL de rang égal ou supérieur à deux.

La quatrième structure, marginale en tant que série mais très fréquente dans les textes, est constituée de deux éléments nettement différenciés entre eux (opposés, hiérarchisés...) (voir l'exemple (4), §3.3). Ce n'est donc pas un principe de similitude qui réunit de prime abord ses items, d'où l'absence de relation de groupement. Le rang des items n'étant pas pertinent, seul importe le besoin de les différencier entre eux. Sur le plan de l'expression linguistique, cette structure ne contient ni d'amorce, ni de rétro-évaluation, mais, deux balises, dont la deuxième marque une différenciation forte, le plus souvent une opposition (*d'une part – mais, de l'autre...*).

Selon le type de structure, un élément donné peut jouer des rôles différents. Ainsi, dans la structure [1], c'est l'amorce qui joue le rôle du déclencheur. Dans la structure [3], elle n'est qu'un indice complémentaire, dans la structure [4], c'est son absence qui constitue un indice.

En résumé, la tâche d'identification des séries a mis à l'épreuve les principes de la méthode d'exploration contextuelle destinée à formaliser les connaissances linguistiques relatives à cette structure. Il a été montré qu'il n'est pas possible d'exhiber un seul type de marque comme indice déclencheur pour l'identification des séries. Cette structure admet plusieurs variantes, résultant chacune d'une manière particulière dont se composent les opérations organisationnelles (balisage, groupement, mise en ordre) et certaines opérations sémantiques ou argumentatives (opposition, étayage...). A chaque ensemble d'opérations correspond une configuration de marques particulière, au sein de laquelle on distingue des indices d'emblée saillants et des indices complémentaires.

6. Conclusion

Les marqueurs de sériation font partie des routines cognitives ayant pour but d'alléger le travail d'interprétation. Ces balises semblent parfaitement univoques. Comme nous venons de le voir, leur évidence est trompeuse.

Analyser des configurations « mouvantes » de marques, témoins des intrications complexes d'opérations discursives multiples, mais non toutes nécessaires, en intégrant de plus la dimension des usages et des normes, n'est pas chose facile. Cela ne devient possible qu'avec l'aide d'outils permettant de recueillir des données, de les quantifier et d'aider à les analyser par des jeux de paramètres appropriés aux phénomènes étudiés.

Et c'est bien un processus en boucle qui se met en route : (i) observation des séries collectées « à la main » et choix des contributions pertinentes dans la littérature, (ii) première expérience d'extraction sur corpus et analyse des résultats, (iii) approfondissement théorique de la notion (si nécessaire), (iv) deuxième expérience d'extraction sur corpus et analyse des résultats... C'est un cercle « vertueux » où l'analyse linguistique aide à la mise en place d'un outil informatique sur mesure, lequel facilitant la collecte des observables contribue à son tour à alimenter cette même analyse.

Il faut se rendre à l'évidence que cette situation idéale n'est pas toujours possible. Il y a un temps nécessaire et incompressible pour pouvoir avancer sur un tel projet d'étude. Or les temporalités de la recherche ne sont pas toujours propices à des projets cumulatifs, en plusieurs étapes. Pour ce qui concerne ce travail sur la sériation, la réalisation de l'étape 4 n'a pas été possible, du fait très prosaïque de la reconfiguration de l'équipe de recherche.

III. DISCOURS RAPPORTÉS

A. Etat des lieux

Les études consacrées aux modalités de prise en charge énonciative dans l'expression de la causalité m'ont sensibilisée à l'importance de celles qui attribuent l'information à une autre source de savoir ou de dire.

Une information causale n'a clairement pas la même valeur, et de ce fait pas le même intérêt pour un utilisateur donné, suivant qu'elle est prise en charge par l'auteur du texte ou attribuée à un autre locuteur avec, par exemple, une marque de distanciation de la part de l'auteur. Plus généralement, nous avons vu qu'une telle information pouvait s'inscrire dans un cadre (univers) déterminé (de connaissance, de croyance, spatial, temporel...). Identifier et traiter automatiquement des segments textuels (en tant que définition, événement, séquence causale...) demande de tenir compte de ce « contexte de validité ». Il est important de savoir extraire de l'information ainsi contextualisée pour être en mesure de l'interpréter correctement et pouvoir statuer sur sa légitimité et sa valeur opératoire.

La problématique des discours rapportés, se trouvant au cœur des recherches sur l'énonciation et apte à fédérer des réflexions sur le discours, son organisation et ses usages sociaux, m'est apparue comme extrêmement intéressante d'un point de vue linguistique, et particulièrement importante, voire stratégique pour des travaux appliqués en TAL. En effet, la palette des applications concernées directement ou indirectement par le discours rapporté est large, on peut mentionner entre autres : (i) la mesure de l'impact de la production d'un chercheur (mesures bibliométriques) (Hirsch 2005), (ii) les systèmes question-réponse (*que pense X au sujet de Z ?*) (Somasundaran *et al.* 2007), (iii) l'amélioration des performances des systèmes de résumé automatique (Farzindar *et al.* 2004). L'engouement pour le discours rapporté se trouve confirmé par deux autres courants de recherches appliquées, très porteurs en ce moment : celui de la fouille des opinions et celui qui se préoccupe de la maîtrise des informations diffusées.

Les citations véhiculent souvent des opinions jugées significatives, des avis ou des informations ayant un impact social. De ce fait, elles peuvent répondre à des préoccupations fort diverses : suivi de l'opinion politique, des intérêts culturels, maîtrise de l'image et la réputation, etc. Depuis quelques années, ce besoin d'identifier des contenus subjectifs exprimés sur le Web ou dans des bases textuelles conduit à la réalisation d'applications informatiques dites de fouille d'opinions, d'analyse d'attitude, de sentiments ou d'émotion (Wiebe *et al.* 2005). Or, on constate que les stratégies (de repérage, d'analyse ou d'extraction) effectivement mises en œuvre sont souvent très empiriques et ne font que rarement et très indirectement le lien avec des travaux linguistiques touchant à l'énonciation. Cette question sera abordée dans la dernière partie de ce document.

Le deuxième courant réunit les applications concernées par la maîtrise de l'information diffusée, comme la détection de plagiat ou le suivi d'impact d'une information. Ces problématiques ont été étudiées dans le cadre du projet ANR PIITHIE²², porté par la société Sinequa en collaboration avec le laboratoire LINA de Nantes et la société Advestigo. Dans le contexte de la recherche du plagiat, (Poulard *et al.* 2008) proposent un système de repérage des citations directes et indirectes, en identifiant la source et son propos. La problématique de l'identification des textes dérivés à partir d'un texte source est abordée dans (Poulard *et al.* 2010). Les auteurs définissent un cadre théorique posant les concepts de la dérivation et proposent un modèle multidimensionnel cadrant les différentes formes de dérivation.

Rappelons que les premiers travaux en TAL sur le repérage automatique de citation en français sont ceux de (Mourad et Desclés 2004) et de (Giguet et Lucas 2004). Ils s'attachent à reconnaître des schémas réguliers et essentiellement phrastiques des discours rapportés directs et indirects. Dans le monde anglo-saxon, (Teufel *et al.* 2006) se sont intéressés de manière plus ciblée à des citations de travaux autres dans des articles scientifiques.

²² PIITHIE (Plagiats et Impacts de l'Information Textuelle recHerchée dans un contexte InterlinguE) <http://www.capdigital.com/piithie/>

Ces différentes problématiques appliquées placent la citation au cœur des études, contrairement à celles où elle ne sert qu'à contextualiser une autre information, en les orientant de plus en plus vers les phénomènes de circulation. Les discours que l'on cherche à analyser aujourd'hui sont ceux qui participent des nouvelles formes de communication. Les nouvelles pratiques discursives sur le Web (blogs, wikis, forums...) diversifient les modes d'emprunt à des discours d'autrui, dont la réitération, plus aisée et plus rapide, multiplie les horizons énonciatifs et les trajectoires argumentatives de ces discours. Ceci en tension permanente entre le devoir de fidélité à leurs emplois d'origine et la recherche d'effets rhétoriques inédits *via* des transformations parfois substantielles. Ces écrits du Web offrent un champ d'observation illimité, notamment pour des recherches pluridisciplinaires. C'est dans une telle optique que j'ai réalisé une première étude sur la détection automatique de formules politiques à fort impact, en collaboration avec Emmanuel Quintin dont la compétence première est la science politique (Quintin *et al.* 2011).

Les travaux d'obédience linguistique et littéraire sur le discours rapporté ont une longue tradition. Il existe de nombreuses descriptions, minutieuses et abondamment exemplifiées, portant sur les différents modes de citation, les mécanismes de transposition entre discours, les formes de dialogisme. Nous les devons, sans être exhaustive, à (de Gaulmyn 1983), (Ducrot 1984), (Authier-Revuz 1995), (Charaudeau 1997), (Rosier 1999, 2008), (López Muñoz *et al.* 2004), (Bres *et al.* 2005), et plus généralement à la vaste communauté Ci-dit. Des travaux plus formels sur les verbes de citation et les aspects syntaxiques de l'incise citationnelle ont été conduits par Olivier Bonami et Danièle Godard (Bonami et Godard 2008) et dans le contexte du TAL par Laurence Danlos, Benoît Sagot et Rosa Stern (Danlos *et al.* 2010).

Le champ d'études du discours rapporté s'est progressivement étendu en partant des approches philologiques et linguistiques vers une approche discursive sensible aux genres de discours et aux pratiques sociales associées. De plus, depuis quelques années, la problématique du discours rapporté s'élargit à l'étude de la parole qui fait l'objet de plusieurs transmissions. On s'intéresse aux configurations énonciatives des discours circulants et au vocabulaire qui leur est associé. On interroge également les mécanismes et les enjeux de la propagation des discours (López Muñoz *et al.* 2010). Sans oublier des questionnements plus transversaux sur ce que ces discours peuvent révéler sur notre société (construction du sens commun, reproduction de normes socio-culturelles...). « Le discours en circulation apparaît comme l'expression d'une conscience collective », pour reprendre les mots de (Rabatel 2004, p.8). Cette question de circulation des discours devient aujourd'hui un nouveau point de convergence entre travaux linguistiques et travaux appliqués en TAL.

Mes propres recherches sur le discours rapporté ne sont qu'une incursion dans ce vaste champ de recherches. Trois problématiques ont été abordées.

Dans un premier temps, j'ai entrepris un travail sur la construction d'une typologie des discours rapportés linguistiquement plausible et exploitable par des applications informatiques. Afin de pouvoir rendre compte de l'hétérogénéité constitutive du DR, je me suis concentrée sur la description des opérations énonciatives et sémantiques constitutives des différents modes de citation. Ce travail a été réalisé en analysant les marques de l'activité citationnelle observée dans les articles de presse. Au moment où les recherches linguistiques françaises se saisissent de la problématique de l'effacement énonciatif²³, c'est donc, au contraire, un travail approfondi sur les introducteurs de DR et sur plusieurs autres indices de la citation qui est entrepris.

J'ai construit une typologie des discours rapportés (Jackiewicz 2005, 2006), fondée sur une double classification : celle des opérations énonciatives (restitution, authentification, interprétation, évaluation, explication) constitutives du DR et celle des rapports intersubjectifs entre différentes instances de parole impliquées dans le DR. Ce travail, mené sur corpus, a produit également un ensemble de ressources linguistiques associées : il s'agit des verbes et des adverbiaux énonciatifs, introducteurs de DR, ainsi que des marques de reformulation. Dans ce contexte, je me suis intéressée de manière plus approfondie à l'étude des formes de responsabilité et d'engagement propres aux pratiques journalistiques dans la presse d'information et d'opinion (Jackiewicz 2011). Cette étude est synthétisée dans les pages qui suivent. Le lecteur ne manquera pas de noter des similitudes d'approche avec le travail mené sur la causalité. Les actes d'attribution de cause et d'attribution de dire impliquent différentes formes d'engagement de la part de celui qui les réalise.

²³ Voir le numéro 156 de la revue *Langages*, « Effacement énonciatif et discours rapportés » (Rabatel, 2004).

Dans un deuxième temps, sur la base de cette première étude à visée descriptive et systématisante, j'ai proposé dans une optique plus formelle, en collaboration avec Philippe Laublet, une esquisse d'ontologie linguistique de l'énonciation textuelle, appelée OntoDiscours (Jackiewicz et Laublet 2007). Cette ontologie est conçue pour annoter des segments textuels conjointement ou non avec une ontologie d'un domaine. Six classes de spécialisation permettent d'annoter finement : (i) le niveau énonciatif (direct, rapporté), (ii) les actants et les circonstants de l'énonciation, (iii) les constituants fonctionnels, (iv) les constituants modaux, (v) les éléments paratextuels, ainsi que (vi) les marques de surface de ces différents objets. L'ontologie ne se limite pas à une simple catégorisation hiérarchique. Elle contient également des relations destinées à relier les instances des classes qu'elles soient ou non le support d'annotation. Par exemple la classe « Discours_Rapporté » possède quatre relations : locuteur, allocutaire, date et lieu. Par héritage, tel discours rapporté direct pourra avoir au moins un locuteur précisé, par exemple *Eva Joly*, un lieu comme *Paris*. *Eva Joly* et *Paris* peuvent correspondre à des annotations de segments textuels du même texte ou être simplement des objets de la base de connaissance, non présents explicitement dans le texte. Le démonstrateur, construit sur Protégé et Knowtator, permet d'illustrer les apports de l'approche. Un tel modèle peut être utilisé dans des recherches de corrélations entre tel et tel aspect discursif ou énonciatif sur des corpus différenciés. Cela permet de construire de « micro » grammaires sémantiques, caractéristiques de certaines visées communicatives liées à des contextes précis. Un exemple de tel pattern récurrent serait un discours rapporté évalué et reformulé.

Enfin, plus récemment, nous avons proposé une méthode itérative de repérage d'énoncés à fort impact (Quintin *et al.* 2011). Cette recherche vise la construction d'outils informatiques permettant de répondre à deux besoins clairement identifiés aujourd'hui : (i) identifier des discours en circulation (repris, commentés, critiqués, détournés...), (ii) évaluer et analyser l'impact d'une citation particulière.

Ce travail suit une démarche itérative, partant de l'analyse des traits caractéristiques d'un énoncé à fort impact. En première approximation, il revient à l'analyste d'isoler ce type de formule. À partir d'exemples recueillis sur le net entre avril et juin 2009, nous avons constitué un corpus de soixante énoncés composé de posts, de messages et de billets autour d'une phrase devenue célèbre, énoncée en 1990 par Michel Rocard : « *La France ne peut accueillir toute la misère du monde, mais elle doit savoir en prendre fidèlement sa part* ». L'impact des formules est corroboré par des observations quantitatives et qualitatives sur corpus permettant d'attester de la circulation. Le procédé mis en œuvre s'appuie sur une exploitation systématique des formules contextuelles (« réplique qui passera à la postérité » ; « célèbre réplique » ; « phrase qui restera dans la mémoire collective » ; « une expression (que nous a) léguée... » ; « phrase culte », « déjà culte version » ; « polémique » ; « version qui cartonne » ...). Nous analysons également finement les transformations que subit le discours qui circule. Enfin, nous tentons de systématiser les objectifs communicationnels et représenter les horizons énonciatifs de ces discours. L'étude de faisabilité et le prototype associé ont été réalisés avec la plate-forme SemioLabs développée par la société Noopsis de Caen.

Ce type d'étude permet d'interroger les nouvelles pratiques énonciatives sur le Web autour des énoncés répétés. Le Web constitue un laboratoire des fonctionnements discursifs et permet d'observer l'évolution dynamique (échanges d'opinion) et statique (reprise d'énoncés) des représentations culturelles : le processus de stéréotypage au sens de (Amossy 2010). Les nouveaux médias sont en effet très fortement doxiques. Outre cet objectif, l'étude peut avoir comme prolongement l'évaluation du degré de figement du discours politique. Les études sur la stéréotypie et la « langue de bois » sont déjà nombreuses, citons (Pineira et Tournier 1990) entre autres. Adéquatement outillées et menées sur des corpus actuels, en grand volume, construits selon des paramètres précis, elles pourront livrer une image plus précise et plus complète des mécanismes à œuvre. Une autre piste permettant de prolonger cette étude serait de mesurer l'impact réel d'un personnage politique à travers celui de ses énoncés. Ces études offriraient l'avantage de pouvoir exploiter à la fois les méthodes qualitatives et quantitatives.

B. Responsabilité et prise en charge dans le discours rapporté

1. Problématique

Le discours rapporté est à la fois pratique de discours, pratique intersubjective et pratique sociale. Dans de nombreux contextes sociaux, traiter des discours rapportés pose de manière forte la question des responsabilités et manipulations des discours d'autrui. Cette question touche aux différentes contraintes de littéralité ou de concision, de l'identification, l'opacification ou la dilution de la source énonciative (Rosier 2005). Elle concerne aussi, ce qui échappe souvent aux analystes, les postures évaluatives et explicatives des énonciateurs en charge de la transmission de discours.

Dans les pages qui suivent, je synthétise une étude qui cherche à systématiser les formes de responsabilité dans les citations journalistiques.

Je propose d'aborder la notion de responsabilité à travers l'étude des relations intersubjectives s'établissant entre plusieurs instances énonciatives dans et à travers le discours rapporté (DR). Plus précisément, étant donnée une séquence de discours rapporté, explicitement attribuée à un tiers, la question est de savoir quelles sont les modalités de l'intervention de l'énonciateur (celui qui cite) dans ce même discours, quels en sont les lieux d'inscription et les marques linguistiques correspondantes.

Sans être l'objet même de l'analyse, la responsabilité occupe une position centrale dans ce travail dans la mesure où elle constitue, comme je m'attache à le montrer, un moyen de classification opératoire pour les différentes séquences de discours rapportés que nous cherchons à décrire. Inversement, je défends l'idée que l'examen attentif de discours rapportés extraits d'un corpus de presse fournit en retour des éléments utiles permettant de mieux cerner l'étendue des actions énonciatives dont la responsabilité incombe à l'énonciateur.

Mon propos n'est pas de théoriser la notion de responsabilité en linguistique, mais d'en décrire les manifestations dans un contexte où elle apparaît comme particulièrement saillante. A minima, il s'avère éclairant de la situer par rapport à la notion de prise en charge énonciative, mieux décrite, tout en mesurant la difficulté de se confronter à une notion particulièrement complexe. En effet, les différentes facettes de la prise en charge analysées finement, notamment par la ScaPoLine (source, point de vue, responsabilité, jugement, accord...), Ducrot (construction d'un jugement, acquiescement à un jugement), Laurendeau (prise en charge, prise en compte) ou Desclés (énonciation « simple », différents types d'assertions) ne se laissent pas articuler dans un modèle unique et cohérent.

2. Choix préliminaires

2.1 Observables

Explorer la dimension intersubjective des pratiques citationnelles est une tâche difficilement envisageable sans tenir compte des usages réels de la citation. J'ai choisi d'examiner des séquences citationnelles telles qu'elles fonctionnent dans des situations réelles et clairement déterminées de communication réalisée via le discours rapporté. A cet effet, j'ai pris comme corpus de référence deux ans d'articles du *Monde* sur CD-ROM (1^{er} juillet 2002 - 30 juin 2004). Le guide rédactionnel « *Le style du Monde* » (*Le Monde* 2002) fournit des indications précieuses sur les recommandations que doivent suivre les journalistes en matière de citations, et sur les types d'articles et sur les caractéristiques qui leurs sont associées, notamment en ce qui concerne le rapport à la subjectivité et à l'expression des opinions. Du côté de la littérature linguistique, on trouve dans (Rabatel et Chauvin-Vileno 2006) et (Rabatel 2006), une étude fine et très éclairante des différentes facettes (juridique, morale, linguistique...) de la responsabilité dans l'écriture de la presse.

« Les discours de presse offrent des *corpora* d'un grand intérêt pour traiter de la question de la responsabilité, tant ils se caractérisent par une double contrainte plus ou moins réglée selon les genres : ils doivent d'une part préciser certaines sources, d'autre part en protéger certaines autres, s'astreindre à une certaine neutralité dans la présentation des opinions

comme dans celle des faits, bref, osciller entre la nécessité de présenter les points de vue des autres (ou de certains autres) et l'obligation de ne pas donner le point de vue personnel des journalistes. Fort ancrage énonciatif dans un cas, effacement énonciatif dans l'autre sont mis au service d'une déontologie qui repose sur le trépied de l'impartialité, de la neutralité et de l'objectivité. Cette double contrainte pose de redoutables questions en matière de responsabilité». (Rabatel et Chauvin-Vileno 2006, § 22).

La mission des médias d'information et d'opinion, tels que *le Monde*, n'est pas seulement de transmettre ou reproduire des paroles qui circulent dans l'espace social (être un miroir social), mais également (ou surtout) de les rendre intelligibles et de montrer l'impact qu'elles produisent dans cet espace (médiation et instruction). Les pratiques citationnelles servent principalement à rendre compte des interactions qui s'opèrent dans et à travers l'activité linguistique des sujets communicants. Et cela d'une manière doublement responsable. Tout d'abord, face aux locuteurs d'origine, dont les propos doivent être restitués, sans être altérés, en mettant en évidence leurs aspects saillants, et notamment les effets qu'ils produisent sur leurs destinataires (ou allocutaires). Ensuite, face aux lecteurs (co-énonciateurs), en anticipant sur leurs besoins en information et sur les connaissances éventuelles dont ils auraient besoin pour comprendre et interpréter correctement les paroles citées²⁴. C'est sur ces rapports intersubjectifs que nous concentrerons notre attention.

2.2 Séquences citationnelles

Tout discours rapporté est par essence recontextualisé, c'est-à-dire détaché de la situation d'énonciation initiale et inséré dans un autre discours. Le contexte d'accueil qui entoure les paroles rapportées proprement dites demande à être exploré pour livrer les éléments qui rendent interprétables ces paroles. Ces éléments contextuels n'appartiennent pas nécessairement à l'unité phrastique qui comprend le discours reproduit (1).

