Texte et contexte : pour une approche fonctionnelle et empirique

Shirley Carter-Thomas

Maître de conférences à Télécom et Management SudParis (TMSP)

Mémoire de synthèse
en vue de l'obtention de l'Habilitation à diriger des recherches

sous la direction de Monsieur Michel Charolles, Professeur des universités

Université Sorbonne Nouvelle - Paris 3

2009

Remerciements

Je remercie en premier lieu Michel Charolles d'avoir accepté ce travail d'encadrement et de conseil et lui exprime ma reconnaissance ainsi qu'à l'ensemble des rapporteurs et membres du jury : David Banks, Claude Delmas, Catherine Fuchs, Lita Lundquist et Marie-Paule Péry-Woodley.

Mes activités de recherche ne se sont pas déroulées dans l'isolement et j'éprouve une profonde gratitude envers tous ceux qui m'ont accompagnée. Je souhaite remercier tous les membres de l'équipe LaTTiCe, avec lesquels j'ai eu le plaisir de travailler au cours de ces quatre dernières années, pour l'intérêt porté à mes travaux et leur accueil toujours chaleureux. Ma gratitude va aussi aux amis du GERAS auprès desquels je trouve une source de soutien depuis de nombreuses années. Au sein de cette société savante, je tiens à remercier tout particulièrement Elizabeth Rowley-Jolivet pour sa générosité et sa rigueur scientifique. C'est grâce à Elizabeth que j'ai pu développer certains aspects de ma recherche et notre collaboration a été très précieuse.

Je remercie Télécom et Management SudParis (TMSP) de m'avoir permis de poursuivre mes recherches dans les meilleures conditions. Au sein de TMSP je tiens à remercier notamment la direction de la Recherche représentée par Stanislav Budkowski ainsi que François Simon pour leur soutien et la confiance qu'ils m'ont accordée et Fabienne Canal, la responsable de mon département, pour sa compréhension.

Je remercie enfin tous ceux qui m'ont aidé à relire et à améliorer ce manuscrit : Claire, Florence, Michèle, Monique et Sophie.

Table des matières

Présentation 6

Parcours 8

Partie A : Le cadre théorique

1 Aborder le texte 17

1.1 De la phrase vers le texte 18
1.1.1 Procédés de cohésion formelle 20
1.1.2 D'autres formes de connexité 22
1.1.3 Cohésion et cohérence : pour une approche intégrée 24

1.2 Du texte au discours 27
1.2.1 Distinction texte/discours 27
1.2.2 Deux notions complémentaires 29

1.3 Dangers du « tout pragmatique » 30
1.3.1 Prises sur le texte 31
1.3.2 Articulation phrase/texte/discours 32

2. Analyser la structure informationnelle 34

2.1 L'approche de l'École de Prague 37
2.1.1 Délimitation d'un niveau informationnel 37
2.1.2 Noyau thème-rhème 39
2.1.3 Dynamisme communicatif 41
2.1.4 Du thème de la phrase au texte : progression thématique 43

2.2 L'approche de Halliday 48
2.2.1 Séparation de thème et « donné » 48
2.2.2 Critère de position 51
2.2.3 Thèmes marqués 52
2.2.4 Thèmes composites 54

2.3 L'approche de Lambrecht 56
2.3.1 Indentifiabilité et accessibilité des référents 56
2.3.2 Un topique pragmatique 59
2.3.3 Taxinomie de structures focales 60

2.4 Pour une approche opérationnelle 63

Partie B : De la théorie à la pratique

3. Marqueurs de la structure informationnelle 67

3.1 Thème et sujet grammatical 68

3.2 Thème et position initiale 70

3.2.1 Détachement à gauche 72
3.2.2 Adverbiaux cadratifs 73

3.3 L'ordre des mots 78
3.3.1 Ordre syntaxique *vs* ordre pragmatique 78
3.3.2 Le cas d'une langue polysynthétique : l'inuit 82

4. Une syntaxe au service du texte 88

4.1 Les structures dites spécialisées 90
4.1.1 Passif(s) 90
4.1.2 Énoncés existentiels 92
4.1.3 L'extraposition 93
4.1.4 Structures détachées 94
4.1.5 Structures hypotaxiques 95
4.1.6 Constructions clivées 96

4.2 Le cas des phrases clivées : comparaison anglais/français 97
4.2.1 Corpus et méthodologie 98
4.2.2 Traduction des phrases clivées en anglais 104

Partie C : Pour une approche empirique

5. Travailler sur corpus 111

5.1. Le corpus comme objet d'étude 112
5.1.1 Choix et constitution 112
5.1.2 Représentativité 115
5.1.3 Corpus *vs* base de données 117

5.2 Linguistique de corpus comme méthode 118
5.2.1 Principes d'exhaustivité et de quantification 118
5.2.2 Analyse textuelle et linguistique de corpus 120

5.3. Corpus et textes spécialisés 123

6. Maîtriser les variations : l'approche de genre 129

6.1 Cohérence et généricité 130

6.2 Démarches classificatrices 133
6.2.1 Une notion biface 135
6.2.2 Distinction type/genre 136

6.3 Genre et discours spécialisés 138
6.3.1 Genres scientifiques en contraste 141
6.3.2 Genres et pédagogie 143

Partie D : Analyser les textes spécialisés

7 Applications de la structure informationnelle 148

7.1 La progression thématique appliquée..... 149
7.1.1 Applications en didactique 149
7.1.2 Applications aux textes spécialisés..... 155

7.2 Structures spécialisées et genre..... 159
7.2.1 Contraintes de genre : communication de congrès vs article de recherche 160
7.2.2 Corpus et méthodologie..... 162
7.2.3 Distribution de structures spécialisées dans les deux genres..... 164
7.2.4 Comparaison entre chercheurs anglophones et non anglophones..... 173

8 Organisation rhétorique et textes spécialisés 178

8.1 Une analyse de l'introduction de la communication de congrès..... 180
8.1.1 Corpus et méthodologie..... 181
8.1.2 Une proposition pour un modèle de « *moves* » 185
8.1.3 Impact de la culture disciplinaire *vs* impact de la culture linguistique 189

8.2 Syntaxe et relations interpersonnelles dans la communication de congrès 192
8.2.1 Relations interpersonnelles et choix de pronom 193
8.2.2 Relations interpersonnelles et syntaxe 197

9 Une syntaxe au service des genres : Le cas des phrases conditionnelles 204

9.1 Emploi des phrases conditionnelles dans les articles scientifiques en médecine 206
9.1.1 Corpus et méthodologie..... 207
9.1.2 Fréquences, formes et fonctions..... 209
9.1.3 Comparaison français/anglais..... 214

9.2 Emploi des phrases conditionnelles dans les communications de congrès..... 218
9.2.1 Fréquences, fonctions et formes (contrastes avec l'article de recherche)..... 219
9.2.2 Comparaisons : conférenciers anglophones/conférenciers non anglophones.... 222

9.3 Emploi des phrases conditionnelles dans les éditoriaux spécialisés 223
9.3.1 Fréquences, fonctions et formes (contrastes avec l'article de recherche)..... 224
9.3.2 L'étude de quatre variants..... 227

9.4 De la théorie à la pratique..... 230

10. Texte et contexte 234

10.1 Options méthodologiques..... 234

10.2 Aspects applicatifs..... 236

Bibliographie..... 239
1. Références aux travaux personnels 239
2. Références bibliographiques 242

Présentation

Ce mémoire de synthèse porte essentiellement sur les recherches que j'ai effectuées depuis la thèse de linguistique générale et appliquée que j'ai soutenue en 1998 à l'université Paris V René Descartes. Il comprend différents travaux en linguistique générale comme en anglais de spécialité. Toutefois, étant donné que ces axes de recherche sont étroitement liés, soit par leur mise en œuvre, soit par le corpus par le biais duquel ils sont abordés, je n'ai pas cherché à séparer le point de vue du *linguiste* de celui de l'*angliciste*. Tout itinéraire implique inévitablement un certain va-et-vient : une nécessaire ouverture vers des idées nouvelles mais également des retours en arrière. De ce fait, l'ordre de présentation dans ce mémoire ne correspond que partiellement à l'ordre chronologique de la réalisation de ces recherches entrecroisées.

La première partie, **Cadre théorique,** situe le cadre général de mes recherches. Dans le chapitre 1**,** je fais état de différentes théories et approches sur la structuration des textes en linguistique textuelle qui ont influencé mon parcours. Dans le chapitre 2**,** je me penche tout particulièrement sur trois théories de la structure informationnelle afin de situer ce que je considère être les principaux enjeux de la problématique informationnelle.

Dans la deuxième partie, **De la théorie à la pratique**, les questionnements sur la structuration des textes et sur les traces laissées par cette structuration se poursuivent. Je m'attache à explorer les réalisations linguistiques de la structuration informationnelle. Dans le chapitre 3, j'examine les rapports privilégiés que la notion de thème/topique peut entretenir avec certains positionnements phrastiques ou avec certaines fonctions grammaticales. Le chapitre 4 est consacré au rôle important joué par la syntaxe dans la signalisation de la structure informationnelle.

Une constante pour moi est la nécessité de travailler sur des textes situés, afin de mettre en relation les formes examinées avec une situation d'énonciation précise. Dans la troisième partie de cette synthèse, **Pour une approche empirique,** je m'attarde sur deux thèmes qui parcourent l'ensemble des travaux présentés : celui des études sur corpus au chapitre 5 et celui de la notion de genre au chapitre 6.