(1) « *Il y a désormais deux sortes de logements sociaux : ceux que l'on démolit et ceux qui sont mis en vente* ». Colportée, **sous le sceau de l'anonymat, par un observateur, la formule, délibérément provocatrice, résume** les inquiétudes suscitées par la politique du gouvernement en matière d'habitat. (*Le Monde*, 25 mars 2004)

L'objet « citation » est clairement une entité discursive multiforme, potentiellement composée de segments fonctionnellement hétérogènes (comme on le verra dans le point 4), non tous nécessaires, mais participant chacun à l'intelligibilité de l'ensemble.

La visée ultérieure d'extraction de citations conduit à privilégier l'examen des séquences citationnelles que l'on peut qualifier de « complètes » et d'autonomes. Cela pose la question de la délimitation de cet objet multiforme. Les marques linguistiques pertinentes pour l'étude correspondent donc à des configurations d'indices où les expressions introductrices de DR interagissent avec d'autres éléments contextuels également impliqués dans les repérages énonciatifs de la citation (voir le §2.4).

2.3 Ressources linguistiques disponibles et acquises

J'ai privilégié et exploité les travaux qui mettent clairement l'accent sur le signalement de l'activité citationnelle en surface du texte. Concernant les ressources linguistiques pré-existantes et exploitables dans notre étude, j'ai retenu la liste des verbes de parole constituée par (Lamiroy et Charolles 2008), qui a été croisée avec les « prédicats de parole » de (Eshkol 2002) et les verbes performatifs tirés de (Vanderveken 1988). Les verbes de parole sont les principaux marqueurs aptes à introduire des paroles étrangères. Cette liste de verbes a été complétée par une collecte sur corpus. En même temps, ont été prélevées sur corpus des expressions adverbiales de manière qui participent également à l'interprétation sémantique des citations (*doctement, d'un ton ironique, avec une emphase péremptoire...*), ainsi que des marques de reformulation (*en clair, autrement dit...*), au sens de (Rossari 1997).

²⁴ « Les lecteurs occupent une place centrale dans les préoccupations de la rédaction du *Monde*. Le goût de la pédagogie, la nécessité de hiérarchiser l'actualité, l'honnêteté et l'effort constant pour équilibrer le traitement de l'information sont tenus pour qualités indispensables. Il s'agit de faire partager aux lecteurs l'information, les faits et les chiffres les plus fiables ; de leur rendre plus lisible un univers souvent obscur. » (*Le Monde* 2002 : 7).

3. De la prise en charge à la responsabilité

3.1 La prise en charge énonciative et les notions affines

A la fois intuitive et complexe, la notion de prise en charge est généralement analysée ou expliquée à l'aide d'autres notions jugées plus élémentaires.

C'est la notion de vérité qui constitue le point d'attraction vers lequel convergent la majorité des travaux consacrés à la prise en charge. Nous devons à (Culioli 1980) une définition de référence qui donne le « sens technique de prendre en charge : dire ce que l'on croit (être vrai). Toute assertion (affirmative ou négative) est une prise en charge par un énonciateur ».

Pour (Desclés 2009), la prise en charge énonciative représente au sens large, la dimension modale (modus) emboîtant la dimension prédicative (dictum) de l'énoncé. Son schéma général de l'énonciation « JE-DIS (ce qui est dit) » traduit l'idée que tout énoncé est pris en charge par son énonciateur. Dans ce schéma, *JE* désigne l'énonciateur, *JE-DIS* l'opérateur métalinguistique de prise en charge de « ce qui est dit », « *ce qui est dit* » le dictum (lexis, proposition, ou relation prédicative). L'opération complexe de prise en charge peut, selon (Desclés et Guentchéva 2000), se décomposer en plusieurs opérations élémentaires : « Je dis que ce qui est dit est vrai (assertion) », etc²⁵. La notion de vérité, centrale dans les trois premières opérations (parmi lesquelles l'assertion fait figure d'opération prototypique), n'est pas pour autant absente des suivantes. Elle y est implicite, car il s'agit de la vérité pour quelqu'un ou de la vérité hypothétique (celle qui reste à établir).

Les travaux de Laurendeau (Laurendeau 1989, 2009) sur la prise en charge préservent le rôle fondamental de l'assertion. Les différentes valeurs de l'assertion (assertion polémique stricte, assertion constative stricte, désassertion constative stricte, désassertion polémique stricte) qu'il pose comme « inhérentes à une compréhension globale et dialectique de la question de prise en charge » l'amènent à introduire la notion de *prise en compte*. L'analyse de la « tension prise en compte / prise en charge », révèle les différents positionnements possibles des interlocuteurs interagissant face à un contenu prédicatif donné (exemples : *S₀ seul prend en charge ce que S₀' prend en compte ; S₀ prend en charge ce que S₀' ne prend qu'en compte, etc.*).

La vérité reste importante dans les travaux qui cherchent à articuler la prise en charge avec la notion d'engagement. Pour (Desclés 2009), prendre en charge c'est s'engager sur la vérité de ce qui est dit, mais il est également possible de s'engager par un acte performatif « en transformant le monde externe ». L'engagement, à son tour, n'est pas défini, mais mis en rapport avec le « non engagement », l'« engagement non négociable » et le « désengagement négociable ».

La ScaPoLine (Nolke *et al.* 2004), qui comme on le verra plus loin, propose une construction théorique basée sur la responsabilité, n'évacue pourtant pas la vérité. « [...] dans la ScaPoLine, la plupart des jugements sont du type vrai/faux – simples ou modalisés (p.ex. *généralement vrai, peut-être vrai*). On peut en conclure que la définition de la responsabilité en terme de source n'est peut-être qu'un « détour », qui retrouve au final la notion de vérité » s'interrogent (Dendale et Coltier 2005 : 137).

C'est la ScaPoLine qui a théorisé la notion de responsabilité, dont la prise en charge est, dans ce cadre, synonyme. La responsabilité renvoie au cas où un énonciateur émet (est source de) un point de vue, c'est-à-dire formule un jugement à propos d'un contenu. Le lien de responsabilité unit une source à un point de vue. (Dendale et Coltier 2005) montrent que cette définition, à première vue satisfaisante, est à plusieurs titres problématique. Son statut opératoire dépend de la manière de définir à leur tour les notions de source, de jugement et de contenu. Or ces notions s'interdéfinissent les unes par les autres : la source est dite avoir un point de vue, c'est-à-dire un jugement sur un contenu, les jugements étant caractérisés par quelques valeurs (les plus typiques ?) qu'ils peuvent prendre. On peut remarquer aussi que la notion de responsabilité, telle qu'elle est définie dans la ScaPoLine, perd une part importante du sens qu'elle véhicule dans les emplois courants conformes au sens commun. Nos interactions sociales quotidiennes sont constamment saturées par les modalisations par lesquelles nous nous trouvons assignés et rappelés à la responsabilité, du genre *tu n'aurais pas dû, tu l'as voulu, tu pouvais le faire, tu ne le savais pas, pourtant...* Etre responsable, c'est aussi (ou surtout ?) avoir à répondre devant l'autre.

²⁵ Voir dans ce même document, le chapitre « La causalité prise en charge » au paragraphe 3.1.

Enfin, un point notable est que tous les travaux cités précédemment utilisent explicitement ou implicitement (notamment dans les exemples illustratifs d'interactions) la notion d'accord (au sens d'acquiescement ou d'adhésion). L'accord renvoie à un positionnement par rapport à une assertion : « *cette proposition est vraie, je suis d'accord, cette autre ne l'est pas, je suis en désaccord* ». L'attitude caractéristique à l'égard d'un point de vue pris en charge est de pouvoir se montrer en accord ou en désaccord avec ce point de vue.

3.2 La responsabilité et le jugement de responsabilité

Constitutive de la vie sociale humaine, la responsabilité s'applique à des personnes morales, capables de comprendre certaines exigences, attentes et règles et d'agir sur la base de cette compréhension. De plus, la responsabilité renvoie à une disposition à répondre, comme l'explique (Chrétien 2007). Se montrer responsable, c'est être disposé à répondre de quelque chose et devant quelqu'un (un tribunal, des témoins de son engagement, ses pairs, l'humanité, Dieu...). La responsabilité est donc une notion relationnelle, car elle fait appel à l'autre qui est le pôle à partir duquel se construit la validité normative sous-jacente. Tenir quelqu'un pour moralement responsable d'un bienfait ou d'un dommage, c'est avoir envers lui une attitude d'approbation ou de désapprobation sur la base d'un jugement. Appelons, suite à (Neuberg 1997), « jugement de responsabilité » le jugement qui s'il est correct, justifie nos attitudes d'approbation ou de désapprobation. Ce jugement peut être remis en question : un reproche peut s'avérer infondé, un blâme injuste, une louange injustifiée, des félicitations précipitées. Le jugement de responsabilité ou l'anticipation responsabilisante imputent sur le mode performatif les qualités subjectives qui font d'autrui un agent moral, sujet de ses actes, susceptible d'en répondre. Cette attitude caractéristique d'approbation ou de désapprobation face à un acte responsable est clairement différente de celle qui pourra s'exprimer en termes d'accord ou de désaccord face à l'énoncé d'un point de vue pris en charge.

3.3 Le discours rapporté : actions et responsabilités de l'énonciateur

Sur le plan linguistique, il ne s'agit pas de savoir si les locuteurs sont responsables de tout ce qu'ils disent, Le cas des discours rapportés par des médias d'information est plus directement analysable, même s'il n'échappe pas à certaines difficultés, liées notamment aux marges de manœuvre des différences instances productrices de discours que nous ne sommes pas en mesure d'apprécier.

Si, pour simplifier, nous sommes moralement responsables des actes qui affectent directement ou indirectement l'autre, s'emparer du discours d'un tiers pour parler avec, le transmettre, l'expliquer ou le critiquer est un acte nécessairement responsable, car il concerne à la fois son locuteur d'origine que celui à qui l'on s'adresse, tant les choix liés à la sélection et à la recontextualisation critique de ces paroles sont chargés d'impact, particulièrement quand ces actions sont réalisées dans l'espace public. Par conséquent – et les extraits (2-4) en apportent une illustration – l'exercice de la citation dans les médias est *a priori* ouvert à un jugement public, et soumis à l'obligation de rectification ou de réparation (tout aussi publique), si ce jugement met en lumière un préjudice.

(2) *RECTIFICATIFS. NUCLÉAIRE. Nous avons attribué à Jacques Munier, directeur délégué de la centrale nucléaire de Flamanville (Le Monde du 19 février), des propos qu'il n'a pas tenus. Il expliquait bien que les contrats entre EDF et ses sous-traitants étaient conclus sur la base d' « accords gagnant-gagnant ». En revanche, la deuxième partie de la phrase – « qui peuvent toutefois se révéler perdant-perdant pour les salariés qui se retrouvent sur la touche » - était un commentaire du journaliste, et non une citation de M. Munier. (Le Monde, 23 février 2004)*

(3) *La citation de Rocard. Dans l'éditorial du Monde daté du 22 juin, vous citez les propos de Michel Rocard : « La France ne peut pas accueillir toute la misère du monde. » Encore eût-il fallu que cette affirmation fût correctement rapportée. [...] M. Rocard a précisément déclaré, devant la Cimade, en 1990 : « La France ne peut accueillir toute la misère du monde, mais elle doit savoir en prendre fidèlement sa part », ce qui revêt une tout autre signification. C'est d'ailleurs le « destin imprévisible » de cette phrase qui a conduit son auteur à une mise au point dans vos colonnes sous le titre « La part de la France » (Le Monde du 24 août 1996). En vain, semble-t-il, car cette citation, perpétuellement tronquée, reprise sans précaution ni vérification, semble être devenue un postulat indiscutable ; elle sert désormais d'argument d'autorité à ceux qui la propagent et s'en détachent opportunément, pour paraître plus modérés que son auteur. » (Le Monde, Repères, 27 juin 2003)*

(4) *AU Courrier DES LECTEURS (...) A propos de la politique de la droite sur la RTT, vous reprenez l'expression gouvernementale « assouplissement des 35 heures » (Le Monde du 10 septembre). Il n'est pas dans le rôle d'une presse dite « libre » de répercuter les euphémismes du pouvoir, car il est clair qu'il*

s'agit ici d'autre chose que d'un assouplissement. On ne vous demande pas de parler de « suppression » ou de « remise en cause » de la baisse du temps de travail, mais simplement de ne pas faire écho au discours politicien qui, par définition, déforme la réalité. Pour remédier à ce problème, vous pourriez par exemple utiliser des guillemets ; trouver une expression plus neutre ; ou encore choisir un terme plus proche des intentions réelles, et non pas affichées, du gouvernement. (Le Monde, 13 septembre 2002)

Dans le discours rapporté, l'action fondamentale de l'énonciateur est de signaler que ce discours provient d'un autre locuteur ou, plus précisément, est issu d'une autre énonciation, car il est possible de se citer soi-même.

J'emprunte à (Desclés et Guentchéva 2000), le schéma général du discours rapporté : JE ('énonciateur') DIS < X ('locuteur') DIT « ce qui est dit » > où le *locuteur* désigne l'énonciateur qui prend en charge directement la relation prédicative « ce qui est dit ». Dans ce schéma, l'énonciateur (JE) attribue au locuteur (X) le contenu prédicatif exprimé (« ce qui est dit »). C'est cet acte d'attribution qui est pris en charge par l'énonciateur rapportant. En tant que tel, il peut être mis en question : *c'est faux, non, X n'a jamais dit ça ou tu te trompes, ce n'est pas X qui l'a dit, mais Y* ou, au contraire, confirmé : *c'est bien vrai, seul X est capable de dire une chose pareille...* Attribuer un propos à une source (5) peut constituer une information en soi.

(5) **DEVINETTES : Qui prétend ces temps-ci que « la France a besoin d'immigrés ».** Noël Mamère ? Non, Nicolas Sarkozy. **Quel discours récent met en avant la « lutte contre toutes les formes de discrimination », sexisme et homophobie inclus ? Celui de Marie-George Buffet ? Non, la citation est de Jacques Chirac. Qui prône le passage à l'acte à propos du droit de vote des étrangers ? Vincent Peillon ? Non, Philippe Douste-Blazy. Qui vient de déclarer « ouvert » le débat refusé par la gauche depuis plus de dix ans sur la « double peine » ? Encore l'inévitable M. Sarkozy. (Le Monde, 30 octobre 2002, p.1)**

Ce schéma de base qui appelle un positionnement en termes d'accord/désaccord relève donc de la prise en charge énonciative.

Mais la tâche de l'énonciateur rapportant ne s'arrête pas là. Comme nous le verrons, l'action de citer peut se charger (ou s'enrichir) de plusieurs opérations différentes qui relèvent, quant à elles, de sa responsabilité. Résultant d'un choix, voire d'une véritable stratégie de communication, ces actions peuvent être appréciées (approuvées ou désapprouvées) de différents points de vue, la vérité des propos rapportés et l'adhésion à ses propos n'épuisant pas la question (6).

(6) **Avec sa clairvoyance coutumière, Ségolène Royal a bien résumé le problème, en déclarant : « Aux yeux des garçons, le string réduit les jeunes filles à un postérieur. » En d'autres termes, le string ne serait pas un problème si les garçons n'existaient pas, ou (hypothèse plus réaliste) s'ils étaient dépourvus d'yeux pour voir. (Le Monde, 17 octobre 2003)**

4. Responsabilité énonciative dans différents modes d'intervention de l'énonciateur rapportant

4.1 Paroles restituées

Commençons par des segments minimaux où l'énonciateur ne fait que signaler l'origine différente (car due à un autre énonciateur ou à autre situation d'énonciation) des paroles qu'il restitue. Ces paroles sont totalement coupées de leur contexte énonciatif d'origine. On peut gloser l'intervention de l'énonciateur par la formule : « Je dis que ce qui est dit a été dit par un autre, dont je n'indique pas l'identité ou je signale que son identité m'est inconnue. »

Selon les cas, l'énonciateur endosse la responsabilité du choix de ne pas indiquer la source de ces paroles ou celle d'avoir choisi de rapporter des paroles issues d'une source inconnue. Le « risque » encouru est que le lecteur ne dispose pas de connaissances suffisantes pour décoder la teneur réelle des paroles rapportées et mesurer leur effet rhétorique dans le nouveau contexte où ces paroles se trouvent intégrées. C'est le cas de l'extrait (6) :

(6) *Ce n'est pas l'avis des économistes qui jugent, dans leur majorité, que la BCE, malgré des maladroites de communication, un manque de transparence, et en dépit d'une certaine lenteur de réaction aux variations de la conjoncture, peut revendiquer, dans l'ensemble et en raison de sa relative jeunesse, un ... « bilan globalement positif » ! (Le Monde, 28 octobre 2003)*

Plusieurs cas peuvent être distingués, selon que le locuteur et le contexte d'énonciation d'origine (i) sont supposés connus du lecteur (co-énonciateur), faisant partie d'un savoir partagé (6) ; (ii) ne sont plus disponibles dans la mémoire collective ou ne sont pas importants, comme c'est le cas

d'un grand nombre de « participations » au sens de (Maingueneau 2004) (7) ; (iii) sont accessibles au lecteur dans un contexte plus large, il s'agit de citations minimales contenues dans des titres et dans des accroches, qui sont généralement issues des discours rapportés plus amplement dans le corps de l'article ; (iv) ne sont pas connus, c'est le cas des bruits et des rumeurs, qui dans la presse doivent être signalés comme tels (*selon certaines rumeurs...*) ; (v) doivent être protégés et donc rester anonymes, ce qui en général est également dit explicitement.

(7) *Malheur aux vaincus ? A l'évidence, la maxime ne s'applique pas aux principaux cadres du Parti des travailleurs (PT) du Rio Grande do Sul. (Le Monde, 3 janvier 2003)*

4.2 Énonciation restituée

Pour comprendre la teneur des paroles rapportées et juger de leur légitimité, il est souvent nécessaire de les authentifier, c'est-à-dire de connaître les principaux paramètres de l'acte d'énonciation, à savoir : qui a parlé ? où ? quand ? à l'attention de qui ? Dans la presse d'information, c'est le cas de la majorité des citations, qui restituent ainsi leur énonciation, attestant de cette façon qu'il s'agit d'un acte de parole effectif. En effet, l'interprétant comprend qu'il ne s'agit pas d'une reconstitution des intentions d'un tiers, qui peuvent être communiquées par d'autres biais (un dessin, un signe gestuel, une mimique...), mais d'un acte de parole (événement énonciatif) que ce dernier a effectivement réalisé.

L'Énonciation restituée peut être glosée de la manière suivante : « Je dis que ce qui est dit a été dit par un autre, dans une situation d'énonciation, dont je précise certains paramètres (locuteur, moment, lieu, allocutaire...) ».

Il est de la responsabilité de l'énonciateur de donner les coordonnées exactes et informatives. Différentes combinaisons de ces principaux paramètres énonciatifs peuvent être linguistiquement exprimées et chacun d'entre eux peut être spécifié d'une manière plus ou moins riche. Parmi ces coordonnées, le statut privilégié revient toutefois au locuteur, à la fois auteur de l'énoncé rapporté et source des repérages spatio-temporels et personnels au sein de cet énoncé. Individuelle ou collective, cette entité peut être présentée avec différents degrés de détermination, de manière à renseigner sur sa notoriété et par voie de conséquence sur la légitimité et l'impact possible de ses paroles (8).

(8) « Il est extrêmement difficile de faire une loi contre 70 000 médecins qui reçoivent trente patients par jour. » **La citation est de Konrad Adenauer, l'ancien chancelier allemand, et date des années 1960.** Quarante ans plus tard, elle pourrait s'appliquer à la situation française. Depuis les premières mesures sociales de 1930, jamais une profession comme celle des médecins libéraux n'a autant résisté au changement. (Le Monde, 20 janvier 2004)

4.3 Énonciation interprétée

Dans les textes de presse, on trouve souvent des séquences citationnelles composées, où les paroles restituées ne font qu'illustrer un discours plus large, dont le but communicatif devient intelligible grâce à un travail interprétatif de l'ensemble de la situation d'énonciation d'origine. L'énonciateur choisit de transmettre ce qu'il a identifié et reconstruit comme essentiel dans l'intention communicative manifestée par le locuteur.

La glose qui met en évidence la participation de l'énonciateur est : « Je dis que ce qui est dit a été dit/évoqué par un autre et j'en donne une interprétation fondée sur la connaissance que j'ai de l'ensemble du discours d'origine. »

Le travail interprétatif de l'énonciateur rapportant peut porter sur : (i) le rapport interpersonnel établi par le locuteur avec son allocutaire : la valeur d'illocution ou l'effet perlocutoire de l'acte illocutoire d'origine (*conseiller, ordonner, exiger, innocenter...*), (ii) les moyens d'expressions choisis par le locuteur : le style, le mode d'argumentation, le genre du discours dont la citation est issue (*argumenter, ironiser, dissenter...*), (iii) le rapport du locuteur au contenu de son propos : l'adhésion (rapport à la vérité), l'importance qu'il accorde à certains éléments, l'attitude affective (*affirmer, soutenir que, douter, déplorer, insister sur, se réjouir...*).

La responsabilité de l'énonciateur tient dans l'exactitude du décodage des intentions du locuteur, en particulier quand il s'agit de rapporter un discours oral. En effet, la moindre nuance de voix du locuteur peut faire basculer la valeur de son message d'une catégorie illocutoire vers une autre : s'agit-il d'une question, d'un conseil, d'un ordre ? La subjectivité du locuteur (jugement de valeur, implication affective...) doit trouver une lecture la plus fidèle et précise possible. Notons à ce

propos, la difficulté d'opérer des estimations justes dans des domaines où la quantification ne va pas de soi (force d'une conviction, degré d'importance...). De même, le décodage des émotions, on le sait, est loin d'être aisé.

Les citations interprétées ont pour moyen d'expression privilégié le mode narrativisé, suivi éventuellement d'un court extrait rapporté directement. Notons l'importance de marques qui soulignent le caractère objectif de l'interprétation, telles que *visiblement*, *ouvertement*, *manifestement*... Dans (9), les paroles rapportées illustrent, mais ne permettent pas *a priori* de déduire le fait que Bernard Tapie a pu être innocenté. Cette information engage entièrement le journaliste.

(9) *Les enquêteurs du SRPJ de Marseille, dans leur rapport de synthèse, **innocentent totalement** Bernard Tapie des accusations de Pierre Dubiton. « Les mouvements (achats-ventes) de joueurs ont été nombreux durant l'intersaison 2001-2002 et le mercato d'hiver, écrivent-ils, et, comme l'a reconnu M. Dubiton, aucun document engageant le club n'a été signé par Bernard Tapie. » (Le Monde, 14 juin 2003)*

4.4 Discours évalué

Dans certaines citations, les paroles rapportées sont assorties d'une évaluation explicite, incidente à ces paroles ou à l'ensemble du discours dont elles sont extraites et qu'elles illustrent. La responsabilité de ce jugement incombe entièrement à l'énonciateur.

La glose qui rend compte de l'activité de l'énonciateur est la suivante : « Je dis que ce qui est dit a été dit par un autre et j'en donne une évaluation, conformément aux règles (normes, lois ou principes) morales, esthétiques ou logiques en vigueur dans l'espace (social, politique, culturel, linguistique...) auquel j'appartiens (ou que je représente), espace que je partage éventuellement avec mon co-énonciateur. »

Le travail évaluatif de l'énonciateur rapportant peut porter sur : (i) l'attitude du locuteur face à son allocutaire (jugement le plus souvent d'ordre moral) *embobiner, tromper, insulter, endoctriner...* ; avec un brin de condescendance, non sans mépris, avec tact... (ii) l'adéquation des choix des moyens d'expression ou des modes d'organisation du discours au contenu du discours ou à la situation d'énonciation (jugement parfois esthétique, souvent praxéologique) *insinuer, enjoliver, broder, jargonner, divaguer, digresser, rebattre, rabâcher ...* ; dans une drôle de tautologie /langue de bois/... ; en forme de lapalissade... (iii) le positionnement du locuteur face au contenu exprimé (jugement logique focalisé sur la vérité) *prétendre, fabuler, spéculer...* ; avec raison, à tort ...

Cette tripartition n'est pas sans rappeler les présupposés incontournables à l'activité communicationnelle de (Habermas 1987) qui sont l'intelligibilité, la vérité, la justesse et la sincérité. Le plus souvent, c'est le non respect de ces présupposés dans le discours rapporté qui sera mis en évidence. Le jugement évaluatif, qu'il soit moral, esthétique ou logique pose l'épineux problème de la nature des critères (et des conditions de leur applicabilité) sur lesquels se base ce jugement. L'impartialité, la neutralité et l'objectivité sur lesquelles repose la déontologie des médias d'information imposent aux journalistes de toujours rechercher la plus large adhésion possible à leurs points de vue, notamment en fournissant des arguments dont la validité puisse être largement reconnue. Tout en sachant que la prétention à la valeur absolue de ces jugements, sur des fondements objectifs qui devraient s'imposer à tout esprit rationnel, serait une marque d'une attitude dogmatique, tout aussi inadaptée que le subjectivisme ou le relativisme.

Deux extraits (10-11) pour illustrer l'intervention évaluative ciblant la manière de s'exprimer du locuteur cité. Les journalistes ne portent que rarement des jugements sur la valeur de vérité des paroles qu'ils citent. Ils s'autorisent en revanche des appréciations, parfois fort tranchées et souvent esthétisantes, sur la manière de s'exprimer de leurs sources. Des statistiques précises comparatives sur ces deux aspects seraient toutefois nécessaires pour confirmer cette observation.

(10) *Et Canal ? « Je trouve que, malgré tout, ils ne s'en sortent pas trop mal », juge-t-il **dans un grand élan de langue de bois**, avant de dispenser toutefois quelques conseils : « La chaîne a surtout besoin de redresser son image... Elle doit créer la différence en s'appuyant sur ses fondamentaux, à savoir le sport et le cinéma. » (Le Monde, 15 décembre 2003)*

(11) *Comme souvent, **il avait prononcé les mots justes et juste ce qu'il faut de mots** : « Je crains la mort, et je crois à l'immortalité des amis. Ils reviennent sans cesse [...]. Je les retrouve tous les jours au coin d'une bêtise. » Pour retrouver le magnifique et tendre Jean Carmet, en fugue depuis le 20 avril 1994 (73 ans d'état civil), il suffit de suivre le fil éclairant de cette bien belle soirée diffusée sur Festival. (Le Monde, 15 février 2003)*

4.5 Discours expliqué

Dans de nombreuses citations issues de notre corpus, les paroles rapportées se trouvent intégrées dans un discours qui « informe » ces paroles, c'est-à-dire les explique, reformule ou éclaire par des compléments d'informations, en fonction des besoins présumés des lecteurs (un co-énonciateur « type »). L'énonciateur prend alors l'entière responsabilité de la validité et de l'utilité de ces informations. On peut caractériser son intervention par la formule: « Je dis que ce qui est dit a été dit par un autre et je rends ce discours intelligible à mon co-énonciateur en disant ce qui suit. »

Cet aspect de la citation est très saillant dans les articles du *Monde* où la fonction de guidage (au sens de formation du citoyen) est importante. Le journaliste se positionne en médiateur, celui qui déchiffre, dévoile et explique les discours pour les rendre accessibles au plus grand nombre de lecteurs. Ce travail suppose en outre une connaissance la plus complète possible du sujet et de tous les paramètres explicatifs, condition souvent difficile à satisfaire en pratique. Par conséquent, la reformulation ou l'explication qu'il donne de ce discours peut, elle-même, être jugée pertinente ou non, juste ou fautive (12).

La séquence (13) illustre pour sa part une double intervention de la part de l'énonciateur : l'appréciation sur la manière de s'exprimer (« discours évalué ») (*le doux euphémisme*) du locuteur cité, et une reformulation interprétative (*en clair*) adressée à des lecteurs *a priori* peu habitués à la rhétorique communiste chinoise.

(12) *Mercier qui décidément a marqué les Moscovites : une autre de ses oeuvres faisait la couverture de la revue Art Kronika, avec la mention « Moscou livrée au Français ». C'est une citation de Pouchkine, qui évoquait la bataille de Borodino, une victoire des troupes de Napoléon.* (Le Monde, 14 juin 2004)

(13) *Li Qinfu sait parfaitement qu'une cellule communiste stipendiée peut lui être d'une grande utilité dans ses relations avec les autorités. « Le grand problème en Chine dans les affaires, dit-il, c'est l'absence de transparence des politiques officielles. Ces politiques changent tout le temps. Le comité du parti nous aide à mieux comprendre. ». Plus important encore est l'ordre social que fait régner ce même parti. « Il nous aide à évaluer la performance des employés. » Le doux euphémisme. En clair, le parti, c'est le syndicat maison. Sous le dôme du Capitole de Pinghu, le néo-capitalisme chinois croisé de tradition communiste germe. Une culture en serre d'un nouveau type, idéologiquement modifié.* (Le Monde, 15 janvier 2002)

5. Typologie des citations

Dans (Jackiewicz 2005, 2006) nous avons proposé de typer les séquences de discours rapportés en distinguant les citations minimales, les citations authentifiées, les citations interprétées et les citations dialogiques. Cette typologie de nature énonciative est construite à partir des différents modes d'intervention de l'énonciateur dans le discours qu'il rapporte :

- énonciateur rapporteur des seules paroles, restituées et signalées comme étrangères, dans la *citation minimale* (« paroles restituées » du 4.1.),
- rapporteur soucieux de l'authenticité des paroles citées, dont il restitue les principaux paramètres, dans la *citation authentifiée* (« énonciation restituée » du 4.2),
- observateur et interprétant attentif des aspects saillants du discours d'origine dans la *citation interprétée* (« énonciation interprétée » du 4.3),
- véritable médiateur investi, dialoguant avec la parole rapportée et avec le co-énonciateur, dans la *citation dialogique* (*évaluative* ou *intelligible*) (« discours évalué » du 4.4. et « discours expliqué » du 4.5).

L'intervention de l'énonciateur rapportant peut être interrogée selon trois axes.

Le premier concerne l'orientation de l'action rapportante : est-elle destinée en priorité au locuteur d'origine ou au co-énonciateur ? Les citations minimales et dialogiques sont orientées préférentiellement vers l'horizon du co-énonciateur. Si dans la citation minimale, l'énonciateur compte sur les connaissances partagées avec son co-énonciateur, dans la citation dialogique ce sont les connaissances que ce dernier ne possède pas qui sont au cœur de l'intervention de l'énonciateur. Cette intervention est régie par le principe de « pertinence énonciative », au sens de (Soutet 2005 p.172), ayant trait ici à la qualité et la quantité de l'information fournie. Les citations authentifiées et interprétées sont, quant à elles, focalisées sur le locuteur : la situation d'énonciation dont il est le repère fondamental, ses choix en matière d'expression et ses intentions communicationnelles. Autrement dit, la manière de rapporter un discours renseigne sur la nature

des rapports que l'énonciateur entretient avec les autres protagonistes de la situation de communication (proximité intellectuelle ou culturelle...).

Le deuxième renvoie à l'objet précis de cette intervention : la situation d'énonciation d'une part et les trois aspects saillants de l'énoncé rapporté qui sont sa forme d'expression, son contenu, la relation interpersonnelle qu'il établit. Chacun de ces aspects peut être restitué, soumis à une interprétation, une évaluation ou une explication. Prenons un exemple. La trace de la sollicitation de l'autre, qui est une des valeurs possibles de la relation interpersonnelle établie par le locuteur par son discours (*interpeller, solliciter ...*) peut s'enrichir d'une appréciation de la part de l'énonciateur, mettant en évidence son caractère abusif (*racoler, bassiner...*), la volonté d'imposer un modèle de pensée (*endoctriner...*) ou le fait de chercher par de belles paroles à obtenir des avantages personnels (*enjôler, fayoter...*).

Enfin, le troisième axe correspond au degré de connaissance du discours d'origine en rapport avec la réalité à laquelle ce discours fait référence. Celle-ci n'est pas obligatoirement la même, quand il s'agit de restituer ou d'authentifier un discours, de l'interpréter, de l'évaluer ou de le rendre intelligible. S'il est possible de restituer des paroles sans comprendre leur teneur, la médiation journalistique de qualité demande une maîtrise du contenu et du contexte linguistique et extra-linguistique.

L'implication de plus en plus grande de l'énonciateur, manifestée respectivement dans les actions énonciatives de restitution, d'interprétation, d'évaluation et d'explication du discours étranger entraîne une responsabilité croissante, susceptible à chaque niveau d'être jugée en termes d'approbation ou de désapprobation. Globalement, la responsabilité de l'énonciateur qui rapporte le discours d'un tiers est de trouver un mode d'intervention optimal, au vu de la connaissance qu'il a de ce discours et en adéquation avec la situation de communication à laquelle il participe.

6. Pour conclure

Je viens de présenter une analyse des différentes actions énonciatives réalisées ou réalisables par un énonciateur qui rapporte le discours d'un tiers. Il s'agit d'une contribution, fondée sur l'étude d'un corpus spécifique, à la compréhension de ce que peut recouvrir la notion de responsabilité énonciative dans le contexte des citations journalistiques.

Contrairement à certains auteurs, et notamment à la ScaPoLine, je propose de différencier la responsabilité de la prise en charge. Un acte que l'on peut qualifier de responsable appelle un jugement en termes d'approbation ou de désapprobation alors que la prise en charge d'une assertion donne lieu à un accord ou à un désaccord. La figure de l'autre fait ainsi partie intégrante du schéma définitoire de ces deux notions. Cette première esquisse de la notion de responsabilité reste toutefois à réexaminer pour d'autres actes langagiers.

L'exercice de la citation repose sur un schéma de base où l'énonciateur attribue un discours à un tiers (déterminé ou indéterminé). Différentes actions énonciatives peuvent s'y greffer selon les choix communicationnels de l'énonciateur, elles relèvent de sa responsabilité. En premier lieu, l'énonciateur est responsable de la manière de présenter et rendre intelligible ce discours. Il décide d'indiquer ou de ne pas indiquer la source, de restituer ou non le contexte énonciatif d'origine, d'aller jusqu'à reformuler et/ou expliquer les paroles rapportées, etc. Ensuite, il est responsable des jugements qu'il porte et choisit de faire connaître sur ce discours. Et ces jugements ne concernent pas uniquement la valeur de vérité des énoncés cités, mais peuvent porter également sur la forme d'expression ou sur la relation (visée) interpersonnelle dont ils étaient porteurs dans leur situation d'énonciation originelle.

IV. JUGEMENTS D'ÉVALUATION

A. Bref état des lieux

Le questionnement que j'aborde dans cette dernière partie du mémoire poursuit la problématique de la subjectivité langagière en se focalisant sur les jugements d'opinions et de valeur. Il s'agit d'une recherche en cours, dont l'objectif est de décrire systématiquement d'un point de vue linguistique et de modéliser certaines catégories de jugements d'évaluation. Ce travail est mené dans une perspective d'application, car il s'agit de formaliser et d'adapter le savoir linguistique ainsi constitué pour nourrir un système informatique de fouille et d'analyse d'opinions.

Mon ouvrage « original » étant entièrement consacré à cette problématique, je me contenterai ici d'un état des lieux relativement succinct.

En schématisant, dans les deux champs concernés, c'est-à-dire en linguistique et en traitement automatique des langues, nous avons affaire à un tableau contrasté.

En TAL et plus largement dans le monde de l'ingénierie linguistique, on constate aujourd'hui un vif intérêt pour le repérage de certaines formes de subjectivité dans le discours, et notamment des opinions énoncées à propos de certains types de cibles (objets culturels, services...). Les solutions et les outils existants, déjà nombreux, font état d'une conception relativement simple du « langage de l'évaluation ». Pour l'essentiel, ils s'inscrivent dans trois grandes catégories : (i) constitution de ressources lexicales ; (ii) classification de textes d'opinions ; (iii) analyse d'opinions dans les textes. On peut consulter notamment (Bloom *et al.* 2007) et (Pang and Lee 2008). Dans le contexte français, une série de travaux récents ont fait l'objet d'une publication thématique dans (El-Bèze *et al.* 2011). Les articles qui y sont rassemblés abordent les traitements automatiques des opinions, des sentiments et des jugements d'évaluation essentiellement d'un point de vue applicatif (par exemple, modéliser le déroulement affectif d'une conversation téléphonique dans un centre d'appel). En somme, ce sont les méthodes classiques du TAL, essentiellement numériques, qui sont employées pour traiter cet objet multiforme et encore fortement sous-déterminé. Dans le courant des approches symboliques ou mixtes, les travaux développés au LINA (Vernier *et al.* 2009a, b ; Vernier et Monceaux 2010) font partie des approches les plus élaborées sur le plan de la modélisation sémantique des énoncés évaluatifs. Ils proposent (i) une méthode de détection et de catégorisation des évaluations exprimées dans des blogs multi-domaines, en partant des classifications de (Charaudeau 1992) et de (Galatanu 2000), (ii) ainsi qu'une méthode d'apprentissage visant l'enrichissement de lexiques subjectifs.

En linguistique française, malgré des apports ponctuels significatifs, fouillés et finement élaborés, on constate que peu de travaux, en dehors de (Charaudeau 1992) et (Kerbrat-Orecchioni 2002), font référence. Le domaine est peu balisé et assez faiblement investi par les linguistes. L'ensemble des études disponibles, en rapport direct ou indirect avec l'évaluation, est fort hétérogène. L'axe « pragmatique – argumentation – rhétorique » peut en réunir certaines (actes de langage, polyphonie, prise en charge, image de soi...). Certaines études sur le lexique livrent des classes de marques faisant partie des architectures plus générales (verbes psychologiques, adverbess orientés vers le sujet...). Les recherches les plus poussées et les plus « techniques » concernent de toute évidence la problématique de l'émotion (études lexicales et grammaticales des termes d'émotion, étude de la communication émotionnelle...). Il n'existe pas encore de cadre théorique suffisamment solide pour asseoir des recherches appliquées.

Mes propres travaux sur le discours évaluatif se déroulent suivant deux axes. Le premier concerne l'étude des constituants détachés à caractère axiologique. Le deuxième se rapporte au travail collectif, mené dans le cadre du projet ANR OntOpiTex (voir l'Annexe « Valorisation de la recherche »), sur la construction d'un modèle applicatif de l'évaluation.

L'objectif principal d'OntOpiTex est de chercher à identifier et agréger des segments textuels porteurs d'opinions, et surtout de les classer par type (affect, jugement, appréciation, ...) et de les caractériser par des critères plus fins : polarité, intensité, prototypicalité, cible(s), source(s), engagement de l'auteur, etc. Cette problématique est abordée, dans une approche pluridisciplinaire et à partir d'observation de corpus, sur trois fronts : (i) modélisation linguistique et ontologique afin d'éclairer conceptuellement les notions en jeu à l'aide de théories linguistiques existantes, comme

Appraisal (Martin et White 2005) ; (ii) développement et mise en oeuvre d'un modèle opératoire pour l'analyse automatique dans les textes ; (iii) évaluation et étude de cas d'utilisation, à la fois sur un corpus de travail et dans un cadre applicatif réel. Dans ce contexte, mon apport concerne notamment l'élaboration d'un bilan détaillé sur la théorie *Appraisal*.

Le travail sur les constituants détachés de nature axiologique a permis d'initier la réflexion sur le discours évaluatif et la possibilité de son traitement automatique. Il a précédé le démarrage du projet OntOpiTex et se poursuit pour une partie dans ce cadre et pour une partie dans deux autres contextes (collaboration avec Max Silberstein sur les adverbes orientés vers le sujet et dans le projet InVivo).

Dans les grandes lignes cette recherche peut être résumée comme suit. J'exploite les travaux proprement linguistiques sur l'apposition, les constructions détachées et les compléments circonstanciels, décrits dans (Neveu 1998, 2000, 2002), (Guimier 1996) et (Combettes 1998). Au plan sémantique et énonciatif, je m'appuie sur la modélisation des attitudes évaluatives proposée par la théorie *Appraisal* (Martin and White, 2005). Du point de vue discursif, j'observe les relations rhétoriques, au sens de la *Théorie de la Structure Rhétorique* (Mann and Thompson 1988), qui peuvent s'ancrer sur un jugement évaluatif. Sur un versant plus formel, je dégage, à partir une observation de corpus, des patrons linguistiques sous-jacents aux constructions et aux expressions axiologiques les plus saillantes, ainsi qu'un lexique constitutif de ces expressions. Sur cette base, des expérimentations informatiques sur corpus plus vastes sont réalisées, afin de tester et d'amender les patrons, d'acquérir des lexiques et de collecter des observables. Ces ressources sont destinées à guider l'analyse sémantique du fait évaluatif dans des textes des corpus applicatifs, et préparent à l'implémentation d'outils d'analyse d'opinions.

Les analyses sur les constituants détachés composent, à ce jour, trois études. La première est englobante et la plus générale (Jackiewicz 2010). Elle met en évidence différents types de constituants détachés à caractère axiologique, dont elle propose une première description (formes, significations, rapport au contexte...). Elle inclut quelques expériences ciblées, pouvant être vues comme des études de faisabilité, destinées à recueillir des observables (Jackiewicz *et al.* 2009 a, b). Des grands axes d'analyse sont dégagés, ainsi que des types de faits linguistiques à observer (aspects modaux, structures énumératives, segments autonomes...).

La deuxième étude porte sur deux patrons choisis, associés à l'expression de la réputation. Leurs spécificités, l'appartenance catégorielle pour l'un et la sélection d'attribut saillant et (dé)valorisant pour l'autre, permettent d'amorcer une réflexion sur les rapports entre évaluation et catégorisation et de suggérer des études à entreprendre. Le travail est mené dans l'optique de la production de lexiques à partir de corpus. S'il est relativement facile, à partir de ces données, d'identifier les champs sémantiques représentés, l'objectif premier qui est la distinction entre emplois qualitatifs et emplois axiologiques s'avère hors d'atteinte. Les données quantitatives confirment l'existence de routines d'expression associées aux deux patrons analysés. Les extractions ont été réalisées avec l'outil LinguaStream (GREYC), en collaboration avec Stéphane Ferrari et Thierry Charnois.

La troisième étude, de plus grande ampleur, contrairement à la deuxième, part d'un vaste lexique déjà constitué et linguistiquement fondé : il s'agit des adverbes en *-ment*, orientés vers le sujet, issus des travaux de (Molinier et Levrier 2000). L'analyse, quantitative et qualitative, des attestations de ces adverbes ouvre un vaste champ d'observation, faisant apparaître notamment une importante polysémie des adverbes attestés. L'incidence du détachement et l'influence des éléments modalisateurs sont observées. Les auteurs des écrits explorés emploient une très faible part des adverbes disponibles, au sein des configurations récurrentes avec des lexiques restreints. Les collocations ainsi mises en évidence sont à étudier plus finement avec des outils lexicométriques adaptés. Les extractions ont été réalisées avec l'outil Nooj en collaboration avec Max Silberstein (Laseldi).

Dans l'ensemble, les constructions et les lexiques collectés ou attestés sur corpus sont soumis à des analyses fines (impliquant notamment des codages de propriétés saillantes, des analyses croisées, des calculs de fréquences...) pour dégager des catégories sémantiques dominantes, reflétant des systèmes de valeurs en usage, ainsi que les principales facettes des cibles.

B. Constituants détachés et jugements d'évaluation

La mise en discours de l'acte évaluatif constitue un champ d'études vaste et complexe. Mon choix s'est porté sur les jugements d'évaluation ayant pour cible des individus. Le cadre de l'étude a été retreint au phénomène textuel du détachement, « espace » privilégié mais, à l'évidence, non exclusif de l'expression des évaluations.

L'apposition détachée, « poste fonctionnel destiné prioritairement à l'expression de l'analyse, de l'interprétation, du jugement », comme l'a clairement montré (Neveu 1998), offre un angle d'approche original, à la fois particulier et complémentaire des méthodes plus globales.

1. Expressions qui ciblent des personnes

1.1 Formes récurrentes

Les éléments périphériques à la prédication principale dénotent souvent des appréciations, sous forme réduite et clairement délimitée. On y trouve, sans exhaustivité, des expressions stéréotypées, des locutions figées ou semi-figées, des qualifications métaphoriques, des antonomases du nom propre, des adverbiaux d'attitude ou de manière, des groupes adjectivaux sous différentes formes... (1). On peut y isoler des paradigmes d'expressions au fonctionnement équivalent et formellement proches (2).