Dans la quatrième et dernière partie, **Analyser les textes spécialisés,** je me focalise plus particulièrement sur la structuration des textes spécialisés, pour la plupart des textes scientifiques (écrits et oraux) en anglais. Le chapitre 7 est consacré aux marqueurs de la structuration informationnelle dans certains textes scientifiques. Le chapitre 8 est consacré aux modèles d'organisation textuelle axés sur la rhétorique. Dans le chapitre 9, j'évalue le rôle structurant joué par les phrases conditionnelles dans ces textes spécialisés. Enfin, dans le chapitre 10, je termine par une réflexion sur quelques constantes qui sous-tendent l'ensemble des travaux présentés.

Au préalable, en guise de préambule, dans la section intitulée **Parcours**, je retrace brièvement mon développement intellectuel en tant que chercheur, à travers les différentes théories, milieux intellectuels, cadres institutionnels et rencontres qui m'ont influencée.

Parcours

D'origine anglaise, après un début de parcours universitaire (licence de français) en Grande-Bretagne, je suis venue en France préparer une maîtrise et un DEA de linguistique générale à l'Université Paris V René Descartes dans les années 1980. C'était l'époque de la grammaire fonctionnelle d'A. Martinet. Cependant, si ma formation initiale en linguistique était la linguistique fonctionnelle française, avec l'accent qui était mis sur la description rigoureuse des langues et le fonctionnement des petites unités à un niveau phrastique, je me suis vite aussi intéressée aux « 'textes » et aux problèmes posés par la mise en texte dans des situations concrètes.

La linguistique systémique de M. Halliday, la linguistique textuelle nordique (N. Enkvist; L. Lundquist) et en France les travaux sur les linguistiques textuelles de B. Combettes et de M. Charolles m'ont beaucoup influencée pendant toute cette première période. Dans mes tout premiers travaux, et notamment dans mon mémoire de DEA par exemple, j'ai examiné les rapports entre cohésion et cohérence dans les récits d'enfants.

Cet intérêt grandissant pour le texte et la linguistique textuelle a aussi été motivé en partie par ma situation professionnelle – d'enseignante d'anglais auprès d'étudiants francophones. À cet égard, j'ai été très marquée par certains travaux de M. Charolles sur les potentielles applications de la linguistique textuelle en milieu scolaire. Charolles (1978) décrit le malaise ressenti par les enseignants (en l'occurrence des enseignants de français langue maternelle en classe primaire) devant ce qu'ils perçoivent souvent comme l'inadéquation des corrections qu'ils apportent aux copies des élèves. Si les erreurs phrastiques sont généralement localisées et expliquées avec précision, en revanche, les malformations constatées à un niveau plus global ne reçoivent aucun commentaire précis. L'efficacité didactique des appréciations des enseignants semblerait très limitée quand il s'agit d'aborder des questions relatives à l'organisation globale du texte. J'ai trouvé des échos à ces préoccupations

exprimées par Charolles devant les copies de mes propres étudiants. Comment le professeur de langue peut-il analyser et évaluer de façon pertinente des travaux d'étudiants, acceptables sur le plan grammatical mais difficilement interprétables en tant qu'ensemble textuel ? Est-ce qu'il existe des outils analytiques capables de rendre compte de certains dysfonctionnements constatés au niveau organisationnel des textes ?

Ce sont de tels questionnements qui ont débouché sur ma thèse portant sur le rôle joué par la structuration informationnelle dans l'évaluation de la cohérence, soutenue en 1998, ainsi que sur certains travaux entrepris dans le prolongement. En prenant appui sur toute une série de devoirs écrits produits par mes étudiants dans un cours d'anglais, je me suis efforcée d'établir les bornes des pratiques correctives limitées à la seule surface textuelle. J'ai démontré que très souvent des textes dépourvus en surface d'erreurs perceptibles peuvent être considérés comme « défaillants » parce que leur structuration informationnelle ne répond pas aux attentes de leur récepteur. C'est seulement lors d'une analyse de leur structuration informationnelle que certaines causes de ce manque de clarté et de fluidité ont pu être déterminées avec plus de précision.

Dans les travaux ultérieurs à ma thèse, j'ai raffiné la méthodologie adoptée notamment en ce qui concerne les aspects cognitifs de la structure informationnelle. En m'appuyant sur les travaux de W. Chafe et surtout de K. Lambrecht (1994), j'ai cherché à cibler plus précisément la notion cruciale de connaissance partagée. J'ai ensuite réutilisé ces analyses pour d'autres types de textes (journalistiques, scientifiques), écrits et oraux [THOM 00, 01, 02, 04c]. Afin de mieux comprendre les différents phénomènes mis en jeu dans la structuration informationnelle, il me semble essentiel en effet de prendre en considération un éventail de situations discursives distinctes (intralinguales et interlinguales). J'ai entrepris toute une série d'études et de comparaisons afin de cerner de plus près l'impact précis de la situation d'énonciation (par exemple, le mode de transmission, les exigences au niveau de la rhétorique et

des différentes relations interpersonnelles) sur les stratégies informationnelles employées.

Pendant cette même époque post-thèse, j'ai également participé activement à la mise en place et à l'organisation des différentes activités scientifiques du laboratoire Osterlits (UMR 7108). En son sein, j'étais, en effet, coresponsable, depuis sa création en 1999, du domaine « Oralité, cognition, construction du sens », dont la première opération thématique traitait de la topicalisation. C'est à ce titre, par exemple, que j'ai organisé en Juin 2001, en collaboration avec M.M.J. Fernandez-Vest, directrice de l'équipe Osterlits, une journée d'étude dont le thème était « Structure informationnelle, typologie, genres du discours », manifestation à laquelle ont participé plusieurs spécialistes français et internationaux du domaine – dont M. Charolles, B. Combettes, K. Lambrecht et L. Mondada. Une partie de ces travaux, regroupés avec ceux d'une autre journée d'étude tenue antérieurement, a donné lieu à la publication d'un ouvrage collectif (Fernandez-Vest & Carter-Thomas 2004).

Mon orientation vers l'étude de textes dits « spécialisés » se confirme par la suite : articles de recherche scientifique ; communications de colloque, éditoriaux de revues scientifiques spécialisées. Plutôt que de fonder mes analyses sur les grands corpus généralistes (cf. chapitre 5 *infra)*, de tels ensembles textuels ont l'avantage de permettre une certaine stabilisation des conditions pragmatiques et ainsi de mieux évaluer les variations textuelles. L'un de mes objectifs principaux étant de pouvoir cerner de plus près les contraintes d'ordre pragmatique qui pèsent sur les stratégies informationnelles adoptées, le discours scientifique fournit un terrain très propice pour ces analyses.

Une fois encore mes activités de recherche et mes activités d'enseignement se sont trouvées étroitement imbriquées. Mon intérêt pour les textes scientifiques est aussi motivé par ma situation professionnelle. Depuis 1991, j'occupe un poste d'enseignant-chercheur (et depuis 1999 un poste de maître de conférences) dans le département Langues et sciences humaines de Télécom

et Management SudParis (anciennement Institut national des télécommunications) – un établissement d'enseignement supérieur qui fait partie de l'Institut Télécom. Télécom et Management SudParis regroupe sur son site à Evry une grande école publique de management (Télécom Ecole de management) et une grande école d'ingénieurs (Télécom SudParis). J'enseigne l'anglais scientifique (écrit, mais aussi oral) en formation initiale et en formation doctorale aux étudiants inscrits dans ces deux établissements. Je fais également partie de la cellule « relecture » de mon département, et à ce titre je relis fréquemment les articles scientifiques pour mes collègues francophones. Je côtoie ainsi quotidiennement des étudiants et chercheurs dans les sciences exactes (physique, mathématiques, télécommunications) aussi bien qu'en gestion et en sciences humaines. J'essaie d'exploiter dans la mesure du possible le matériel (les corpus) à ma disposition et aussi de réutiliser certains résultats de mes travaux dans mes cours.

Un parcours est jalonné de rencontres. Je ne présenterai certainement pas le même bilan aujourd'hui si j'avais poursuivi une recherche isolée. Une première rencontre décisive pour moi fut celle du réseau des anglicistes de l'association GERAS (Groupe d'étude et de recherche en anglais de spécialité[1]). Depuis 1994, année durant laquelle j'ai participé pour la première fois à leur colloque annuel à Rennes, je trouve une source constante de soutien et d'inspiration auprès des membres de ce réseau, en ce qui concerne l'enseignement comme la recherche. C'est à la demande de l'ancien président du GERAS, M. Petit, que j'ai rejoint l'équipe encadrante du DEA et ensuite du Master recherche d'anglais de spécialité (dont l'équipe principale est l'EA 2025 de Bordeaux 2). Depuis j'ai encadré plusieurs mémoires de troisième cycle en linguistique anglaise et en anglais de spécialité. Je suis aussi responsable au sein du GERAS du groupe thématique « Sciences et academia » qui comprend une vingtaine de membres.

C'est par le biais du GERAS que j'ai également rencontré E. Rowley-Jolivet en 1999, avec qui j'ai depuis cosigné de nombreux articles. E. Jolivet est

[1] http://www.geras.fr/welcome/index.php.

spécialiste de l'anglais scientifique oral et a constitué un corpus important de communications scientifiques, enregistrées en vidéo dans des congrès internationaux. Nous avons entrepris ensemble toute une série de comparaisons entre l'anglais scientifique écrit et l'anglais scientifique oral, ainsi qu'un volet de recherche sur la rhétorique scientifique. Notre collaboration très fructueuse a été récompensée par le comité scientifique de la revue international *English for Specific Purposes* du prix Horowitz en 2005.