(1) *femme de tête ; pur produit du sérail ; magicien de la langue ; agressivement misogyne ; une main de fer ; le Robin des bois du Web ; d'une courtoisie extrême ; en bon élève ; couronné de lauriers pour... ; à force d'obstination ; avec courage ; savamment ; puriste certes, mais pas puritain ; cheveux longs, idées libertaires et tenues brouillonnes...*

(2) *Prudent, /Avec prudence, /D'une prudence exagérée, /Regard prudent, /Prudemment, /Par prudence, /En homme prudent, /Connu pour sa prudence, Tom...*

Les constituants périphériques peuvent dénoter des propriétés objectives ou subjectives, associées à des entités de tout type ontologique, pouvant avoir, dans le cas de ces dernières, une axiologie stable ou instable (3). Leur interprétation appartient au lecteur, amené à décider en fonction du contexte, mais aussi de son point de vue propre et son système des valeurs.

(3) ***Chanteur, homme d'affaires, repreneur, politique, animateur de télévision, président de club, ministre, navigateur, failli sans faillite, Zorro justiciable, radicalement radical, sa vie est un long sprint.***

Dans cette configuration, les jugements d'évaluation, syntaxiquement libres, sont sémantiquement liés par des rapports de cause, d'opposition, d'élaboration (au sens de la RST) essentiellement, à l'état des choses dénoté par la prédication principale.

1.2. Prédication seconde : trace d'un désengagement

Les constituants extra-prédicatifs jouent clairement un rôle dans la hiérarchisation des informations. De l'ordre de la digression ou du présupposé, les contenus qu'ils dénotent sont généralement relégués au second plan, car ils sont vus comme acquis ou relativement consensuels. Neveu parle à ce propos d'« espace cognitif consensuel ». Cela est bien évidemment valable pour les segments de nature axiologique. Ainsi, dans (4), il n'est pas dit « X est *hypocondriaque* », mais cette information introduite par le biais de la prédication réduite (*hypocondriaque*) est présentée comme allant de soi ou comme étant communément partagée (dans ce cas précis, il s'agit d'un cinéaste mondialement connu). C'est une sorte de présupposé, considéré comme acquis, sur lequel il est possible de s'appuyer. L'auteur ne s'en trouve pas engagé de la même manière, l'intention première n'étant pas de juger mais d'informer. Cela a pour effet de minimiser sa responsabilité énonciative.

(4) ***Hypocondriaque, le clarinettiste amateur le plus fameux de Manhattan est un hédoniste contrarié : « on peut vivre centenaire si l'on abandonne toutes les choses qui donnent envie de vivre centenaire ».***

La question de l'engagement de l'auteur se joue également relativement à la fonction du segment détaché dans l'énoncé complet. La simple juxtaposition des prédications principale et secondaire a

pour conséquence de laisser sous-spécifiée la relation sémantique s'établissant entre ces constituants. Le choix de l'interprétation précise incombe alors au lecteur, libérant l'énonciateur de la charge d'assumer certaines assertions pouvant paraître délicates ou abusives. C'est notamment le cas des relations causales lesquelles, une fois explicitement posées, peuvent facilement s'interpréter en terme d'attribution de responsabilité, voire d'accusation. Dans (5), le rapport causal entre l'arrogance de la compagnie en question et son échec commercial n'est pas explicité. Mais il est clair que ce rapprochement suffit pour suggérer une telle lecture.

(5) **Volontiers arrogant, mal remis du scandale de corruption de 2004, trop sûr de son fait**, l'avionneur de Seattle s'est fait battre à plates coutures par un outsider d'autant plus motivé qu'il joue son entrée sur le premier marché du monde.

1.3. Ecriture condensée et dynamique

L'écriture avec constituants détachés est couramment pratiquée dans la presse. Elle présente, outre le confort d'un moindre engagement pour les scripteurs, des caractéristiques formelles utiles et intéressantes. Elle est condensée, fluide et dynamique, propice à des effets de style, supportant l'humour et la dérision (6).

(6) *Enfant intellectuel du Capital et de Rerum Novarum ; maestro de la désinflation ; puriste, certes, mais pas puritain ; sec, mais pas sectaire ; cool mais pas baba, beau mais pas bobo ; homme de glace et de glaciation ; furieusement rétro ; doucement violent, violemment secret...*

Cette manière de présenter l'information s'apparente à un enchaînement d'images (ou flashes) sur le référent concerné, rendant saillantes certaines de ses facettes. Pour reprendre la phrase de (Neveu, 1998), « Le souffle est court, le rythme alerte, le ton vif, les effets de densité et de resserrement sont très accentués ». On note que les articles à vocation biographique (portraits...) abondent d'énoncés à constituants détachés multiples, y compris dans des phrases averbales (7). Les segments détachés (adjectivaux, adverbiaux, sous forme de constructions absolues...), dès lors qu'ils remplissent des fonctions similaires, se trouvent combinés sans difficultés (8). Ce sont de véritables condensés de qualifications ou d'appréciations permettant de broser en peu de mots des portraits vivants, souvent drôles, détaillés et nuancés.

(7) *L'acteur d'Anouilh et de Chabrol, de Diderot et de Pinter, reste un voyageur traqué de l'âme. Inquiétant, ambigu, fascinant. (...). Laconique et inquiétant, discret et courtois, il cultive l'anonymat jusqu'à l'étrangeté.*

(8) **Homme de concept, pédagogue infatigable, militant sincère, sans l'ombre d'une complaisance pour l'argent**, Rocard prolonge à bien des égards la méthode qui séduisait tant chez Pierre Mendès-France.

2. Des expressions aux patrons

2.1. Démarche

Une première étude de ces constituants, préalable à toute expérimentation informatique, a été réalisée sur un corpus de textes issus du journal d'économie *les Echos* en accès libre sur www.lesechos.fr. L'exploration de ces textes a été manuelle ; elle s'est déroulée en trois temps. La première étape a consisté à prélever systématiquement tous les énoncés avec constituants détachés axiologiques dans les textes des rubriques « En vue » et « Portraits » en accès libre (110 textes au moment de la collecte). Lors de la deuxième étape, nous avons recherché sur l'ensemble du site des occurrences des marques (traduites en français) associées à la catégorie « Jugement » dans la théorie *Appraisal* (Martin and White, 2005, p.53) en prélevant les énoncés où les marques en question figurent dans un constituant détaché. Sur la base de ces attestations nous avons formulé un premier ensemble de patrons linguistiques (une dizaine environ). Dans un troisième temps, nous avons recherché, toujours sur l'ensemble du site, certaines configurations précises de marques dans le but de systématiser ces patrons. A chaque étape, les marques lexicales constitutives des segments évaluatifs étudiés ont été collectées et organisées dans des classes selon leur catégorie grammaticale et leur fonction dans le segment.

Les expressions linguistiques correspondantes aux patrons dégagés doivent vérifier conjointement deux conditions : (i) figurer en position détachée : à l'initiale, en incise (entre sujet et verbe) ou détachée à droite, (ii) apparaître en cooccurrence avec un syntagme qui réfère à un individu, un groupe d'individus ou une institution : *Obama, il, le président, cet homme politique, l'Américain...* Ce syntagme dénote la cible du jugement.

2.2. Patrons, lexiques et attestations

Les régularités observées mettent en évidence l'existence de différentes formes de surface, où des constructions peuvent être précisées par des lexiques spécifiques. 20 patrons linguistiques ont été dégagés, sans aucun principe d'uniformité ou de complémentarité entre eux. Ils représentent des schémas sous-jacents aux expressions axiologiques les plus saillantes et récurrentes. Un corpus d'exemples associés aux patrons linguistiques a également été créé, ainsi qu'un premier lexique (550 termes) constitutif de ces expressions (Jackiewicz, 2009). Aucune forme d'exhaustivité n'est revendiquée à ce stade de l'étude. Chaque patron nécessite à son tour une étude linguistique approfondie. A ce jour, cinq de ces patrons (P1, P11, P16, P17 et P18) ont été analysés sur corpus.

- P1 : {homme, femme} de N – *femme de caractère* – génitif de qualité
- P2 : (art){homme, femme} à GN – *un homme à la forte personnalité* – épithète
- P3 : art N Adj - *l'obstination bretonne* – construction absolue (pour les titres)
- P4 : art {oeil, nez, tête, visage...} Adj – *l'œil vif* – constr. abs. avec parties du corps
- P5 : N de GN – *maestro de la désinflation* – expression métaphorique – < excellence >
- P6 : N adj / adj N – *économiste brillant* – GN sans dét. – < savoir-faire >
- P7 : {fine mouche, petit soldat, canard boiteux...} – expressions figées métaphoriques
- P8 : art NP de GN – *un Rimbaud de l'ère numérique* – antonomase du nom propre
- P9 : {parangon, modèle...} de GN – *parangon du cynisme*
- P10 : (Adj) {adepte, chanteur, amateur,...} de (Adj) N – *amateur de bonne chaire* -
- P11 : GAdj – *imprévisible et fantasque* – groupe adjectival
- P12 : GAdj à/pour Vinf – *habile à construire...* – groupe adj. à expansion infinitive
- P13 : de artInd Adj N / N Adj – *d'une extrême fermeté* – génitif de qualité
- P14 : de art-ind {voix, geste, regard, œil...} Adj – *d'un regard ironique*
- P15 : {fort, fier, riche...} de art_poss (Adj) N – *fort de ses succès* – < expérience >
- P16 : {réputé, salué...} pour art_poss (Adj) N – *réputé pour son sang froid*
- P17 : en N Adj / Adj – *en industriel responsable* – structure prépositionnelle en *en*
- P18 : {avec, sans, non sans} N/ adverbe en 'ment' – *avec pragmatisme* – adverbial
- P19 : {grâce à, à force de...} (art_poss) N – *à force d'obstination* – locution causale
- P20 : {en mauvaise posture, en vogue, au sommet de art-poss...} – localisation abstraite

Tableau 1 : Patrons linguistiques avec leurs catégories et exemples illustratifs

Les patrons identifiés renvoient à des formes de surface récurrentes, mais non nécessairement stables. La plupart de ces patrons peuvent admettre des variantes (dues à la présence/l'absence d'adjectifs, d'adverbes, d'expansions en *que/qui*, de coordinations). Ils ne correspondent donc pas à des constructions maximales. Par exemple, le groupe adjectival peut être décliné en une multiplicité de formes, plus ou moins complexes (*populaire ; imprévisible et fantasque ; plus instinctif qu'érudit ; ni trop sentimental, ni trop énergique...*).

Les patrons correspondent à des expressions à sens souvent compositionnel, même si des formes figées en font également partie, saturés par des lexiques à étendue variable. Selon la nature de leurs composants, on distingue : (i) les patrons génériques constitués de catégories grammaticales et de prépositions (P3, P6, P8, P11, P12, P13, P17), (ii) les expressions figées construites sur un modèle particulier (P7, P20), (iii) les patrons faisant appel à des lexiques plus ou moins spécifiques, plus ou moins étendus (tous les autres patrons). De ce fait, les règles correspondantes, utilisées pour la recherche des patrons dans des corpus, ne sont pas toutes comparables. Les règles qui exploitent des listes d'entrées lexicales finies pourront entraîner du silence, d'autres, plus génériques, sont susceptibles d'amener du bruit. Cette question de précision et d'ambiguïté doit être considérée pour chaque patron indépendamment.

Suivant les catégories des constituants, on distingue : (i) les constructions détachées au sens de (Combettes 1998) correspondant aux expressions adjectivales, participiales, de type construction absolue (P3-P4 ; P11-P16) (*prudent, ...*) (ii) les formes adverbiales et apparentées (P17-P20) (*avec persévérance et beaucoup de diplomatie, ...*), (iii) les constituants nominaux avec ou sans déterminants (P1-P2 ; P5-P10) (*journaliste expérimenté, ...*). Les différences de fonctionnement sémantique entre ces trois catégories de patrons sont notables. Par exemple, la famille des adverbes et des formes prépositionnelles se distingue des deux autres par son rôle essentiellement circonstanciel. Elle présente, avec des variantes possibles, la situation dénotée par la prédication principale comme une manifestation d'une qualité attribuée à l'actant support.

Enfin, le dernier critère renvoie au type (ontologique et/ou sémantique) de la cible de l'évaluation. Certains patrons capturent des qualités spécifiquement humaines, d'autres, formés de catégories

grammaticales et de prépositions, peuvent s'appliquer à des cibles de différentes natures. Dans le premier ensemble notons : (i) des patrons qui contiennent un terme catégorisant spécifique (P1, P2, P5, P10...) *femme de tête, apôtre du capitalisme familial...* (ii) des patrons qui exploitent le langage du corps humain (P4, P14) *une main de fer, un rire de velours, l'œil vif...* (iii) des antonomases du nom propre (P8) *un Rimbaud de l'ère numérique, ...*

3. Explorations outillées

Cette première analyse des expressions évaluatives en apposition menée sur un corpus des *Échos* a permis de proposer un ensemble de patrons ainsi qu'un lexique associé. A partir de ces ressources, plusieurs tests informatiques sur corpus analogues (*Le Monde*) ont été réalisés ou sont en cours, en collaboration d'une part avec Thierry Charnois et Stéphane Ferrari (Jackiewicz *et al.* 2009 a ; 2009 b ; Ferrari *et al.* 2009), et d'autre part avec Max Silberstein. Ces tests et explorations ont porté sur une sélection de 11 patrons.

3.1. Collectes ciblées des observables et premières tendances

Les testes opérés sur les différents patrons ont permis d'apprécier globalement leur fréquences d'emploi et la qualité des résultats ramenés. Le groupe adjectival étant parmi les plus productifs, il a fourni un ensemble conséquent d'énoncés²⁶ qui ont servi à des observations plus poussées. L'examen des concordances obtenues a permis de dégager une série de caractéristiques notables des énoncés comportant des qualifications adjectivales. Parmi les plus significatives, il y a :

- la présence significative de séquences à constituants détachés multiples (énumérations) et de groupes adjectivaux complexes (*Chaleureux, provocateur, soupe au lait, terriblement affectif, M. K...*)
- les jeux d'opposition, de renforcement, de complémentation par la présence de connecteurs ou de marqueurs temporels, (*Militant mais opportuniste, franc-tireur mais habile, sociable mais anticonformiste, ...*)
- la présence de groupes adjectivaux autonomes,
- les phénomènes de contagion contextuelle influençant la polarité (*Tenace, solide, décidé, mordant.*),
- l'influence de la position détachée sur l'interprétation axiologique de certains adjectifs (*ouvert, européen décomplexé*).

Dans un deuxième temps, une analyse approfondie impliquant une phase de codage a été réalisée par Nadia Zerguini²⁷ sur un échantillon de 100 premières attestations, afin de vérifier ces tendances. 5 % des groupes adjectivaux sont graphiquement autonomes. 60 % des énoncés comportent un groupe adjectival composé. Dans ces 100 énoncés, on dénombre au total 153 éléments axiologiques. Sur le plan sémantique, les catégories notionnelles d'*Appraisal* ont été jugées pertinentes pour caractériser environ 50% des constituants. Les tendances observées constituent autant d'éléments à prendre en considération pour la mise en place de traitements informatiques futurs, notamment pour la délimitation des segments minimaux dénotant des jugements et pour l'interprétation de la polarité de ces jugements.

3.2. Acquisition et analyse des lexiques

Certains constructions se prêtent relativement aisément à la collecte automatique de lexiques potentiellement axiologiques. C'est notamment le cas des patrons P1 et P16. La démarche d'acquisition de lexiques pour ces deux patrons est incrémentale. Elle consiste à fixer la partie gauche du patron (la tête), formée par une amorce de lexèmes, pour collecter le lexique de la partie droite (l'expansion), puis selon le degré de figement du patron, s'appuyer sur le lexique acquis de la partie droite pour enrichir les constituants de gauche. Nous avons réalisé cette expérimentation sur les éditions complètes couvrant 16 ans (pour P16) et 20 ans (pour P1) du journal le *Monde*. Les

²⁶ 1 662 attestations ont été identifiées dans 535 articles de catégorie portrait ou biographie du *Monde* (sur un total de 884 articles, 4 Mo, environ 700 000 mots).

²⁷ Zerguini N., Etude des groupes adjectivaux extra-prédicatifs à valeur axiologique, Mémoire de Master 1 Informatique et Ingénierie de la Langue pour la Gestion de l'Information, Université de Paris-Sorbonne, 2009.

attestations collectées ont permis de décrire les variantes formelles et les lexiques saturant les places « libres » de ces patrons, et sur cette base, identifier les valeurs sémantiques dominantes associées aux expressions correspondantes. Les données quantitatives obtenues ont servi à dégager certaines tendances.

Ces deux familles d'expression servent à exprimer des jugements assurés, où l'attribution de la qualité est catégorique, ne comportant en règle générale ni condition, ni alternative, ni restriction temporelle. Elles diffèrent entre elles sur un certain nombre de points. On peut noter par exemple que les qualités prédiquées par la construction P1 « (*homme|femme*) de N » sont assignées directement à l'individu cible. Le tour « *Homme de courage, X...* » ne dit pas comment se manifeste le courage de X (*propos courageux, actions courageuses, décisions courageuses*). La construction P16 « *réputé|connu... pour Art_poss GN* » permet, quant à elle, de sélectionner des aspects particuliers de la cible sur lesquels s'est focalisé le jugement.

Voici, en quelques phrases, ce que l'on retient au sujet de la construction « homme|femme de N ». Les expressions construites sur ce schéma et employées en position détachée sont dédiées à l'acte de qualification, lequel dans près de 40% des cas se charge d'une axiologie positive ou négative. Elles s'insèrent souvent dans des structures appositives multiples, entièrement évaluatives ou panachant évaluation et qualification. L'axiologie positive est largement dominante. Les valeurs axiologiques les plus fréquentes touchent à la qualité des rapports sociaux, de l'expérience professionnelle et de l'implication dans l'action. Les emplois qualifiants, axiologiquement neutres, se caractérisent par une plus grande spécialisation sémantique, avec une focalisation sur l'appartenance experte à des univers artistiques ou politiques. Dans les deux cas, la construction « homme de N » est nettement plus fréquente que son analogue « femme de N ». En somme et en simplifiant, dans le corpus analysé, cette expression sert à valoriser des individus de sexe masculin pour leurs capacités relationnelles et leur connaissance investie du monde des arts ou de la politique.

La structure en question appartient de toute évidence à une palette d'éléments langagiers préformés ou « préfabriqués », dans laquelle les auteurs (ici les journalistes) puisent pour exprimer des jugements évaluatifs. On remarque que les lexiques effectivement employés pour la saturer sont relativement restreints. Nous avons noté quelques associations privilégiées entre unités lexicales. Des tours tels que *homme de confiance, homme de conviction, homme de terrain, homme de gauche* se détachent par leur fréquence au sein du corpus retenu, mais on peut s'attendre à les retrouver dans des proportions comparables à une échelle plus vaste.

3.3. Etude de constructions polysémiques, partiellement instanciées

De la même façon, il est possible d'interroger les emplois de certaines constructions quand elles sont instanciées par des ensembles lexicaux particuliers (Jackiewicz *et al.* 2009b). Nous avons réalisé une étude détaillée sur le patron P17 « en N+<exp> » dans une réalisation particulière : « ponct *En|en (très) adj mot{1,15} ponct* » avec *adj* → *bon | vrai | digne | authentique | véritable | simple | vieux | jeune* (+ variations morphologiques). Cette configuration, même construite sur un sous-ensemble restreint et homogène d'adjectifs enclosures, reste ambiguë.

Sur un corpus formé de quatre ans *du Monde* (1Go), toute catégorie confondue, 398 attestations ont été trouvées. La construction n'est pas dédiée à l'évaluation des sujets humains, mais cet emploi reste majoritaire dans le corpus : 59 % des attestations prennent pour cible des individus ou des organisations. Les expressions qui apportent du bruit sont peu nombreuses et fréquentes (* *en bonne logique, en bonne voie, en bonne santé, en bonne partie, en bonne et due forme...*). Cela permet de les écarter à l'avance du traitement. Le deuxième constat met en lumière une très inégale représentation de ce patron selon les adjectifs (*bon* : 174 ; *vrai* : 18 ; *véritable* : 13 ; *vieux* : 13 ; *digne* : 10 ; *simple* : 3 ; *authentique* : 2, *jeune* : 1). Ce résultat mérite d'être vérifié sur d'autres corpus.

Ces attestations peuvent être questionnées plus finement. Par exemple, quelles valeurs peut prendre l'expression « en bon N » quand N renvoie à un humain ? L'observation des champs lexicaux couverts par les substantifs en question, corrélée avec l'examen des prédications principales, met en lumière un résultat notable. 57 % des emplois dénotent l'idée de conformité (ou de conformisme), 32 % expriment l'appréciation d'une capacité ou d'une aptitude, dans 12 % des cas ces deux valeurs sont cumulées. Le jugement de conformité renvoie à l'attitude de quelqu'un qui règle ses idées, son comportement, sur ceux de son milieu (religieux, politique, social, culturel,

familial) : *en bon fils de la bourgeoisie brestoise, en bon citoyen helvétique, en bon spinoziste...* Dans la terminologie *Appraisal*, il s'agit de la catégorie « Gradation » (focus renforcé). Le jugement de qualité quant à lui correspond à un jugement positif porté sur des capacités (savoir-faire, habileté) manifestées par une personne : *en bon contrebassiste, en bon avocat, en bon professionnel...* . Ce cas tombe sous la catégorie « Attitude », dans les rubriques « Jugement », « Estime sociale » et « Capacité ».

3.4. Exploiter une ressource pré-existante à l'étude

Il arrive que des études linguistiques conduites indépendamment de la problématique de l'évaluation puissent fournir des ressources lexicales pertinentes, à la fois de grande ampleur et solidement documentées. C'est le cas des adverbes en – *ment*, appartenant à deux classes fonctionnelles nommées « adverbes disjonctifs d'attitude orientés vers le sujet » et « adverbes de manière orientés vers le sujet » (OS). Nous les devons à (Molinier et Levrier 2000). Cet ensemble représente environ 880 adverbes.