D'autres rencontres très formatrices ont eu lieu pendant le temps que j'ai passé au sein du laboratoire Osterlits. En effet, les échanges à la fois larges et riches que j'ai pu avoir avec les finno-ougristes de l'équipe (dont M.M.J. Fernandez-Vest et J.-L. Léonard) autour des questions typologiques relatives à la structure informationnelle m'ont permis de faire avancer ma réflexion. J'ai également eu la chance pendant la période de délégation d'une année au CNRS (20002-2003) de travailler avec N. Tersis, spécialiste des langues de l'Arctique. A partir d'un corpus de contes inuits, nous avons cherché à décrire les rapports entre la structure informationnelle et l'ordre des mots dans cette langue très polysynthétique. C'était un projet très enrichissant qui a donné lieu à deux publications.

Suite à la fermeture de l'équipe Osterlits, je suis ravie d'être accueillie depuis 2006 par l'équipe LaTTiCe (UMR 8094). Je retrouve ici les thèmes de recherches centraux à mes propres problématiques, à savoir la structuration des textes et l'étude des traces linguistiques de la cohérence textuelle. Je retrouve également au sein de l'équipe LaTTiCe, M. Charolles, l'un des linguistes qui m'a le plus influencée au début de mon parcours.

Au cours du précédent quadriennal j'ai participé activement à deux opérations de l'équipe LaTTiCe, celle dirigée par M. Charolles et M. Piot sur l'expression des relations hypothético-conditionnelles, et celle de S. Prévost portant sur l'ordre des mots et la structure informationnelle. Les travaux et discussions entrepris dans cette dernière opération m'ont notamment permis de confronter

différentes approches à même de rendre compte de la structuration informationnelle. Dans la prolongation de ces travaux, S. Prévost et moi-même avons proposé une nouvelle opération pour le prochain quadriennal (2010-2013) portant spécifiquement sur les fonctions pragmatico-informationnelles des éléments initiaux dans la phrase, l'opération EIOMSIT (Éléments initiaux, ordre des mots, structures informationnelle et textuelle). A partir d'une classification fine, nous projetons de mener diverses études sur les combinaisons possibles entre les différents éléments initiaux ainsi que sur les corrélations spécifiques entre les éléments initiaux et la position du sujet. La notion de genre occupera une place centrale dans ce projet car différentes comparaisons sont prévues entre la prose littéraire (extraits de romans/nouvelles), les articles scientifiques et les articles de presse.

Je collabore aussi depuis sa création au projet ANR Blanc, intitulé « Spatial Framing Adverbials », dirigé par M. Charolles et L. Sarda (2006-2010), dont l'objectif est de décrire les adverbiaux introduits par des prépositions spatiales et leur changement de sens et de fonction selon leur position. Après avoir collaboré à l'étude quantitative sur corpus, je participe à l'élaboration d'un numéro spécial de *Travaux de linguistique* pour le second semestre de 2009, tandis qu'avec L. Sarda nous avons aussi entrepris plusieurs travaux contrastifs (français/anglais) sur les fonctions des groupes adverbiaux en tête et en fin de phrase. Pour clôturer le projet, nous organisons actuellement un grand colloque international : « Linguistic and Psycholinguistic approaches to Text Structuring », qui aura lieu à l'ENS ULM en septembre 2009.
Un autre événement clé de mon parcours récent a été la création en 2007 de la revue *Discours*[2], une revue électronique interdisciplinaire qui s'intéresse tout particulièrement à la cohésion du discours. Bien que représentant un investissement personnel assez lourd pour les membres de équipe éditoriale, l'expérience s'avère très enrichissante et notre jeune revue commence à prendre son envol.

[2] http://discours.revues.org/

Cette brève description de mon parcours me permet de souligner la grande diversité des courants intellectuels, des cultures linguistiques et des cadres institutionnels qui m'ont influencée.
De l'anglais scientifique aux contes inuits, je suis aussi consciente que cet itinéraire peut apparaître un peu hétéroclite par moments. Cependant, à travers cette diversité, il est possible d'identifier des constantes qui sous-tendent ma réflexion (des constantes qui seront davantage explicitées dans le mémoire qui suit).

1° La linguistique fonctionnelle et le texte au sens large

Une première constante est pour moi la nécessité de décrire (et d'expliquer) les expressions linguistiques dans le cadre fourni par l'environnement pragmatique de l'interaction verbale. Je veux examiner l'impact de certains choix et d'agencements de constituants (des choix qui ne sont pas déterminés par le système de la langue) sur l'interprétation du texte en cours et en même temps évaluer l'influence des facteurs contextuels sur les choix qui sont faits.

Parmi les modèles linguistiques existants, ce sont ainsi les approches fonctionnelles qui correspondent les mieux à mes options théoriques, dans la mesure où ces théories cherchent à relier les modules syntaxiques aux modules sémantiques et pragmatiques. Les approches fonctionnelles que je connais les mieux sont celles de l'Ecole pragoise, la grammaire fonctionnelle de S. Dik, la linguistique systémique de M. Halliday et le *Role and Reference Grammar* (surtout en ce qui concerne la partie pragmatique de cette dernière – une approche fondée essentiellement sur les théories de K. Lambrecht).

2° Des textes « situés »

Afin de mettre en relation les formes examinées avec une situation d'énonciation précise, un autre impératif est la nécessité de travailler sur les corpus situés – d'où l'importance que j'accorde à la notion de genre. Les corpus, constitués de différents genres de **textes scientifiques** (article de recherche primaire ; article de vulgarisation, communication de colloque) sont le point de départ d'un bon nombre de mes travaux.

3° Structure informationnelle

Une autre constante qui sous-tend une large part de mes recherches en linguistiques est l'étude de la structuration informationnelle – et surtout l'identification des marqueurs de la structure informationnelle au ras du texte. Ces réalisations de la structure informationnelle sont aussi mises en relation avec une étude du genre textuel (cf. *supra*).

4° Problématiques descriptives et applicatives

Si mes recherches sur la syntaxe fonctionnelle et sur la structure informationnelle impliquent nécessairement une analyse syntaxique fine, ce n'est pas dans une visée seulement descriptive. Des considérations d'ordre applicatif (pour la pédagogie ou la modélisation) sous-tendent aussi la majorité de mes travaux.

Partie A

De la théorie à la pratique

.

Chapitre 1 – Aborder le texte

Résumé

Dans ce premier chapitre, je passe en revue différentes théories et approches en linguistique textuelle qui ont influencé mon parcours intellectuel (1.1). Ces influences (tout comme celles présentées dans les chapitres suivants) ne doivent pas être considérées comme correspondant forcément à un cheminement intellectuel d'une parfaite linéarité sur le plan chronologique. Tout parcours de recherche implique inévitablement un certain va-et-vient, des remises en cause, en plus d'une nécessaire ouverture à des idées nouvelles. Certaines théories invoquées dans ce chapitre, comme celles portant sur la cohésion intra et interphrastique de M..A.K. Halliday et R. Hasan (1976) ont surtout marqué mes premières recherches sur l'évaluation des travaux écrits en classe de langue. D'autres, et en particulier tout ce qui concerne la structure informationnelle des phrases, ont influencé l'ensemble de mes travaux

J'aborde également dans ce chapitre l'articulation texte/discours (1.2). Après avoir discuté des strictes distinctions parfois maintenues entres les notions de texte et de discours, je souligne ce que j'estime être la nécessaire complémentarité des deux notions. Dans mes travaux sur l'analyse des textes scientifiques (de types et de disciplines variés), j'effectue un va-et-vient constant entre des considérations textuelles et des considérations plus pragmatiques liées à l'insertion de l'objet textuel en question au sein d'un genre et d'un discours. Autrement dit, je relie une approche résolument ascendante, où j'analyse les choix linguistiques faits par les locuteurs au ras du texte, à une approche descendante qui évalue ces choix en fonction du contexte extralinguistique spécifique.

Je termine le chapitre par une mise en garde contre certaines approches cognitivistes qui privilégient le « tout pragmatique » (1.3). Si le rôle du récepteur dans l'interprétation d'un texte est à prendre en compte, il est essentiel également de ne pas dévaloriser le texte lui-même. Les choix opérés par le locuteur dans la construction du texte laissent forcément des traces à la surface textuelle. Une fois mises en relation avec une analyse de la situation extralinguistique qui a engendré le texte, ces traces fournissent une aide précieuse au récepteur dans son parcours interprétatif.

Chapitre 1 – Aborder le texte

Texte : du latin *textus*, « chose tissée », de *texere*, « tisser »

1.1 De la phrase vers le texte

Le point de départ de mes travaux de recherche en linguistique a été le *texte* et les notions de textualité ou de *texture*. Qu'est-ce qu'un texte ? Comment font les locuteurs pour créer un bon texte ? Pouvons-nous mesurer la qualité textuelle ? Avec de tels questionnements on dépasse automatiquement la notion de phrase avec son accent sur la grammaticalité *per se*, pour s'intéresser à un niveau « supérieur » – un niveau qui se situe à l'interface de la syntaxe et de la pragmatique.

L'étymologie du mot *texte* donne une image assez juste de l'usage que j'en fais dans mes travaux : une chose tissée, en l'occurrence des mots tissés, des paroles tissées, des phrases tissées, des paragraphes tissés. Dans cette image, le locuteur est le tisseur, et son travail est celui de bien tisser un ensemble cohérent, un ensemble qui sera perçu comme communicatif pour son/ses recépteur(s).