Le choix de la catégorie adverbale oriente l'investigation sur les jugements ciblant des actions, des comportements, des attitudes. Ces deux ensembles d'adverbes particuliers, « orientés vers le sujet », dénotent des appréciations portées sur des situations agentives et des qualités déployées par les agents qui y sont impliqués intentionnellement (9).

(9) *Depuis quatre ans, Alain Carignon s'évertue **methodiquement, consciencieusement, minutieusement**, à préparer les conditions de son retour.*

On constate que l'orientation vers le sujet induit un fonctionnement spécifique de l'adverbe d'attitude. Les attitudes y sont dotées d'une efficience causale (10).

(10) ***Prudemment**, il n'a pas repris la promesse de M. Chirac de réduire la pression fiscale de 30 milliards d'euros sur cinq ans (dont 12,5 milliards pour l'impôt sur le revenu).*

Or, l'explication des actions par les attitudes ou les dispositions (plus ou moins stables ou inhérentes) des agents soulève toutes sortes de difficultés, à commencer par son caractère, invérifiable et potentiellement abusif. Cela peut expliquer la très faible part des adverbes d'attitude orientés vers le sujet effectivement présents dans le corpus. Les adverbes utilisés dans le corpus tiré du *Monde* sont majoritairement positifs et couvrent un ensemble restreint d'attitudes, lesquelles sont plus intellectuelles que morales²⁸. De leur côté, les adverbes de manière orientés vers le sujet sont employés plus couramment, même si une forme de retenue semble s'appliquer, du moins dans le corpus considéré, par rapport aux adverbes les plus lourdement négatifs.

Concernant les adverbes de manière OS, on retiendra principalement que le détachement met en lumière la polyfonctionnalité de ces adverbes et leurs rapports privilégiés avec la quantification. Le deuxième fait à noter la présence significative des structures énumératives ; cette propriété est utile pour la désambiguïsation des adverbes. Mais au delà de cet aspect opératoire, on constate que ces segments complexes encodent des aspects particuliers de l'acte évaluatif. Enfin, les groupes adjectivaux ou nominaux avec adverbes OS que l'on rencontre en apposition sont clairement de nature à exprimer des jugements évaluatifs (11), certains pouvant être fortement stéréotypés (sous forme de locutions). Dans cette configuration, ces adverbes remplissent souvent deux fonctions étroitement liées : de qualification et de quantification.

(11) *Moi, c'était la catastrophe : j'étais à la fois foncée et crépue. Et **désespérément maigre** !*

Observer les fréquences d'emploi de ces adverbes, souvent en co-occurrence avec d'autres éléments, conduit à identifier leurs emplois effectifs, courants, voire stéréotypés, plus marginaux... Il devient également possible de mettre en évidence les structures dans lesquelles ces adverbes apparaissent préférentiellement, ainsi que les rapports entre structures et significations. L'emploi des séquences autonomes (12), comme pour les segments adjectivaux est attesté et peut être apprécié précisément.

(12) *Permettez ici qu'on ricane. **Gentiment mais fermement**. Une vieille expérience des sondages peut-être.*

²⁸ Les ouvrages littéraires issus de Frantext, les adverbes d'attitude OS sont (i) mieux représentés (ii) deux fois plus fréquents et surtout (iii) présents de manière bien plus équilibrée, quant à la balance des adverbes positifs et négatifs.

Force est de constater qu'en choisissant les adverbes orientés vers le sujet comme fenêtre sur le discours évaluatif, notre problématique s'est multipliée. Nous avons été confrontée tout au long de l'étude à deux phénomènes linguistiques particulièrement complexes : la polysémie et le figement.

Le travail en corpus (l'année complète 2002 du *Monde*) a permis de mesurer le degré de polysémie des adverbes et constater leur propension à former des locutions. Des associations privilégiées avec des verbes particuliers, présence ou absence de certaines modalités, co-occurrences avec d'autres adverbes (notamment ceux qui sont fonctionnellement moins ambigus) sont autant d'indicateurs à prendre en considération. Les régularités ainsi décrites donnent des indications sur les traitements à mettre en œuvre pour augmenter la pertinence des résultats attendus, sans prétendre à résoudre toutes les difficultés qui pourront se présenter. Le recours au corpus a également permis de compléter et parfois d'amender sur certains points l'étude théorique originelle. Dans les tableaux livrés par Molinier et Levrier, la palette des fonctions possibles est clairement sous-estimée pour certains adverbes.

4. Caractérisation ontologique, branchée sur le social et le culturel

Les pistes de recherche ouvertes par les différentes analyses réalisées sur les constituants détachés à caractère axiologique sont nombreuses.

L'étude approfondie des expressions telles que *femme de pouvoir* ou *homme de lettres* m'a amenée à amorcer une réflexion sur les rapports entre évaluation et catégorisation. La catégorie est un puissant opérateur d'intelligibilité. Elle semble intervenir également et à plusieurs titres dans les processus de qualification et d'évaluation axiologique. Il est à noter que parmi les patrons dégagés, six associent de différentes manières évaluation et catégorisation (*femme de caractère, patron à poigne, brillant économiste, en politicien expérimenté... Monsieur Jourdain de la décroissance, maestro de la désinflation...*).

Les sujets humains sont évalués ou jugés en tant que membres de leurs catégories, mais aussi en raison de cette appartenance, et parfois même au moyen d'une « catégorisation subjective » par identification à des individus dotés de qualités remarquables. Certains attributs catégoriels semblent être choisis préférentiellement pour motiver ou supporter un jugement d'évaluation. La sélection d'attribut peut viser, selon les cas, le trait qui singularise l'individu au sein de sa catégorie (normalité), le trait qui en fait un bon représentant (typicalité) (focus), le trait qui en fait un individu désirable, conformément à la hiérarchie des valeurs en cours pour la catégorie en question (qualité).

(13) *Tout au long d'une vie entièrement consacrée au service de l'Etat, il fut tour à tour **homme de gestion, homme de décision, homme de pouvoir, homme de réflexion**, c'est-à-dire un " politique " au vrai sens du terme.*

Avec le patron P1, nous avons mis en évidence des représentations courantes (plus ou moins stéréotypées) associés aux genres. Par la même méthode, il est possible d'étudier dans des corpus constitués à cet effet les qualités associées, par exemple, à des catégories professionnelles (14). Il s'agit du patron P6, dont le constituant nominal appartient à un ensemble de termes préconstruit, composé de noms de métiers, de professions et de fonctions.

(14) *brillant industriel, opiniâtre patron, patron opérationnel, chanteur intelligent, chirurgien rigoureux et élégant, écrivain rare, instituteur pointilleux, comédien médiocre, petit comptable, petit employé, austère professeur, artiste prestigieux...*

La faisabilité d'une telle étude a été montrée par Laurie Acensio²⁹, sur un corpus est constitué d'articles journalistiques de type « portrait », « biographie » et « nécrologie » parus dans le journal *Le Monde* entre 1987 à 1994. Le travail d'exploration a été réalisé à l'aide de Lexico3 et d'extracteur de collocations Coocs.

Une tout autre forme de catégorisation est à l'œuvre dans le procédé antonomastique. Il s'agit dans ce cas d'une forme particulière de « ressemblance de famille » ou encore d'une structure prototypique, construite à partir des qualités, d'un parcours ou d'un destin particulier d'un individu qui devient le parangon de la catégorie. Etre assimilé à un individu modèle revient à être jugé positivement ou négativement, selon l'image sociale attachée à cet individu référent. Le cas des

²⁹ Dans le cadre d'un projet intégration en Master 2, IILGI, à Paris-Sorbonne (2010-2011), intitulé « Groupes nominaux à fonctions catégorisantes et qualifiantes ».

antonomases est à considérer également en raison du fait que ce type de jugement mobilise fortement des références culturelles (*Ségolène Royal, en Jeanne d'Arc des familles ; Borloo : Monsieur Jourdain de la décroissance ; « Alcatel, Sisyphé des télécoms »*...). Il semble intéressant d'étudier les antonomases telles qu'elles fonctionnent dans certains univers (culturels, professionnels...) particuliers, pour en découvrir les valeurs importantes et les personnes qui en sont l'incarnation.

5. Conclusion

Le mode d'expression appositif est le fait d'une pensée analytique, interprétative et souvent critique. Il permet de mettre en évidence, visuellement ou auditivement, des attributs saillants ou socialement valorisés des individus cibles. Il les présente, et c'est un fait notable, de manière faiblement modalisée, comme relativement sûrs et consensuels. Ces caractéristiques font des segments apposés à la prédication principale un espace à explorer en priorité, quand on vise la détection automatique de jugements d'évaluation.

Au stade actuel, du point de vue sémantique, l'un des apports de ce travail est d'avoir montré que certaines structures renvoient à des aspects particuliers des actes évaluatifs (portée du jugement, vision catégorielle de la cible...). Nous avons mis en évidence des valeurs spécifiques associées à certains patrons, par exemple jugement de conformité vs jugement de qualité, pour « en bon N ». Deux familles de constructions associées à l'expression de la réputation ont été analysées. Les observables accumulés dans certaines expériences d'extraction sur patron ont permis d'initier un examen critique du modèle *Appraisal*, du point de vue de la couverture notionnelle proposée.

L'étude sur les constituants détachés a fait émerger des structures récurrentes au moyen desquelles il est possible de capter l'expression des qualités, positives ou négatives, associées à des individus. Plus généralement, le système appositif fournit, au travers des expressions que l'on rencontre dans son cadre, des informations sur : (i) la visibilité des catégories de cibles (selon les genres, les professions...), (ii) les facettes ou les composantes de la réputation des cibles (formes de sociabilité, opinions...), (iii) les champs sémantiques des valeurs actualisés (formes d'interaction sociale...).

Il s'agit en quelque sorte de témoins linguistiques des valeurs que la société considère comme désirables ou indésirables, au cours d'une époque. Analyser le fonctionnement de ces structures, grâce à des explorations systématiques sur des corpus appropriés, permet de construire une certaine image de ces valeurs. Pour cette raison, nos résultats peuvent intéresser toute une gamme d'applications portant sur la visibilité, la notoriété et la réputation des individus et des institutions. Il devient également possible d'analyser sous différents angles la perception des catégories socio-professionnelles (les caractéristiques qui les rendent attractives, celles qui les singularisent, leurs représentants les plus illustres...). Un des prolongements que je souhaite donner à ce travail serait, grâce à une collaboration avec des socio-linguistes, sociologues ou spécialistes de science politique, de pouvoir (ré) interroger ces données à la lumière des problématiques sociales concrètes.

BILAN ET PERSPECTIVES

L'étude des traces de l'acte énonciatif laissées par l'énonciateur dans l'énoncé offre au linguiste un champ d'observables d'une grande richesse. Je pense avoir montré que les études outillées sur corpus conduisent à mieux prendre la mesure des régularités et de la variabilité des réalisations effectives. Cela permet de revenir vers les conceptualisations avec une vision plus riche et plus complète des faits énonciatifs. La deuxième idée générale est que de telles analyses sur corpus, théoriquement fondées et centrées sur l'observation systématique des formes d'expression, contribuent à améliorer notablement la qualité des ressources que l'on destine à des applications informatiques. Les progrès des études sur l'énonciation sont en partie liés à leur automatisation. De son côté, le TAL a beaucoup à gagner à accroître la part des connaissances sur les phénomènes énonciatifs dans les traitements informatisés, quelle que soit leur visée.

D'un point de vue méthodologique, j'ai tenté de rendre compte aussi clairement que possible des problèmes que j'ai rencontrés et que l'on rencontre souvent en linguistique informatique. Je pense qu'il est important d'indiquer précisément les limites des travaux conduits. Cela ouvre la porte aux améliorations potentielles. En pratique, on recherche souvent une forme de consensus entre solutions empiriques et solutions fondées théoriquement. Les besoins concrets suscitent et stimulent les recherches, mais ne disent pas sur quels objets précis et avec quelles méthodes travailler. Les applications mettent à l'épreuve les modélisations linguistiques, mais ne peuvent en assurer la validation. L'implémentation implique toujours une simplification voire une dégradation des connaissances linguistiques employées. Il est important de noter qu'il s'agit en général de travaux s'inscrivant dans la durée, bénéficiant d'une multitude d'essais et de contributions collectives. Tendre vers une meilleure capitalisation des savoirs et des ressources linguistiques, autorisant leur transmission et leur réutilisation me paraît être une voie à la fois indispensable et prometteuse.

La synthèse qui s'achève vient de présenter quatre recherches concernées par la texture énonciative des discours. Elles contribuent à la description fine de quelques faits linguistiques ayant une dimension subjective avérée et à la construction de ressources linguistiques associées. Ces travaux livrent un ensemble catégorisé de ressources lexicales assorties, pour certaines, de données statistiques et de règles matérialisant les contraintes et les tendances constatées (voir la section 9 des Annexes). Le rôle des analyses préalables à des applications informatiques est d'apporter des solutions opératoires, mais linguistiquement valides, « au plus près » de l'objet considéré, dans la réalité de ses emplois. Tel a été mon objectif tout au long de ces années de recherche.

Je ne reviendrai pas sur les acquis respectifs de ces différents travaux. En dernier lieu, je voudrais seulement faire remarquer deux choses.

J'ai examiné avec beaucoup d'attention les formes qui cumulent plusieurs significations, autrement dit, qui sont trace de plusieurs opérations réalisées conjointement. En discours, certaines marques, à l'image de celles que l'on rencontre dans les cadres organisationnels par exemple, sont multifonctionnelles. L'étonnante économie des discours tient de toute évidence à cette condensation de fonctions. C'est un point auquel le TAL n'accorde pas encore suffisamment d'attention.

Le travail sur les adverbes en *-ment* orientés vers le sujet est celui qui marie, à mes yeux, le mieux la linguistique « classique » et la linguistique informatique. Le fait de pouvoir s'appuyer sur une étude linguistique solide, claire, systématique et fort bien documentée, comme celle de (Molinier et Levrier 2000), est un atout et un avantage indiscutable. Une telle ressource permet, grâce à des outils d'exploration de corpus, comme Nooj ou LinguaStream par exemple, d'accumuler des observables que l'on peut alors coder et interroger selon de multiples critères. Sur cette base fiable, toute une chaîne d'analyses peut alors se mettre en place et pour permettre également de revenir vers l'étude initiale en la questionnant et l'enrichissant. C'est un cercle vertueux auquel j'aimerais pouvoir contribuer plus souvent.

Bien des pistes se sont ouvertes durant ce travail que je n'ai pas eu la possibilité d'explorer ou d'approfondir. Au terme de cette synthèse, l'examen de problématiques passées suscite l'envie de

revisiter et d'approfondir certaines recherches évoquées. Dans un premier temps, j'en vois principalement deux.

Je voudrais tout d'abord poursuivre le travail sur les catégories sémantiques associées aux attitudes de type « Jugement », pour reprendre le terme issu d'*Appraisal* (il s'agit d'évaluations ciblant des personnes, des organisations, etc.), notamment sur la question des compétences sociales. Le champ sémantique des compétences sociales reste encore peu et imparfaitement décrit, contrairement, notamment, à celui des émotions. Ce n'est pas un nouveau modèle qui est visé, mais une méthode reproductible de recueil et d'analyse des observables montrant comment dans un contexte précis, il est possible d'utiliser les corpus pour découvrir les systèmes de valeurs qui y sont véhiculées. La deuxième piste de recherche que j'aimerais poursuivre est, comme je l'ai indiqué dans l'ouvrage nouveau, l'étude des rapports entre évaluation et catégorisation. Dans les deux cas, je souhaiterais pouvoir associer aux méthodes de recherche par patrons que j'emploie habituellement des méthodes issues de la lexicométrie. Le domaine de la statistique textuelle est celui que j'aimerais investir dans les prochaines années.

En ce qui concerne les perspectives à plus long terme, deux chantiers me tiennent à cœur.

Le premier tient à la modélisation des notions énonciatives et leur capitalisation au sein des représentations formelles de nature ontologique. Un premier apport vient d'une étude réalisée en 2007 sur la modélisation des constituants et des relations propres aux discours rapportés (OntoDiscours). Un autre travail de plus grande ampleur est en cours dans le cadre du projet OntOpiTex Il apportera une contribution pour ce qui concerne la modélisation des jugements évaluatifs. Une partie de la thèse de Mathias Lambert que je co-encadre prendra en charge l'élaboration d'une telle ontologie.

Le deuxième, plus programmatique, concerne le rapprochement de mes recherches sur la subjectivité dans les discours avec certaines problématiques sociales (étude des rapports et des conflits sociaux, analyse de la construction identitaire des personnes, observation de la construction des inégalités...). On voit très nettement comment les explorations outillées des productions langagières circulant à l'intérieur des espaces déterminés pourraient participer à une approche transversale de ces questions, en se plaçant dans une démarche résumant pluridisciplinaire (économie, droit, philosophie, science politique, sociologie, histoire, psychologie...). Je songe notamment à des initiatives telles que Presage³⁰ à Sciences-Po, où la linguistique n'est pas présente.

L'étude de la subjectivité constitue de toute évidence un sujet à vocation fédératrice. Elle offre matière à des échanges pour la communauté des linguistes, des informaticiens « talistes », des chercheurs en sciences sociales. Ma participation à la mise en place de l'Observatoire des pratiques médiatiques émergentes (Praxiling) est un premier pas dans ce sens.

³⁰ Presage : Programme de recherche et d'enseignement des savoirs sur le genre. Voir : <http://presage-ofce.sciences-po.fr/la-mission-de-presage.html>

BIBLIOGRAPHIE

- Abdalla R., Teufel S., 2006, « A bootstrapping approach to unsupervised detection of cue phrase variants ». *Proceedings of ACL/COLING*.
- Amossy R., 2010, *L'argumentation dans le discours*, Coll. Cursus, Armand Colin, Paris.
- Andler D., Fagot-Largeault A., Saint-Sernin B., 2002, *Philosophie des sciences*, vol1 et 2, Editions Gallimard, Folio Essais.
- Authier-Revuz J., 1995, *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, t.1 et t.2, Larousse.
- Bally C., 1932, *Linguistique générale et linguistique française*, édition de 1944, Berne, A. Francke SA.
- Bärenfänger M., Lobin H., Lungen H., Hilbert M., 2006, « Using OWL ontologies in discourse parsing », Kühnberger, Kai-Uwe / Mönnich, Uwe (Eds.), *Proceedings of the Workshop of Ontologies in Text Technology. Osnabrück*.
- Benveniste E., 1974, *Problèmes de linguistique générale*, II, Paris.
- Berri J., Cartier E., Desclés J.-P., Jackiewicz A., Minel J.-L., « Filtrage automatique de textes », *Conférence internationale Traitement Automatique des Langues et Applications Industrielles TAL + AI 96*, Moncton, NB, Canada, juin 1996.
- El-Bèze M., Jackiewicz A., Hunston S., (Eds), 2011, « Opinions, sentiments et jugements d'évaluation », *Revue TAL*, n°51 :3.
- Bilhaut F., Ho Dac M., Borillo A., Charnois T., Enjalbert P., Le Draoulec A., Mathet Y., Miguet H., Péry-Woodley M.-P., Sarda L., 2003, « Indexation discursive pour la navigation documentaire: cadres temporels et spatiaux dans l'information géographique », *Actes de TALN 2003*, pp. 315-320.
- Bloom K., Stein S., Argamon, S., 2007, « Appraisal Extraction for News Opinion Analysis at NTCIR-6 ». (http://lingcog.iit.edu/doc/bloom_ntcir2007.pdf).
- Bonami O., Godard D., 2008, « Syntaxe des incises de citation », *Actes du premier Congrès Mondial de Linguistique Française*, pp. 2395-2408.
- Boudon R. 1990, *L'art de se persuader*, Fayard, Paris.
- Bres J., Haillet P., Mellet S., Nølke H., Rosier L., (Eds), 2005, *Dialogisme, polyphonie : approches linguistiques*, Paris, Duculot.
- Brunot F. 1922, *La pensée et la langue*, Masson, Paris.
- Calais F., (Dir.), 2006, *Cohérence et discours*, PUPS.
- Charaudeau P., 1992, *La grammaire du sens et de l'expression*, Paris, Hachette Education.
- Charaudeau P., 1997, *Le discours d'information médiatique*, Paris, Nathan.
- Charaudeau P., Maingueneau D., 2002, *Dictionnaire d'analyse du discours*, Seuil.
- Charolles M., 1995, « Cohésion, cohérence et pertinence du discours », *Travaux de linguistique*, 29, pp. 125-15.
- Charolles M., 1997, L'encadrement du discours : Univers, champs, domaines et espaces, *Cahier de Recherche Linguistique 6*, Université de Nancy2.
- Charolles M., 2002, « Organisation des discours et segmentation des écrits », *Inscription Spatiale du Langage : structures et processus*, 29-30 janvier 2002, IRIT, Toulouse.
- Charolles M., Péry-Woodley M.-P., (Eds), 2005, « Les adverbiaux cadratifs », *Langue française*, n°148, Larousse.
- Chrétien J.L., 2007, *Répondre. Figures de la réponse et de la responsabilité*, Paris, PUF.
- Coirier P., Gaonac'h, D., Passerault, J.-M., 1996, *Psycholinguistique textuelle*, Armand Colin, Paris.

- Combettes B., 1998, *Expressions détachées en français*, Ophrys, Collection L'essentiel Français.
- Condamines A., « Sémantique et corpus, quelles rencontres possibles ? », in A. Condamines (dir), *Sémantique et corpus*, Hermès, Lavoisier, 2005.
- Couto J., Ferret O., Grau B., Hernandez N., Jackiewicz A., Minel J.L., Porhiel S., 2004 « RÉGAL, un système pour la visualisation sélective de documents », *Revue d'Intelligence Artificielle*, vol. 18, n° 4/2004, pp.481-514.
- Crispino G., Jackiewicz A., Minel J.-L., 2004, « Spécification et implantation informatique d'un langage de description de structures discursives », *TALN 2004*, Fès, Maroc.
- Culioli A., 1980, « Valeurs aspectuelles et opérations énonciatives : l'aoristique », in David, J. & Martin, R., *La notion d'aspect*, Paris, Klincksieck, pp. 182-193.
- Culioli A., 1990. *Pour une linguistique de l'Énonciation* - Tomes 1, 2 et 3, Ophrys, coll. l'Homme dans la langue.
- Danlos L., Sagot B., Stern R., 2010, « Analyse discursive des incises de citation », *Actes du 3e Congrès Mondial de Linguistique Française*, La Nouvelle-Orléans, Etats-Unis.
- Dendale, P., Coltier, D., 2005, « La notion de prise en charge ou de responsabilité dans la Théorie Scandinave de la Polyphonie Linguistique », in Bres, J. et al. (éds.), *Dialogisme et polyphonie. Approches linguistiques*, Bruxelles, De Boeck/Duculot, pp. 125-140.
- Desclés J.-P., 1990, *Langages applicatifs, langues naturelles et cognition*, Hermès, Paris.
- Desclés J.-P., 1997, « Systèmes d'exploration contextuelle », *Co-texte et calcul du sens*, Claude Guimier, Ed., Presses Universitaires de Caen, Caen, pp. 215-232.
- Desclés J.-P., 2009, « Prise en charge, engagement et désengagement », *Langue française*, 162, pp. 29-53.
- Desclés J.-P., Guentchéva Z., 1998, « Causality, causativity, transitivity », L. Kulikov, H. Vater (Eds), *Typology of verbal Categories*, Niemeyer, pp. 7-27.
- Desclés J.-P., Guentchéva Z., 2000, « Énonciateur, locuteur, médiateur », A. Becquelin et Ph. Erikson (Eds), *Les rituels du dialogue*, Editions de l'Harmattan.
- Desclés J.-P., Jackiewicz A., 2006, « Abduction et prise en charge de la causalité », *La cause : approche pluridisciplinaire*, *Revue LINX*, n°54.
- Ducrot O., 1984, *Le dire et le dit*, Paris, Éditions de Minuit.
- Eshkol I., 2002, *Typologie sémantique des prédicats de parole*, Paris, Thèse de l'Université de Paris XIII.
- Farrar S. O., Langendoen D. T., 2003, « A linguistic ontology for the SemanticWeb », *GLOT International* 7(3).
- Farrar S. O., Lewis W. D., 2005, « The GOLD Community of Practice, An Infrastructure for Linguistic Data on the Web », *E-MELD Workshop on Digital Language Documentation*, Harvard.
- Farzindar A., Lapalme G., Desclés J.P. 2004, « Résumé de textes juridiques par identification de leur structure thématique », *Revue TAL*, vol. 45, n°1, pp. 39-64.
- Ferrari S., Charnois, T., Jackiewicz A., Gardin P., Widlocher A., 2009, « Jugements d'évaluation : des constituants détachés à d'autres formes régulières ». *6^e Journées JLC*, Lorient, 10-12 septembre 2009.
- Ferret O., Grau B., Minel J.-L., Porhiel S., 2001, « Repérage de structures thématiques dans des textes », *Actes TALN 2001*, Tours, pp. 163-172.
- François J., 1988, *Changement, causation, action*, Droz, Genève.
- Galatanu O., 2000, « Signification, sens, formation », in J.M. Barbier et O. Galatanu (Eds) *Education et Formation. Biennales de l'éducation*, Paris PUF.
- Garcia D., 1998, « Exploration, pour l'élaboration de requêtes de filtrage de textes, des connaissances causales détectées par Coatis », *RIFRA'98*, Tunisie.

- Garfinkel, A., 1981, *Forms of Explanation, Rethinking the Questions of Social Theory*, New Haven, Yale University Press.
- Gaulmyn de, M.M., 1983, *Les verbes de communication dans la structuration du discours. Essai sur la réflexivité du langage*, Thèse d'Etat, Paris VIII.
- Giguet E, Lucas N., 2004, « La détection automatique des citations et des locuteurs dans les textes informatifs », In (López Muñoz et al.2004).
- Goecke D., Lüngen H., Sasak F., Witt A., Farrar S. O., 2005, « GOLD and Discourse : Domain and Community-Specific Extensions», *Proceedings of the 2005 E-MELD-Workshop. Boston, MA*.
- Goody J., 1979, *La raison graphique. La domestication de la pensée sauvage*. Les Editions de Minuit, traduit de l'anglais et présenté par J. Bazin et A. Bensa.
- Grevisse M., 1993, *Le bon usage*, DeBoeck&Duculot, 13^e édition.
- Grize B., 1974, « Du groupement au nombre » in Problèmes de la construction du nombre, *Etudes d'épistémologie génétique*, n°XI, pp. 69-96, K.Reprint, Nendeln/Lienchtenstein.
- Gross G., 1996, Une typologie sémantique des connecteurs : l'exemple de la cause, *Studi Italiani di Linguistica Teorica e Applicata*, anno XXV, n°1.
- Gross G., Nazarenko A., 2004, « Quand la langue cause : paramètres d'une analyse linguistique », *Intellectica* n° 38.
- Groupe λ-I, 1975, « Car, parce que, puisque », *Revue Romane*, X, n° II, pp.248-280.
- Guentchéva Z., (Ed), 1997, *La catégorie grammaticale du médiatif*, Peters.
- Guimier C., 1996, *Les adverbes du français*, Ophrys, Collection L'essentiel Français.
- Habermas J., 1987, *Théorie de l'agir communicationnel*, t.1, Paris, Librairie Arthème Fayard pour la traduction française.
- Habert B., *Construire des bases de données pour le français*, Ophrys, 2009.
- Hernandez N., Grau B., 2005, « Détection automatique de structures fines du discours », *TALN2005*, Dourdan.
- Hirsch J.E., 2005, « An index to quantify an individual's scientific research output », *Proceedings of the National Academy of Sciences*, vol. 102, pp.16569-16572.
- Ho-Dac L.-M., Péry-Woodley M.-P., Tanguy L., 2010, « Anatomie des structures énumératives », *TALN 2010*, Montréal, Québec, 19-23 juillet 2010.
- Jackendoff R., 1990, *Semantics and Cognition*, Cambridge (Mass), M.I.T. Press.
- Jackiewicz A., 1996, « La notion de la cause pour le filtrage de phrases importantes d'un texte » *Conférence internationale Traitement Automatique des Langues et Applications Industrielles TAL + AI 96*, Moncton, NB, Canada, juin 1996.
- Jackiewicz A., 1998, *L'expression de la causalité dans les textes. Contribution au filtrage sémantique par une méthode informatique d'exploration contextuelle*. Thèse de doctorat. Université de Paris IV.
- Jackiewicz A., 1999a, « La causalité dans les textes », *Sémantique et confrontation des langues*, n°2, pp.147-164, Académie polonaise des Sciences, Varsovie, Pologne.
- Jackiewicz A., 1999b, « Causalité et prise en charge énonciative », *Etudes Cognitives*, n°3, pp.249-269, Académie polonaise des Sciences, Varsovie, Pologne.
- Jackiewicz A., 2002, « Repérage et délimitation des cadres organisationnels pour la segmentation automatique des textes », *CIFT'02*, Hammamet, Tunisie, pp. 95-107.
- Jackiewicz A., 2004, « La causalité dans la langue : une question de point(s) de vue », *Revue Intellectica*, 2004/1, 38, pp. 43-67.
- Jackiewicz A., 2005, « Les séries linéaires dans le discours », *Revue Langue française*, « Les adverbiaux cadratifs », 2005/4 - n° 148, pages 95 à 110.
- Jackiewicz A., 2005, « Transmission dialogique des paroles d'autrui », *Les adverbiaux cadratifs et l'organisation des textes*, *Revue Verbum*, tome XXVII, n°3, pp. 265-291.

- Jackiewicz A., 2006, « Relations intersubjectives dans les discours rapportés », *Discours et document : traitements automatiques*, Revue TAL, Vol. 47, n°2, pp.65-87.
- Jackiewicz A., 2009, « L'évaluation à la périphérie du prédicat : constructions, lexiques et relations sémantiques », *Arena Romanistica n°4, actes du 28^e Colloque international sur le lexique et la grammaire*, Bergen, Norvège.
- Jackiewicz A., 2010, « Structures avec constituants détachés et jugements d'évaluation », *Document numérique*, vol. 13, n° 3, pp. 11-40.
- Jackiewicz A., 2011, « Formes de responsabilité dans les discours rapportés », in Patrick Dendale (dir), *La notion de prise en charge en linguistique*, Collection Champs linguistiques, De Boeck/Duculot, pp. 93-115.
- Jackiewicz A., Minel J.-L., 2003, « L'identification des structures discursives engendrées par les cadres organisationnels », *TALN 2003*, Batz sur Mer.
- Jackiewicz A., Laublet, P., 2007, « Web sémantique et linguistique du discours. Ontotexte : un tournant énonciatif », *Journée Ontologies et textes, TIA*, Sophia Antipolis.
- Jackiewicz A., Ferrari S., Charnois T., 2009a, « Jugements d'évaluation et constituants périphériques ». *TALN'09*, Senlis.
- Jackiewicz A., Vigier D., Charnois, T., Ferrari S., 2009b, « Vers une analyse automatique des discours évaluatifs. Le cas des constituants détachés 'en N <exp>' ». *LPTS 2009*. Paris, 21-23 septembre 2009.
- Jacob P., 1997, *Pourquoi les choses ont-elles un sens ?* Editions Odile Jacob.
- Jacques M.-P., Rebeyrolle J., 2006, « Titres et structuration des documents », *ISDD2006*, Caen.
- Kerbrat-Orecchioni C., 2002, *L'énonciation*, Armand Colin.
- Kordi E., 1988, *Modal'nye i kausativnye glagoly v sovremennom francuzskom jazyke, (Verbes modaux et causatifs en français contemporain)*, Leningrad Nauka.
- Lamiroy B., Charolles, M., 2008, « Les verbes de parole et la question de l'(in)transitivité », *Discours N° 2*. (<http://discours.revues.org/index3232.html>)
- Largeault J., 1985, *Systèmes de la nature*. J. Vrin, Paris.
- Laublet P., Naït-Baha L., Jackiewicz A., Djioua, B., 2002, « Collecte d'informations textuelles sur le Web selon différents points de vue », *Interaction Homme-Machine et Recherche d'Informations*, pp.261-293, Éditions Hermès, Paris.
- Laublet P., Lambert M., Jackiewicz A., 2010, « Ontologies pour l'annotation d'actes évaluatifs et de leurs cibles » Poster, *Journées Francophones d'Ingénierie des Connaissances (IC)*, Sylvie Despres (Eds.), Nîmes, Juin 2010.
- Laurendeau P., 1989, « Repérage énonciatif et valeur de vérité : la prise en compte, la prise en charge », Vincent, D. & Saint-Jacques, D., (Eds), *Des analyses de discours, Actes du CELAT, 2*, Québec, Université Laval, pp.107-129.
- Laurendeau P., 2009, « Préassertion, réassertion, désassertion : construction et déconstruction de l'opération de prise en charge », *Langue française*, 162, pp. 55-70.
- Le Monde*, 2002, *Le style du Monde*, Guide rédactionnel.
- López Muñoz J.-M., Marnette, S., Rosier L., (Eds.), 2004, *Discours rapporté dans tous ses états : question de frontières*, Paris. L'Harmattan.
- López Muñoz J.-M. , Marnette, S. , Rosier L. , Vincent D., 2010, *La circulation des discours*, Nota Bene (Editions), Collection Langue et pratiques discursives.
- Luc C., Virbel J., 2001, « Le modèle d'architecture textuelle : fondements et expérimentation ». *Verbum*, 23(1), pp. 103-123.
- Maingueneau D., 2004, « Hyperénonciateur et 'participation' », *Langages*, n°156, pp.111-126.
- Maingueneau D., 2009, *Aborder la linguistique*, Editions du Seuil, Paris.
- Mann W.C., Thompson S.A., 1988, « Rhetorical Structure Theory : toward a functional theory of text organization », *Text* 8 (3), p. 243-281.

- Martin J.R., White, P.R.R., 2005, *The language of evaluation. Appraisal in English*, Hampshire, Palgrave Macmillan.
- McCawley J.D., 1976, « Remarks on What Can Cause What, *Syntax and Semantics*», V.6, p. 117-129, Academic Press.
- Miéville D., 1984, *Un développement des systèmes logiques de S. Leśniewski. Protothétique-Ontologie-Méréologie*, Berne/Francfort/New York, Peter Lang.
- Miéville D., 1992, « Objet de discours et organisation raisonnée », *Les organisations raisonnées*, Travaux du Centre de Recherches Sémiologiques, n°60, pp.27-53, Université de Neuchâtel.
- Minel J.-L., Cartier E., Crispino G., Desclés J.-P., Ben Hazez S., Jackiewicz A., 2001, « Résumé automatique par filtrage sémantique d'informations dans des textes », *Technique et Science Informatiques*, n°3, pp.369-396, Éditions Hermès, Paris.
- Minel J.-L., 2002, *Filtrage sémantique. Du résumé automatique à la fouille de textes*, Hermès, Lavoisier.
- Molinier Ch. Levrier F. 2002, *Grammaire des adverbes : description des formes en -ment*, Droz.
- Moscovici S., (Ed), 1994, *Psychologie sociale des relations à autrui*, Paris, Nathan Université.
- Mourad G., Desclés J.P., 2004, « Identification et extraction automatique des informations citationnelles dans un texte », In (López Muñoz et al.2004).
- Nazarenko A., 2000, *La cause et son expression en français*, Ophrys, Paris.
- Nedjalkov V.P., Silnickij G.G., 1969, *Tipologija kauzativnyx konstrukcij*, Xolodovitch (Ed.).
- Neuberg M. (éd.), 1997, *La responsabilité. Les questions philosophiques*, Paris, PUF.
- Neveu F., 1998, *Etudes sur l'apposition. Aspects du détachement nominal et adjectival en français contemporain, dans un corpus de textes de J.-P. Sartre*. Honoré Champion Editeur, Paris, 1998.
- Neveu F. (éd), 2000, « Nouvelles recherches sur l'apposition », *Langue française* n°125, Larousse.
- Neveu, F., 2002, « L'ajout et la problématique appositive », J. Authier-Revuz et M.Ch. Lala, eds. *Figures d'ajout phrase, texte, écriture*, Presses Sorbonne Nouvelle, pp.111-122.
- Neveu F., 2004, *Dictionnaire des sciences du langage*, Armand Colin, 2004.
- Nølke H., Fløttum K., Norén C. (Eds), 2004, *ScaPoLine. La théorie scandinave de la polyphonie linguistique*, Paris, Kimé.
- Pang B., Lee L., 2008, « Opinion Mining and Sentiment Analysis », *Foundations and Trends in Information Retrieval*, vol. 2, p.1-135.
- Park, W., Abraham, M., 2001, « Acquisition of Causal Information from Medical Documents in French for Experts », *World Congress on Medical Physics and Biomedical Engineering*, Turquie.
- Perelman C., Olbrechts-Tyteca, L., 1992, *Traité de l'argumentation*, (1958), Editions de Bruxelles.
- Péry-Woodley M.-P., 2000, *Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle*, Mémoire d'habilitation, Université de Toulouse-le Mirail.
- Péry-Woodley M.-P., Asher N., Enjalbert P., Benamara F., Bras M., Fabre C., Ferrari S., Ho-Dac L.-M., Le Draoulec A., Mathet Y., Muller P., Prévot L., Rebeyrolle J., Tanguy L., Vergez-Couret M., Vieu L., Widlöcher A., 2009, « ANNODIS : une approche outillée de l'annotation de structures discursives », *TALN'09*, 24-26 juin 2009, Senlis, France.
- Piaget J., 1974, « La construction au nombre », Problèmes de la construction du nombre, *Etudes d'épistémologie génétique*, n°XI, pp. 1-68, K. Reprint, Nendeln / Lienchtenstein.
- Pineira,C., Tournier M., 1990, « De quel bois se chauffe-t-on? Origines et contextes actuels de l'expression langue de bois dans la presse française », *Mots* n°21, pp. 5-20.
- Plantin C., 1990, *Essais sur l'argumentation*, Editions Kimé, Paris.

- Pottier B., 2001, *Représentations mentales et catégorisations linguistiques*, Editions Peters, Louvain-Paris.
- Poulard F., Waszak T., Hernandez N., Bellot P., 2008, « Repérage de citations, classification des styles de discours rapport et identification des constituants citationnels en écrits journalistiques », *Actes de TALN 2008*, Avignon, France.
- Poulard F., Hernandez N., Afantenos S.D., Daille B., 2010, « Évaluation de descripteurs statistiques et linguistiques pour la détection de dérivation de texte », *Document numérique*, *Document numérique*, vol. 13, n° 3/2010, pp. 69-93.
- Puech M., 1990, *Kant et la causalité*, J. Vrin, Paris.
- Pustejovsky J., 1995, *The Generative Lexicon*, The MIT Press.
- Quintin E. Jackiewicz A., Roy T., 2011, « Analyse d'énoncés politiques à fort impact et éléments pour leur détection automatique. », in J. Bres, A. Nowakowska, J.-M. Sarale, S. Sarrazin (coord.) *Actes du colloque international Dialogisme : langue, discours* (8-10 septembre 2010, Montpellier), mis en ligne le 10 juillet 2011, consulté le 01/09/2011. URL : http://recherche.univ-montp3.fr/praxiling/IMG/pdf_QuintinJackiewiczRoy1.pdf.
- Rabatel A., (Ed), 2004, « Effacement énonciatif et discours rapportés », *Langages* n°156, décembre 2004.
- Rabatel A., 2006, « L'effacement de la figure de l'auteur dans la construction événementielle d'un « journal » de campagne électorale et la question de la responsabilité, en l'absence de récit primaire », *Semen*, 22. (<http://semen.revues.org/document2843.html>).
- Rabatel A. & Chauvin-Vileno, A., 2006, « La question de la responsabilité dans l'écriture de presse », *Semen*, 22. Consultable sur <http://semen.revues.org/document2792.html>.
- Rezsohazy R., 2006, *Sociologie des valeurs*, Armand Colin.
- Ricoeur P., 1990, *Soi-même comme un autre*, Paris, Editions du Seuil.
- Rosier L., 1999, *Le discours rapporté. Histoire, théories, pratiques*, Bruxelles / Paris, Duculot.
- Rosier L., 2005, « Chaînes d'énonciateurs et modes d'organisation textuels : du discours rapporté à la circulation re-marquée des discours », *Cahiers de Praxématique*, n°45, pp. 103 -124.
- Rosier L., 2008, *Le discours rapporté en français*, Paris, Ophrys.
- Rossari C., 1997, *Les opérations de reformulation*, Berne, Peter Lang.
- Schnedecker C., 2006, *De l'un à l'autre et réciproquement*, De Boeck, Belgique.
- Somasundaran S., Wilson T., Wiebe J. Stoyanov V., 2007, « Qa with attitude : Exploiting opinion type analysis for improving question answering in on-line discussions and news », *International Conference on Weblogs and Social media (ICWSM'07)*.
- Soutet O., 2005, *Linguistique*, PUF.
- Svensson M., 2010, *Marqueurs corrélatifs en français et en suédois*. *Studia Romanica Upsaliensia*.
- Talmy L., 1988, « Force Dynamics in Language and Cognition », *Cognitive Science*, Vol. 12, pp.49-100.
- Teufel S., 2006, « Discourse structure in scientific articles: argumentation and citation », *ISDD2006*, Caen.
- Teufel S., Siddharthan A., Tidhar D., 2006, « Automatic classification of citation function », In *EMNLP*, Sydney, Australia.
- Turco G., Coltier D., 1988, « Des agents doubles de l'organisation textuelle, les marqueurs d'intégration linéaire », *Pratiques*, n°57.
- Vandendorpe C., 1999, *Du papyrus à l'hypertexte. Essai sur les mutations du texte et de la lecture*, Editions de la Découverte.
- Vanderveken D., 1988, *Les actes de discours : essai de philosophie du langage et de l'esprit sur la signification des énonciations*, Bruxelles, Mardaga.

- Vernier M., Monceaux L., Daille B., 2009a, « DEFT'09 : détection de la subjectivité et catégorisation de textes subjectifs par une approche mixte symbolique et statistique », Actes de l'atelier de clôture du 5^{ème} Défi Fouille de Textes (DEFT'09), Paris, France, pp. 101-112.
- Vernier M., Monceaux L., Daille B., Dubreil E., 2009b, « Catégorisation des évaluations dans un corpus de blogs multi-domaine », *Numéro spécial de la revue RNTI – Fouille de données d'opinion*, pp.45-70.
- Vernier M., Monceaux L., 2010, « Enrichissement d'un lexique de termes subjectifs à partir de tests sémantiques », *revue TAL*, n°51:1, pp.125-149.
- Vigier D., Terran E., (Eds), 2005, « Les adverbiaux cadratifs et l'organisation des textes », *Verbum*, Tome XXVII, n°3, Presses Universitaires de Nancy.
- Wiebe J., Wilson, T. Cardie, C., 2005, « Annotating Expressions of Opinions and Emotions in Language », *Language Resources and Evaluation*, 39, 2-3, pp.165-210.
- Wierzbicka A., 1997, « Les universaux d'ordre lexical et grammatical : une fenêtre sur les structures conceptuelles », *XVIe Congrès International des Linguistes*, Paris.

ANNEXES

1 - CURRICULUM VITAE

Agata JACKIEWICZ - DESBERTRAND

Née le 28 septembre 1965, (Gdansk, Pologne)
Mariée, 2 enfants ; nationalité française.

44, rue Raymond Losserand
75014 Paris
Tél. 01 45 41 64 89
Agata.Jackiewicz@paris-sorbonne.fr

SITUATION ACTUELLE ET RESPONSABILITES

Maître de conférences à l'Université de Paris-Sorbonne (Paris IV, UFR ISHA) en Linguistique informatique (section 27), depuis 1998.

Membre de l'équipe STIH, Paris-Sorbonne, (depuis 1994 : UMR 27 CAMS, LALIC, LALICC).

Membre du Conseil de l'UFR Institut des Sciences Humaines Appliquées (ISHA), depuis 2000.

Directrice adjointe de l'UFR Institut des Sciences Humaines Appliquées (ISHA), de septembre 2006 à juin 2011.

Directrice des études du Master 1 « Information et Communication » Spécialité Recherche *Informatique et Ingénierie de la langue pour la gestion de l'information* IILGI et Spécialité Professionnelle *Ingénierie de la langue pour la gestion intelligente de l'information* ILGI depuis septembre 2005.