Dans mes premiers travaux, je me suis surtout intéressée à l'évaluation de la qualité des textes des apprenants (des devoirs écrits) dans le cadre de l'apprentissage de l'anglais comme langue étrangère. Une des premières conclusions de ces travaux [THOM 98 ; THOM 99b et c ; THOM 00] concerne l'inadéquation d'une évaluation portant exclusivement sur les erreurs et maladresses de portée phrastique. Les règles au niveau de la langue peuvent être parfaitement respectées à un niveau local, mais si le destinataire du texte n'arrive pas à suivre le déroulement du texte dans sa globalité, le texte ne sera pas perçu comme une « réussite ». Sans être forcément incohérents, ces travaux d'apprenants peuvent très souvent dégager une impression d'étrangeté – une impression qui ne sera pas facile à cerner (et encore moins à corriger).

Dans tout texte il doit y avoir à la fois mise en œuvre de règles conscientes, celles du système de la langue utilisée, et mise en œuvre de règles souvent moins conscientes, liées à ce processus de tissage au sein d'une situation de communication déterminée. Dans un article sur les corrections en usage chez les maîtres d'école des classes primaires en France, M. Charolles (1978) faisait état de la pénurie d'outils conceptuels et analytiques pour aborder ce deuxième niveau de mise en texte. Cette question semblerait toujours d'actualité aujourd'hui.

En parlant d'un « texte », nous soulignons le fait qu'il s'agit d'un ensemble et non pas d'une suite de propositions ou de phrases disparates et sans rapport aucun. L'accent est mis sur son unité. Cependant il est difficile de cerner de manière précise comment cette unité est créée et de s'approprier des outils analytiques capables de l'appréhender [THOM 00 ; 08d].

Pour les grammairiens du texte dans les années 60 et 70 – par exemple R. Harweg (1973), J.S. Petöfi (1973), T.A. van Dijk (1972) – la définition du concept de texte a été surtout grammaticale. Le principe de compositionalité s'est étendu de la phrase au texte. En s'inspirant de la grammaire générative de N. Chomsky, ces linguistes espéraient fournir une description structurale de la compétence textuelle. Leur projet reposait sur le fait que toutes les suites de phrases acceptables en tant que phrases ne sont pas forcément acceptables en tant que texte. A l'aide d'un système de règles appropriées, ils cherchaient à prédire en quelque sorte les intuitions des sujets parlants, en ce qui concerne la bonne formation textuelle.

Les limites de cette approche sont vite apparues. Les exemples et contre-exemples de ce qui constitue un texte cohérent se sont succédé (van Dijk 1972 ; Enkvist 1978 ; T. Reinhart 1980). Le texte s'est avéré être une entité trop complexe pour être appréhendé par une série de règles. La plupart des chercheurs concordent maintenant avec Charolles quand il déclare :

> *Il n'existe pas au plan du texte, de règles de bonne formation qui s'appliqueraient en toutes circonstances et dont les violations, comme c'est le cas en syntaxe de phrases, « feraient l'unanimité* ». (1988 : 52)

A partir de là, il devient également difficile de maintenir que certaines suites de phrases ou de morceaux de textes seraient cohérents ou incohérents en eux-mêmes.

Parmi les linguistes qui on tenté d'aborder le texte autrement que comme une extension de la phrase, l'approche préconisée par M.A.K. Halliday et R. Hasan reste sans doute la plus connue. Pour eux le texte n'est pas un élément grammatical mais un élément d'« usage du langage » :

> *A text is a unit of language in use. It is not a grammatical unit, like a clause or sentence; and is not defined by its size. A text is not something like a sentence but only bigger, it differs from a sentence in kind.* (1976 : 1-2)

Le fait de souligner aussi la taille variable des textes permet de relativiser l'existence de la phrase comme base de la textualité. Un texte peut contenir plusieurs volumes de même qu'il ne peut comporter qu'une phrase ou même qu'un mot. Mais Halliday et Hasan insistent surtout sur la nature très différente du texte par rapport à la phrase. Il s'ensuit que les règles (ou contraintes) auxquelles est soumis le texte ne sont donc pas du même ordre que celles qui gouvernent la morphosyntaxe.

1.1.1 Procédés de cohésion formelle

Halliday et Hasan (*ibid*.) font un inventaire des moyens linguistiques qui restent théoriquement à la disposition des locuteurs de l'anglais pour créer la « texture » qui permet au texte d'apparaître comme un texte[1]. À travers la notion de cohésion, définie comme *« the means whereby elements that are structurally unrelated to one another are linked together, through the dependence of one on the other for its interpretation »* (*ibid*. : 27), les auteurs proposent un inventaire des relations de cohésion (*« cohesive ties »*). Ils classent ces procédures lexico-grammaticales cohésives sous cinq grandes rubriques : référence (anaphores et cataphores), ellipse, conjonctions (coordinateurs et connecteurs) et cohésion lexicale.