FORMATION ET TITRES UNIVERSITAIRES

Thèse de Doctorat en Linguistique informatique, « *L'expression de la causalité dans les textes. Contribution au filtrage sémantique par une méthode informatique d'exploration contextuelle* », mention « Très honorable avec félicitations, à l'unanimité », Université de Paris-Sorbonne, 21 février 1998.

Jury : Jean-Pierre Desclés, Paris IV, (directeur de thèse) ; Jean-Marie Pierrel, Nancy I, (rapporteur) ; Michel Charolles, Nancy II, (rapporteur) ; Gaston Gross, Paris-Nord, (président) ; Raymond Boudon, Paris IV, membre de l'Institut ; Jacques Blin, société CR2A-DI.

La préparation de la thèse s'est déroulée dans le cadre d'une convention CIFRE (n°004/1994, informatique) entre le laboratoire CAMS (Centre d'Analyse et de Mathématiques Sociales, unité mixte de recherche 17 du CNRS, de l'EHESS et de l'Université de Paris-Sorbonne) et la société EDIAT / CR2A-DI (Courbevoie).

DEA en Sciences Cognitives, EHESS, Paris VI et Ecole Polytechnique, mention Bien, 1992, Mémoire (dir. Jean-Pierre Desclés) : « *Contributions aux problèmes de l'extraction de connaissances : manifestations linguistiques et représentation informatique de la relation partie-tout* ».

Diplôme Universitaire des Sciences Cognitives pour l'ingénierie de la Connaissance, (Composantes : sémantique, informatique, psychologie expérimentale et cognitive, neuro-psycholinguistique), Université de Toulouse-Le Mirail, mention Très Bien, 1988-1990.

Maîtrise en Sciences du Langage, Université de Toulouse-Le Mirail, mention Très Bien, 1990.
Mémoire (dir. Andrée Borillo) : « *Description et classification d'adverbiaux de référence temporelle* ».

DUT d'Informatique de gestion, IUT de Rodez, 1988.

Cybernétique économique et Informatique, option Statistiques et Econométrie, Université de Gdansk en Pologne, première année validée avec mention Très Bien, 1986.

Baccalauréat (Série générale, option Mathématiques), Lycée d'enseignement général n°IV, Tadeusz Kosciuszko, Torun, Pologne, mention Très Bien, 1985.

LANGUES

Polonais, Anglais, Russe

AUTRES EXPÉRIENCES PROFESSIONNELLES

Société **INFORSUD** à Causse Comtal (12) : Conception et réalisation d'un logiciel de gestion d'association en DBASE III (stage IUT), avril – septembre 1988.

Société **EC2** (Edition, Colloques & Conseil) à Nanterre, en collaboration avec le Centre de Terminologie et de Néologie CTN du CNRS : Rédaction du *Vocabulaire d'Intelligence Artificielle*, publié par EC2 en mai 1991, février - juin 1991.

Société **EDIAT** (Européenne d'Intelligence Artificielle) à Boulogne, en collaboration avec le laboratoire CAMS (EHESS/CNRS) :

Contribution à la conception et à la réalisation d'un système expert d'exploration contextuelle SEEK, intégré à l'atelier de génie cognitif, supportant la méthode d'acquisition et de modélisation des connaissances METODAC.

Réalisation d'une pré-étude pour la Direction d'Etudes et de Recherches d'EDF Châtou : « Conceptualisation, modélisation et capitalisation des connaissances extraites des documents techniques », janvier - mai 1992 et avril - novembre 1993.

2 - PUBLICATIONS

CHAPITRES D'OUVRAGE OU ARTICLES DANS OUVRAGE COLLECTIF

- Agata JACKIEWICZ, « La causalité dans les textes », in *Semantyka i konfrontacja jezykowa (Sémantique et Confrontation des Langues)*, Varsovie, Pologne, SOW, Académie des Sciences de la Pologne, 1999, vol.2, pp.147-164, (18 pages).
- Philippe LAUBLET, Leïla NAÏT-BAHA, Agata JACKIEWICZ, Brahim DJIOUA, in Céline Paganelli (dir) « Collecte d'informations textuelles sur le Web selon différents points de vue », ch.8, *Interaction homme-machine et recherche d'informations*, Paris, Éditions Hermès, Lavoisier, 2002, ch.8, pp.261-296, (36 pages).
- Agata JACKIEWICZ, « Formes de responsabilité dans les discours rapportés », in Patrick Dendale (dir), *La notion de prise en charge en linguistique*, Bruxelles, De Boeck/Duculot, Collection Champs linguistiques, 2011, pp. 93-115, (23 pages).

REVUES

- Agata JACKIEWICZ, « L'expression lexicale de la relation d'ingrédience », *Faits de langues* Ophrys, 1996, n°7, pp.53-62.
- Jawad BERRI, Emmanuel CARTIER, Jean-Pierre DESCLÉS, Agata JACKIEWICZ, Jean-Luc MINEL, « Safir, système automatique de filtrage de textes », in Jean-Pierre Desclés et Jong-Chul Soh (dir) *Langages, Cognition et Texte*, vol.2, Université Hankuk des langues étrangères, Université de Paris-Sorbonne, Séoul, Corée du Sud, 1997.
- Agata JACKIEWICZ, « Causalité et prise en charge énonciative », *Etudes Cognitives*, Académie des Sciences de la Pologne, Varsovie, Pologne, 1999, vol.3, pp. 249-270.
- Jean-Luc MINEL, Jean-Pierre DESCLÉS, Emmanuel CARTIER, Gustavo CRISPINO, Slim BEN HAZEZ, Agata JACKIEWICZ, « Résumé automatique par filtrage sémantique d'informations dans des textes », *Revue TSI*, Hermès, 2001, vol.20, n°3, pp.369-396.
- Leïla NAIT-BAHA, Agata JACKIEWICZ, Brahim DJIOUA, Philippe LAUBLET, « Query Reformulation for Information Retrieval on the Web Using the Point of View Methodology : Preliminary Results », *International Journal Knowledge Organization*, 2001/28, n°3, pp.129-136.
- Jawier COUTO, Olivier FERRET, Brigitte GRAU, Nicolas HERNANDEZ, Agata JACKIEWICZ, Jean-Luc MINEL, Sylvie PORHIEL, « RÉGAL, un système pour la visualisation sélective de documents », *Revue d'Intelligence Artificielle*, 2004, vol. 18, n°4, pp.481-514.
- Agata JACKIEWICZ, « La causalité dans la langue : une question de point(s) de vue », *Intellectica*, Revue de l'Association pour la Recherche Cognitive, 2004/1, n°38, pp. 43-67.
- Agata JACKIEWICZ, « Les séries linéaires dans le discours », In *Les adverbiaux cadratifs, Langue française*, 2005/4, n°148, pp. 95 -110.
- Agata JACKIEWICZ, « Transmission dialogique des paroles d'autrui », In *Les adverbiaux cadratifs et l'organisation des textes, Verbum*, 2005/XXVII, n°3, pp. 265-291.
- Jean-Pierre DESCLÉS, Agata JACKIEWICZ, « Abduction et prise en charge de la causalité », In *La cause : approche pluridisciplinaire, LINX*, 2006, n°54, pp.35-47.
- Agata JACKIEWICZ, « Relations intersubjectives dans les discours rapportés », In *Discours et document : traitements automatiques, TAL*, 2006, vol. 47, n°2, pp.65-87.

Agata JACKIEWICZ, « L'évaluation à la périphérie du prédicat : constructions, lexiques et relations sémantiques », *Arena Romanistica*, 2009, n°4, pp.150-159.

Agata JACKIEWICZ, « Structures avec constituants détachés et jugements d'évaluation », In *Le texte : objet d'analyse et vecteur de connaissances, Document Numérique*, 2010, vol.13, n°3, p.11-40.

Agata JACKIEWICZ, Susan HUNSTON, Marc EL-BÈZE, Préface : « Opinions, sentiments et jugements d'évaluation », *TAL*, 2010, vol. 51, n°3, pp.7-17.

CONFÉRENCES INTERNATIONALES AVEC COMITÉ DE LECTURE ET PUBLICATION DES ACTES

Cas particuliers :

- publication dans une revue <R>
- publication dans un ouvrage collectif <O>
- publication d'un résumé long <S>

Agata JACKIEWICZ, « La notion de la cause pour le filtrage de phrases importantes d'un texte », *Conférence internationale Traitement Automatique des Langues et Applications Industrielles TAL + AI 96*, Poster + papier court, Moncton, NB, Canada, juin 1996, (5 pages, double colonne).

Jawad BERRI, Emmanuel CARTIER, Jean-Pierre DESCLÉS, Agata JACKIEWICZ, Jean-Luc MINEL, « Filtrage automatique de textes », *Conférence internationale Traitement Automatique des Langues et Applications Industrielles TAL + AI 96*, Moncton, NB, Canada, juin 1996, (7 pages, double colonne).

Jawad BERRI, Emmanuel CARTIER, Agata JACKIEWICZ, Jean-Luc MINEL, « A Linguistic Method for Text-Filtering », *Procesamiento del Lenguaje Natural, Revista n°19*, Séville, Espagne, septembre 1996, (7 pages).

Jean-Pierre DESCLÉS, Emmanuel CARTIER, Agata JACKIEWICZ, Jean-Luc MINEL, « Textual processing and Contextual Exploration Method », *International and Interdisciplinary Conference on Modeling and Using Context, CONTEXT'97*, Rio de Janeiro, Brésil, février 1997, (10 pages).

Leila NAIT-BAHA, Agata JACKIEWICZ, Philippe LAUBLET « Causalité, recherche et filtrage d'informations sur le Web », *Rencontre Internationale sur l'extraction, le filtrage et le résumé automatiques RIFRA'98*, Tunisie, novembre 1998, (11 pages).

Agata JACKIEWICZ, « Repérage et délimitation des cadres organisationnels pour la segmentation automatique des textes », *Actes de CIFT'02*, Hammamet, Tunisie, 2002, pp. 95-107, (6 pages, double colonne).

Gustavo CRISPINO, Agata JACKIEWICZ, Jean-Luc MINEL. « Spécification et implantation informatique d'un langage de description de structures discursives », *TALN 2004*, Fès, Maroc, 2004, (6 pages).

Brahim DJIOUA, Jorge GARCIA-FLORES, Antoine BLAIS, Jean Pierre DESCLÉS, Gaëll GUIBERT, Agata JACKIEWICZ, Florence LE PRIOL, Leila NAIT-BAHA, Benoît SAUZAY, « EXCOM: an automatic annotation engine for semantic information », *FLAIRS 2006*, Melbourne, Floride, USA, 11-13 mai 2006, (6 pages, double colonne).

Florence LE PRIOL, Antoine BLAIS, Jean-Pierre DESCLÉS, Brahim DJIOUA, Jorge GARCIA-FLORES, Gaëll GUIBERT, Agata JACKIEWICZ, Leila NAIT-BAHA, Benoît SAUZAY, « Automatic annotation of localization and identification relations in platform EXCOM » *FLAIRS 2006*, Melbourne, Floride, USA, 11-13 mai 2006, (6 pages, double colonne).

Agata JACKIEWICZ, « Objets de la prise en charge dans le discours rapporté », Colloque international « Prise en charge / commitment », Anvers, Belgique, 11-13 janvier 2007, <O>, (17 pages)

- Agata JACKIEWICZ. « L'évaluation à la périphérie du prédicat : constructions, lexiques et relations sémantiques », *28e Colloque international sur le Lexique et la Grammaire*, Bergen, Norvège, 28 septembre-2 octobre 2009, <R> (10 pages).
- Stéphane FERRARI, Thierry CHARNOIS, Agata JACKIEWICZ, Pierre GARDIN, Antoine WIDLÖCHER, « Discours évaluatif : une campagne d'annotation pour la validation de patrons », *6^e Journées Internationales de Linguistique de Corpus*, JLC, Lorient, 10-12 septembre 2009, (13 pages).
- Agata JACKIEWICZ, Denis VIGIER, Thierry CHARNOIS, Stéphane Ferrari, « Vers une analyse automatique des discours évaluatifs. Le cas des constituants détachés 'en N <exp>' », Colloque international Approches linguistiques et psycholinguistiques de la structuration des textes, LPTS'09, Lattice, ENS Paris, 21-23 septembre 2009, <S> (3 pages).
- Dominique LEGALLOIS, Stéphane FERRARI, Agata JACKIEWICZ, « Modéliser le dialogisme pour le traitement automatique des langues et l'analyse du discours. », *Colloque international Dialogisme : langue, discours*, Montpellier, 8-11 septembre 2010. <S>, (4 pages).
- Emmanuel QUINTIN, Agata JACKIEWICZ, Thibault ROY, « Analyse d'énoncés politiques à fort impact et éléments pour leur détection automatique. », in J. Bres, A. Nowakowska, J.-M. Sarale, S. Sarrazin (coord.) Actes du colloque international Dialogisme : langue, discours (8-10 septembre 2010, Montpellier), mis en ligne le 10 juillet 2011, consulté le 01/09/2011. URL : http://recherche.univ-montp3.fr/praxiling/IMG/pdf_QuintinJackiewiczRoy1.pdf. (12 pages).

CONFÉRENCES NATIONALES AVEC COMITÉ DE LECTURE ET PUBLICATION DES ACTES

- Daniela GARCIA, Agata JACKIEWICZ, « Aide à l'acquisition des connaissances causales par exploration de textes », *Journées d'Acquisition des Connaissances JAC 95*, Grenoble, avril 1995, (14 pages).
- Agata JACKIEWICZ, « Le filtrage d'informations textuelles par une approche contextuelle de la causalité », *Rencontre des étudiants chercheurs en informatique pour le Traitement Automatique de la Langue RECITAL'96*, Dourdan, septembre 1996.
- Jawad BERRI, Emmanuel CARTIER, Jean-Pierre DESCLÉS, Agata JACKIEWICZ, Jean-Luc MINEL, « SAFIR, système automatique de filtrage de textes », *Conférence Traitement Automatique du Langage Naturel TALN'96*, Marseille, mai 1996, pp.140-149, (10 pages).
- Jean-Pierre DESCLÉS, Emmanuel CARTIER, Agata JACKIEWICZ, Jean-Luc MINEL, « Filtrage Automatique de Textes : L'exemple des Énoncés Définitoires », *Colloque JST Francil '97*, Avignon, avril 1997.
- Leila NAIT-BAHA, Agata JACKIEWICZ, Brahim DJIOUA, Philippe LAUBLET « Reformulation de requêtes pour la collecte d'informations sur le Web à partir de points de vue : premiers résultats », *ISKO'99*, 1999, (7 pages).
- Agata JACKIEWICZ, Jean-Luc MINEL, « L'identification des structures discursives engendrées par les cadres organisationnels », *TALN 2003*, Batz sur Mer, 2003, (10 pages).
- Agata JACKIEWICZ, Philippe LAUBLET « Web sémantique et linguistique du discours. OntoDiscours : un tournant énonciatif », *Conférence TIA'07, Journée Ontologies et textes*, INRIA Sophia Antipolis, 10 octobre 2007, (10 pages).
- Agata JACKIEWICZ, Thierry CHARNOIS, Stéphane FERRARI, « Jugements d'évaluation et constituants périphériques », *TALN09, 16e conférence sur le traitement automatique des langues naturelles*, Senlis, France, 24-26 Juin 2009, (10 pages).
- Philippe LAUBLET, Mathias LAMBERT, Agata JACKIEWICZ, « Ontologies pour l'annotation d'actes évaluatifs et de leurs cibles » Poster, *Journées Francophones d'Ingénierie des Connaissances (IC)*, Sylvie Despres (Eds.), Nîmes, Juin 2010, (6 pages).

COMMUNICATIONS INVITEES

- « Causalité, prise en charge énonciative et argumentation », *Séminaire du groupe Argumentation en sciences sociales*, Laboratoire GEMAS, Université de Paris-Sorbonne, Paris, 14 janvier 1999.
- « Les séries linéaires dans le discours. Application à la segmentation automatique de textes ». *Séminaire de l'équipe ERESS*, Université de Toulouse le Mirail, Toulouse, 16 juin, 2003.
- « Subjectivité et intersubjectivité dans l'expression de la causalité », *Séminaire du Laboratoire LIDILEM*, Université de Grenoble 3, 6 juin 2008.
- « Structures avec constituants détachés et jugements d'évaluation », Katholieke Universiteit Leuven, Département de Linguistique (Langues romanes), Louvain, Belgique, 6 mai 2010.

COMMUNICATIONS DANS DES JOURNÉES D'ETUDE

- « SERAPHIN, système d'activité résumante. Le cas de la causalité », *Journée Filtrage et résumé automatique de textes*, Direction de l'Information Scientifique des Technologies Nouvelles et des Bibliothèques, Paris, 13 mai 1996.
- « La causalité et la prise en charge énonciative », *Rencontres dans le cadre de la Coopération franco-polonaise entre l'Académie Polonaise des Sciences (Varsovie), l'École Normale de Cracovie et l'Université de Paris-Sorbonne*, Paris, 28 novembre 1998.
- « La causalité pour le résumé et le filtrage des textes », *Journée Filtrage automatique des textes organisée dans le cadre de la coopération avec les chercheurs de l'Université Lyon II Lumière*, Université de Paris-Sorbonne, Paris, 25 janvier 1999.
- « La causalité pour la collecte d'informations sur le Web », *Journée d'étude LaLIC, LANCI et IDIST*, Université de Paris-Sorbonne, Paris, 23 février 1999.
- « Le point de vue de la causalité dans le filtrage des informations sur le Net », *Journée Scientifique : Des industries de la langue à la fouille des textes*, ENST Bretagne, Brest, 25 février 1999.
- « L'exploration contextuelle et le système Context0 », *Séminaire de recherche : Les expressions introductrices de cadres de discours et leur portée textuelle*, Paris, Université de Paris III, 23 septembre 1999.
- « Les marqueurs d'intégration linéaire. Etude sur un corpus ». *Projet REGAL*, Laboratoire Limsi, Orsay, 14 juin 2001.
- « Repérage et délimitation des cadres organisationnels. Etude des marqueurs d'intégration linéaire », *Séminaire de recherche : Les expressions introductrices de cadres de discours et leur portée textuelle*, Paris, Université de Paris III, 15 novembre 2002.
- « Repérage et délimitation des séries dans le discours. Etude des marqueurs d'intégration linéaire », *Séminaire franco-albanais (coop avec l'Université de Tirana)*, LaLIC, Université de Paris-Sorbonne, 27 mars 2003.
- « Exploration contextuelle : aspects linguistiques. Le cas de l'énonciation ». *Journée d'étude de l'équipe LaLICC/ Celsa*, Université de Paris-IV, 26 novembre 2003.
- « Les séries linéaires dans le discours : marques, opérations et structures sous-jacentes », *Journée ATALA*, La Rochelle, 30 avril 2004.
- « Repérage d'objets textuels et leur annotation sémantique », *séminaire LaLICC / GREYC / ERSS*, Université de Paris –Sorbonne. Paris, 9 mars 2004.
- « Décrire et modéliser des organisations textuelles. Le cas des séries linéaires », *Journées LaLICC*, 1-2 octobre 2004, Université de Paris –Sorbonne. Paris.
- « La causalité prise en charge », *Journée sur la causalité*, Université de Paris X, Nanterre, 18 juin 2005.
- « Identification automatique de relations causales dans des textes français », *Journée d'étude sur la causalité*, l'École Supérieure de Traducteurs et Interprètes à Gand, Belgique, le 25 avril, 2006.

- « Filtrage sémantique de textes et cadres de discours », *Journée Cadres et TAL*, LIPN, Villetaneuse, le 21 mai 2007.
- « Analyse de discours évaluatifs », Journée d'études *Observatoire des pratiques médiatiques émergentes. In Vivo*. Montpellier, MSH, Praxiling, 8 juin 2011.

3 – ENCADREMENT DOCTORAL ET SCIENTIFIQUE

DIRECTION DE THESE DE DOCTORAT

Co-direction avec Claude Montacié (STIH, Paris 4) :

Mathias Lambert, « Modélisation et annotation sémantique d'opinions et de jugements d'évaluation dans les textes », Université Paris 4, thèse en cours.

Thèse effectuée et financée dans le cadre du projet ANR OntOpiTex.

DIRECTION DE MÉMOIRES ET DE PROJETS D'INTEGRATION (MASTER)

DEA MIASH (Mathématiques et Informatique Appliquées aux Sciences de l'Homme)

Magali Cavanié, « Les relations statiques et leurs utilisations au sein du projet RAP », mémoire DEA MIASH (Mathématiques et Informatique Appliquées aux Sciences de l'Homme), Université de Paris-Sorbonne, septembre 1999.

Asma Bouhafs et Sana-Leïla Chaar, « Intégration, consultation et manipulation des données causales dans la base BDContext. Conception et réalisation des interfaces en langage Java », mémoire DEA MIASH, Université de Paris-Sorbonne, septembre 1999.

(Co-direction avec Jean-Pierre Desclés), Aude Vinzerich, « Etude sur l'assertion : approche énonciative, logique et textuelle », mémoire DEA MIASH, 2003.

Antoine Blais, « Evaluation de la tâche Résumé dans la plate-forme informatique ContextO. Proposition d'étiquetage pour la classe Annonce thématique », Mémoire DEA MIASH, 2004.

DESS ILSI

(Co-direction avec Jean-Luc Minel) Céline Leredde, Lisa Oustinova, « L'opération de reformulation. Analyse linguistique en vue d'une implémentation dans ContextO », DESS ILSI, 2003.

Katy Fraichard, Alexandre Janvier, Julie Livolsi, « La reformulation dans le discours », Mémoire DESS ILSI, 2004.

Inna Merkoulova, Majid Salhi, « Auto-représentation de l'énonciation. Analyse linguistique et codage informatiques des commentaires sur le discours en train de se faire », Mémoire DESS ILSI, 2005.

Master 1 « Information et Communication », Ingénierie de la langue pour la gestion intelligente de l'information (ILGII), Université de Paris-Sorbonne :

(Co-direction avec Claude Montacié), Anne-Laure Delettre, « Relation entre verbes de parole et discours rapporté : Le cas des verbes de cris d'animaux et d'élocution », Master 1 « Information et Communication », (ILGII), 2006.

Myriam Maloron, « Collecte et description des adverbiaux énonciatifs introduits par la préposition *en* », Master 1 « Information et Communication » (ILGII), 2007.

- (Co-direction avec Philippe Laublet), Charles-Henri Arnoux, « Ontologies pour l'étude des réputations des entreprises », Master 2 « Information et Communication » (ILGII), 2008.
- (Co-direction avec Philippe Laublet), Virginie Mouilleron, « Constitution de ressources linguistiques pour la fouille d'opinions dans des textes », Master 1 « Information et Communication » (ILGII), 2008.
- (Co-direction avec Claude Montacié), Thuy Linh Ngo, « Emotions et mode d'élocution – le cas des verbes de communication », Master 1 « Information et Communication » (ILGII), 2008
- Emmanuel Quintin, « Circulation, littéralité et fidélité de l'énonciation : de la reprise au plagiat dans le discours politique », Master 1 « Information et Communication » (ILGII), 2009.
- Nadia Zerguini, « Etude des groupes adjectivaux extra-prédicatifs à valeur axiologique », Master 1 « Information et Communication » (ILGII), 2009.
- (Co-direction avec Philippe Laublet), Luce Lefevre, « Ontologies pour l'étude des réputations des entreprises », Master 1 « Information et Communication » (ILGII), 2009.
- (Co-direction avec Philippe Laublet), Roald Muller, « Le langage de l'évaluation. « Appraisal » en français », Master 1 « Information et Communication » (ILGII), 2009.
- Adriana Amaya « Une analyse automatique des discours évaluatifs. Le cas des adverbes d'attitude orientés vers le sujet », Master 1 « Information et Communication » (ILGII), 2010.
- (Co-direction avec Claude Montacié), Catherine Mathis, « Relations sémantiques des noms d'émotion : étude des cooccurrences », Master 1 « Information et Communication » (ILGII), 2010.
- Kamel Akli, « Vers une analyse automatique des discours évaluatifs. Le cas des constructions détachées avec 'enclosures' », Master 1 « Information et Communication » (ILGII), 2010-2011.
- Clémentine Riquier, « Analyse et traitement des opinions, des sentiments et des jugements d'évaluation : état de l'art, perspectives et propositions », Master 1 « Information et Communication » (ILGII), 2011.
- Manon Vincent, « Jugements d'évaluation et réputation des entreprises », Master 1 « Information et Communication » (ILGII), 2011.
- Gael Guibon, « Analyse d'énoncés diffamatoires et éléments pour leur détection automatique », Master 1 « Information et Communication » (ILGII), 2011.

Master 2 PRO « Information et Communication », Ingénierie de la langue pour la gestion intelligente de l'information (ILGII), Université de Paris-Sorbonne :

Tutorat de stage :

- Emmanuel Quintin, « Analyse d'opinions. Constitution de ressources et expérimentations », Rapport de stage, Société Noopsis, Caen, Master 2 « Information et Communication » (ILGII), octobre 2010.

Projets d'intégration :

< Le projet d'intégration permet de réaliser en binôme ou en trinôme un projet lourd (1 à 2 hommes x mois) à forte composante technologique permettant d'intégrer et d'appliquer les connaissances acquises dans les différents modules du MASTER 2. Le travail donne lieu à la rédaction d'un mémoire et à une soutenance. >

- (Co-direction avec Philippe Laublet), Géraldine Kiyoko Clad, Nicolas Jeunot, « Etude linguistique des relations discursives dans les conversations en ligne en vue de la réalisation d'un outil d'accès aux forums de discussion », (ILGII), 2006.
- (Co-direction avec Philippe Laublet), Claudia Amaya, Benjamin Perbal, « Ontologies linguistiques, annotations pour le Web sémantique et traitements associés », (ILGII), 2007.

- (Co-direction avec Philippe Laublet) Myriam Maloron, Somara Seng, Mohammed-Akli Allouche, « Relations de cohésion dans les textes écrits et ontologies linguistiques pour l'annotation d'articles scientifiques », (ILGII), 2008.
- (Co-direction avec Philippe Laublet), Malika Ould Slimane, Djilali Zoulim, Thuy Linh Ngo, « Jugements d'évaluation dans des articles économiques : analyse de corpus, extraction automatique et caractérisation des cibles », (ILGII), 2009.
- (Co-direction avec Max Silberztein), Charlène Petit, Raudah Aminudin, « Construction et analyse sémantique de lexiques évaluatifs pour l'étude de la réputation. Le cas des adverbes orientés sujet. », (ILGII), 2010.
- (Co-direction avec Claude Montacié), Laurie Acensio, « Groupes nominaux à fonctions catégorisantes et qualifiantes », (ILGII), février 2011.

PARTICIPATION AUX JURYS DE THÈSE

- Wonsil PARK, « Sémantique et représentations formelles de verbes qui expriment les relations causales : *augmenter, conduire, créer, déclencher, diminuer, entraîner, entretenir, pousser, provoquer* », Thèse de doctorat, Laboratoires Image et traitement de l'information (ENST Bretagne) et LaLICC (Université de Paris-Sorbonne), janvier 2006.
- Atefeh FARZINDAR, « Résumé automatique de textes juridiques » thèse de doctorat effectuée aux laboratoires RALI (Université de Montréal) et LaLICC (Université de Paris-Sorbonne), mars 2005.

4 – ACTIVITES D'ENSEIGNEMENT

J'assure mon service complet de maître de conférence depuis la prise de fonction en septembre 1998.

A) ACTIVITÉS D'ENSEIGNEMENT à Paris IV actuelles

- « Composantes textuelles pour le TAL », CM, Master 2 « Information et Communication » Informatique et Ingénierie de la Langue pour la gestion de l'Information (ILGII).
- « Approches de l'énonciation et linguistiques de corpus » CM, TD, TP, Master 1 « Information et Communication » (IILGI et ILGII).
- « Sémantique cognitive » Cours et TD, Master 1 « Information et Communication » (IILGI et ILGII).
- « Bases de données relationnelles », CM, TD, TP, Licence3, Licence « Lettres classiques et modernes, Sciences du langage », Parcours *Langue française et techniques informatiques LFTI*.
- « Programmation impérative », CM, TD, TP, Licence2, Licence « Lettres classiques et modernes, Sciences du langage », Parcours *Langue française et techniques informatiques LFTI*.

Direction de mémoires et de projets d'intégration en Master 1 et 2.

B) ACTIVITÉS D'ENSEIGNEMENT à Paris IV avant septembre 2005

- « Analyse textuelle et traitement informatique », CM, DEA MIASH (Mathématiques et Informatique Appliquées aux Sciences de l'Homme)
- « Exploration contextuelle et filtrage automatique de textes » Atelier informatique, DEA MIASH
- « Linguistique et TAL », mise à niveau, CM, DESS ILSI (Ingénierie de la Langue et société de l'Information)
- « Sémantique des langues naturelles et traitement informatique », CM, Maîtrise, Option LIL (Logique, Informatique et Langage)
- « Techniques informatiques », CM, TD, TP, Deug bi-disciplinaire : Langue française et Techniques informatiques
- « Initiation à l'informatique » EA250, CM/TD/TP, DEUG 2, option.

Direction de mémoires et de projets annuels en DEA et DESS.

C) ACTIVITÉS D'ENSEIGNEMENT hors Paris IV

ENST Bretagne : Séminaire « Approche linguistique de la causalité », (Maryvonne Abraham), Brest, 24 février 2004.

A l'université Clément d'Ohrid de Sofia, en Bulgarie (Mission EGIDE)

Mai 2003 et mai 2004 :

Enseignement dans le Master TAL (CM, 12 h) : « Décrire et modéliser l'organisation discursive des textes en vue de son exploitation informatique ».

A l'Université Hankuk des Etudes Etrangères, Séoul, en Corée du Sud, (**Mission EGIDE**)

1er juillet - 15 juillet 2005 et 29 juin - 15 juillet 2006.

« Composantes et structures textuelles pour le traitement automatique des langues » (CM, 18h), dans le cadre de deux Ecoles d'été à l'attention des étudiants du département de français et du département de pédagogie du français : « Pour la rencontre de la linguistique avec l'informatique », subventionnée par l'Ambassade de France en Corée.

A l'Université de Paris III

Séminaire de Doctorat de sciences du Langage 2006-2007 « L'initiale de la phrase et de l'énoncé, à l'écrit et à l'oral » dirigé par Mary-Annick Morel & Michel Charolles.

« Les expressions introductrices de cadres organisationnels », Université de Paris 3, février 2007.

5 – RESPONSABILITES SCIENTIFIQUES ET ADMINISTRATIVES A PARIS IV

COMMISSION DE SPÉCIALISTES

Membre de la Commission des Spécialistes des sections 7, 26 et 27 à l'Université de Paris IV, depuis 2001.

Vice-Président de la Commission des Spécialistes des sections 26 et 27 à l'Université de Paris IV, depuis 2004.

RESPONSABILITES AU SEIN DE L'UFR

Membre du Conseil de l'UFR Institut des Sciences Humaines Appliquées (ISHA), depuis 2000.

Directeur adjoint de l'UFR Institut des Sciences Humaines Appliquées (ISHA), de septembre 2006 à juin 2011.

COORDINATION D'ENSEIGNEMENTS

Coordination des enseignements d'Initiation à l'informatique, proposé à l'ensemble des étudiants de Paris IV, *SHEA250*, entre septembre 1998 et septembre 2004.

Gestion des ateliers informatiques ouverts en Libre service à Paris IV, entre septembre 1998 et septembre 2004.

Contribution à la création, puis **coordination** des DEUG expérimentaux *Langue française et Techniques informatiques (LFTI)* et *Langues et Techniques informatiques (spécialité Allemand/Russe/Italien/Anglais / Techniques informatiques)*, entre septembre 2000 et septembre 2004.

Coordination des enseignements de l'option LIL (Langues, Informatique et Logique), cours de Logique, de Linguistique et de Programmation, niveau Licence et Maîtrise, entre septembre 2004 et septembre 2005.

Direction des études en Master 1 « Information et Communication » Spécialité Recherche *Informatique et Ingénierie de la Langue pour la Gestion de l'information* et Spécialité Professionnelle *Ingénierie de la langue pour la gestion intelligente de l'information*, à partir de septembre 2005.

6- ACTIVITES DE COLLABORATION INTERNATIONALES

- **Avec la Pologne** (Académie Polonaise des Sciences PAN, Varsovie)

Participation au programme de coopération franco-polonaise (accord institutionnalisé et financé CNRS – PAN) sur la **sémantique cognitive** (quantification, causalité, catégories de l'aspect et du temps). Membres fondateurs : Stanislas Karolak, Violetta Koseska, Antoni Mazurkiewicz, Wieslaw Banys, Hélène Włodarczyk, André Włodarczyk, Jean-Pierre Desclés, Zlatka Guentchéva ; 1997-1999

- **Avec la Bulgarie** (Université Clément d'Ohrid, Sofia)

Echanges sur **l'enseignement de la linguistique et du TAL**, dans le cadre des missions EGIDE, avec Albena Vasilieva et Petar Lazarov, 2002 et 2003.

- **Avec la Corée du Sud** (Université de Hankuk des Etudes Etrangères, Séoul)

Participation au projet de recherche franco-coréen sur **l'identification automatique de citations en français et en coréen**. Collaboration avec les chercheurs de l'équipe du professeur Shiyeon Pak ; depuis juillet 2005.

- **Avec la Belgique** (l'École Supérieure de Traducteurs et Interprètes de Gand)

Initiative puis mise en place d'une coopération scientifique, officialisée par une convention entre l'Université de Paris-Sorbonne et l'École Supérieure de Traducteurs et Interprètes de Gand en Belgique, avec l'équipe Autodoc dirigée par le professeur Rita Godijns. Initiée en novembre 2004, cette collaboration a été officialisée en septembre 2005.

Thème de recherche : **Identification et extraction automatique des énoncés causaux** dans des textes

7 – PRINCIPAUX PROJETS DE RECHERCHE

Projet RAP (Recherche, Analyse, Propose) ; 1997-1999

Financé par le Ministère de l'enseignement supérieur et de la recherche, dans le cadre des actions sur la recherche et le filtrage d'information sur les réseaux comme le Web.

Participants : LALIC, société Pacte Novation, l'Urfist de Paris

But : (i) proposer une méthode originale pour aider des utilisateurs dans des situations de collecte d'information sur le Web, (ii) construire un outil informatique qui permet d'utiliser la méthode proposée.

Contribution : réalisation d'un module de reformulation des requêtes selon le point de vue de la causalité.

Projet REGAL : Action Cognitive ; 2000 – 2002

Laboratoires : LIMSI (UPR), LATTICE (UMR 9094), CEA (LIC2M) et LaLICC (UMR 8139)

But : construire des logiciels permettant de structurer automatiquement un texte et de le visualiser.

Contribution : identification des cadres organisationnels.

Projet E-VEIL : Ministère de l'Industrie ; 2002 – 2004

Laboratoires : LaLICC, Paris 5, Société Xistos

But : projet de recherche d'information et de catégorisation des sites Web ; recherche de séquences textuelles pouvant exprimer un savoir sur le monde

Contribution : identification des relations partie – tout.

Projet OLETT : Action "Société de l'Information" du CNRS ; 2002 – 2005

Laboratoires : LATTICE et LaLICC

But : (i) Construction des représentations spatio-temporelles à partir de textes, (ii) Repérage des phrases exprimant des rencontres (qui a rencontré qui ? quand ? et où ?).

Participation aux travaux sur la conceptualisation de la relation de « rencontre » et ses moyens d'expression dans les textes.

Projet ANR SFA (blanc SHS); 2006 – 2008

« SFA Spatial Framing Adverbials : linguistic and psycholinguistic approaches »

Laboratoire porteur : LATTICE

Etude des relations d'indexation (observation du fonctionnement des adverbiaux dits cadratifs, essentiellement des adverbiaux spatiaux, concrets et abstraits) au moyen de deux types d'investigations : analyses linguistiques sur corpus et expérimentations psycholinguistiques.

Contribution : codage et analyse des adverbiaux en *-en*.

Projet ANR ONTOPITEX (ContInt) 2009 – 2012

« OntOpiTex : Modèles linguistiques et ontologies. Extraction informatique et caractérisation d'opinions et de jugements d'évaluation dans les textes ».

Laboratoires: LaLIC puis STIH (porteur du projet), GREYC, CRISCO, Société TKM (Grenoble), Société Noopsis (Caen)

Responsabilité : tâche 2 « Constitution des modèles linguistiques et de l'ontologie », voir la section 8.

Projet IN VIVO ; 2011 – 2013

Programme MSH Montpellier ; Axe : Société et nouvelles technologies de communication

But : Création d'un observatoire des pratiques médiatiques émergentes

Participants : Praxiling UMR 5267 (porteur du projet), CLEMI Montpellier, Université de Lausanne, Université de Linkoping (Suède), STIH EA 4509 (Paris 4), Société Noopsis (Caen)

Responsabilité : l'axe « Interface, données et métadonnées ».

8 - VALORISATION DE LA RECHERCHE

EXPERTISES NATIONNALES ET INTERNATIONNALES

Membre du comité de lecture : *Workshop « Ontologies et textes » TIA 2007*, Sophia-Antipolis, Octobre 2007.

Membre du comité scientifique : *Journées scientifiques du CRISCO*, « Ce que le texte fait à la phrase », 3-4 décembre 2009, Caen.

Membre du comité scientifique : *Colloque International « Approches Linguistiques et psycholinguistiques de la structuration des textes » (LPTS 2009)*, Lattice, Ecole Normale Supérieure, Paris, France

Membre du comité du programme : *Atelier « Acquisition et modélisation de relations sémantiques » TIA 2009*, Toulouse, 18 -20 novembre 2009.

Membre du comité de lecture : *CRL « Supports et prédicats non verbaux dans les langues du monde »*, Amr Helmy Ibrahim (éd.), Paris, CRL, 2009.

Expert ECOS : Evaluation - orientation de la Coopération Scientifique, Programme de coopération et de partenariat universitaire technique et scientifique, France - Amérique Latine, 2009.

Expert ECOS : Evaluation - orientation de la Coopération Scientifique, Programme de coopération et de partenariat universitaire technique et scientifique, France - Amérique Latine, 2010.

Membre du comité de lecture : *revue TAL n°51 :1 Varia*, 2009-2010.

Membre du comité de rédaction : numéro spécial de la *revue Document Numérique*, «Le texte : objet d'analyse et vecteur de connaissances», 2010.

Membre du comité de lecture spécifique : numéro spécial de la *revue TAL 51 :3*, « Opinions, sentiments et jugements d'évaluation », 2010- 2011.

Membre du comité scientifique : *Colloque International « Approches Linguistiques et psycholinguistiques de la structuration des textes » (LPTS 2011)*, Belgique.

Membre du comité des experts ANR pour le Programme "Contenus et Interactions" 2010.

RESPONSABILITES EDITORIALES

Revue DISCOURS : Revue de linguistique, psycholinguistique et informatique, en ligne en texte intégral sur <http://discours.revues.org/>

- Membre fondateur (septembre 2007),
- Rédacteur en chef (2007-2009), Editeur : Université de Paris-Sorbonne
- Membre du comité éditorial (2007-2009).

N° ISSN : 1963-1723

Portail : [Revues.org](http://www.revues.org) sur la plateforme des « Cahiers » (<http://www.revues.org>).

Domaine : La revue Discours publie des articles scientifiques ayant trait aux phénomènes de cohésion textuelle, assortis de corpus annotés, de ressources lexicales ou de grilles de codage.

La revue adhère à la fédération des revues en sciences humaines et sociales [Revues.org](http://www.revues.org), membre du Centre pour l'édition électronique ouverte CLEO (<http://cleo.cnrs.fr/> - CNRS, EHESS, l'Université d'Avignon et l'Université de Provence). Cette structure participe au

réseau mis en œuvre par le Très grand équipement ADONIS (TGE ADONIS, <http://www.tge-adonis.fr/>).

DIRECTION DES PUBLICATIONS

Revue TAL <http://www.atala.org/-Revue-TAL->

Co-Rédacteur invité : numéro spécial de la *revue TAL* 51 :3, « Opinions, sentiments et jugements d'évaluation », 2011.

RESPONSABILITES AU SEIN DE PROJETS

Projet ANR OntOpiTex

Programme : Contenu et Interaction Edition 2009,

Intitulé : « OntOpiTex : Modèles linguistiques et ontologies. Extraction informatique et caractérisation d'opinions et de jugements d'évaluation dans les textes ».

Responsable de la Tâche 2 : « Constitution des modèles linguistiques et de l'ontologie », 30 mois.

Type de recherche : Recherche industrielle

Durée : Octobre 2009 – 36 mois

Résumé :

L'analyse informatisée des opinions, attitudes ou jugements répond à des enjeux économiques et sociaux importants. Pourtant, les travaux actuels se limitent souvent à la notion de polarité (positive, négative, neutre), quand les phénomènes visés sont en réalité bien plus riches, diversifiés et complexes. L'objectif principal d'OntOpiTex est de chercher à identifier et agréger des segments textuels porteurs d'opinions, et surtout de les classer par type (affect, jugement, appréciation, ...) et de les caractériser par des critères plus fins : polarité, intensité, prototypicalité, cible(s), source(s), engagement de l'auteur, etc. Des outils interactifs seront proposés pour visualiser et naviguer dans ces annotations à l'échelle d'une collection de documents. Sur le plan scientifique (linguistique, TAL et IC), des questions théoriques importantes émergent concernant la nature même de l'objet « opinion », et le choix des méthodes et outils les plus adaptés à leur traitement.

Nous aborderons ces questions sur trois fronts, en adoptant une approche pluridisciplinaire novatrice et originale dans ce champ d'étude : (1) modélisation linguistique et ontologique afin d'éclairer conceptuellement les notions en jeu à l'aide de théories linguistiques comme *Appraisal* (Martin et White 2005) ; (2) développement et mise en oeuvre d'un modèle opératoire pour l'analyse automatique dans les textes ; (3) évaluation et étude de cas d'utilisation, à la fois sur un corpus de travail et dans un cadre applicatif réel. La réalisation de ces objectifs ambitieux sera rendue possible par une étroite collaboration entre laboratoires spécialistes de ces questions, et entreprises dont les métiers allient la veille à l'informatique linguistique.

Partenaires :

- Laboratoire Langages, Logique, Informatique, Cognition (STIH, Université Paris-Sorbonne), **porteur du projet**
- Groupe de REcherche en Informatique, Image, Automatique et Instrumentation de Caen (GREYC, Université de Caen),
- Centre de Recherches Interlangues sur la Signification et le Contenu (CRISCO, Université de Caen),
- Société Noopsis (Caen),
- Société TecKnowMetrix (Grenoble).

9 – RESSOURCES LINGUISTIQUES CONSTITUEES

Présentation quantitative approchée.

RELATION PARTIE - TOUT

E-veil (2002)	support	marqueurs	classes sémantiques	schémas discursifs
	base Access		2350	20

CAUSALITÉ

Context0 / Excom (2006)	support	règles / valeurs	schémas expressions	qualificatifs	verbes causalité (formes conj)	verbes prise en charge
	base Access		64	667	893	6129

RAP	support	expressions causalité	expressions partie-tout
	base Access		175

SERIES LINEAIRES

Cadres organisationnels dans Context0	support	marqueurs	classes	valeurs	règles
	base Access		381	55	2

Base définitive (lexiques catégorisés)

support	balises (MIL)	balises autres	classifieurs	quantifieurs	retroévaluation	connecteurs
base Access	255	13	275	58	20	14

DISCOURS RAPÔRTES

support	adverbes PEC	verbes parole tot	verbes parole DR	noms parole	qualif parole	expr dire	reformulation (tot)	reformulation (DR)
base Access	321	517	356	549	221	78	330	161

CONSTITUANTS DETACHES A CARACTÈRE AXIOLOGIQUE (en cours de constitution)

phase 1 : constitution manuelle de ressources

support	lexiques	expressions	énoncés	patrons
fichiers texte et excel	409	283	300	20

phase 2 : exploration sur corpus

support	patrons	P1	P6	P11	P16	P17	P18
fichiers texte et excel	attestations	2887	100	1662	590/419	398/298	25929
	lexiques	<i>à construire</i>					