[1] Ces moyens ne fonctionnent évidemment pas de la même façon d'une langue à une autre : cf. par exemple L. Lundquist (2005).

Cette taxinomie de la cohésion a été le point de départ de nombreuses études, en particulier en didactique anglaise, sur l'utilisation des procédés de cohésion dans les textes des apprenants. En effet, dans l'ouvrage de Halliday et Hasan, la cohésion semble être présentée comme une cause de la cohérence textuelle. Un texte qui est reconnu en tant que tel, qui est cohérent ou qui possède une *texture* dans la terminologie des auteurs, contiendra certains indices formels de cohésion :

> *If a passage of English ...is perceived as a text, there will be certain linguistic features present in the passage which can be identified as contributing to its total unity and giving it texture.* (*Ibid.* : 2)

Un grand nombre de manuels scolaires et/ou universitaires destinés aux étudiants en anglais comportent un chapitre sur les différents moyens d'assurer la cohésion dans un texte conformément à la classification de Halliday et Hasan. Il existe plusieurs travaux dont le but spécifique est de tester empiriquement le rôle de la cohésion dans la cohérence (Connor 1984 ; Hartnett 1986 ; Erhlich 1988). Cependant, la plupart des ces études concluent à l'absence d'un lien direct de cause/effet entre les deux notions. La présence de liens cohésifs ne suffit pas en elle-même. Il faut d'abord que ces liens soient employés de façon appropriée. Un lien, par exemple, qui ne renvoie pas à un élément jouant un rôle majeur au niveau de la structure informationnelle, ne sera pas cohésif. Dans l'exemple suivant :

> *The first of the antibiotics, the Times reported, was discovered by Sir Alexander Fleming. <u>It</u> corrected its error the following day.* (Ehrlich 1988 : 114)

On constate que le pronom *it* n'est pas cohésif, car il renvoie à un élément qui, en tant que partie de la proposition subordonnée (« *the Times* »), joue un rôle secondaire sur le plan informationnel. Le récepteur d'un tel texte serait désorienté.

Cependant même si cette relation était signalée avec un procédé cohésif approprié d'un point de vue lexico-grammatical, il n'y a pas de garantie qu'elle soit forcément parlante pour le récepteur. La marque de cohésion n'est en fait que la partie visible d'une relation que l'émetteur souhaite signaler. Un texte qui possède beaucoup de marques de cohésion fournit des indications explicites

pour l'interprétation du texte en question, mais la cohésion, à elle seule, ne suffit pas à rendre un texte cohérent. Ce n'est pas un but en soi :

> *Cohesion is a means to an end, not the end itself. The success of a writer's prose depends upon much more than successful use of any cohesive devices.* (Hartnett 1986 : 152)

Cela dit, il ne faudrait pas minimiser non plus le rôle potentiel des marques de cohésion. La présence des liens de cohésion formels peut être d'un grand secours pour le récepteur. Dans certains types de textes, une signalisation explicite de leur organisation interne est très importante. Quand il s'agit par exemple de comprendre certaines explications dans un manuel scolaire, le récepteur/lecteur a besoin d'être guidé au maximum. Il est également vrai que, pour la plupart des textes produits par les élèves ou étudiants en situation d'apprentissage, l'utilisation de marques formelles de cohésion est très valorisée. Il n'empêche qu'une concentration exclusive sur les marques explicites de cohésion interphrastique peut s'avérer inadéquate pour comprendre ce qui crée l'unité d'un texte et l'effet de texture.

1.1.2 D'autres formes de connexité

Il est possible d'imaginer un texte sans aucun lien cohésif formel du type délimité par Halliday & Hasan, qui pourrait néanmoins être considéré comme cohérent pour un récepteur donné dans une situation spécifique. Ainsi G. Brown et G. Yule fournissent-ils plusieurs exemples de textes qui ne possèdent pas de liens *explicites* de cohésion (1983). Cependant, ils estiment que les lecteurs pourront facilement trouver dans les phrases contiguës des relations sémantiques, ce qui leur permettra de conférer une cohérence à ces textes.

> *It seems to be then that 'texture', in the sense of explicit realisation of semantic relations, is not critical to the identification and co-interpretation of texts.* (1983 : 196)

Autrement dit, les relations sémantiques entre les phrases peuvent être établies sans recours à des marques formelles explicites.

Le processus de tissage peut aussi faire intervenir d'autres séries de relations ou de connections, comme celles désignées par le terme *cadre* (Charolles

1997). Charolles, en fait, envisage l'existence de deux types de marques de cohésion : la relation de connexion et la relation d'indexation. Les relations de connexion sont à rapprocher des cinq procédures de cohésion de Halliday et Hasan. Il s'agit de relations linéaires, évoluant de proposition en proposition.

> D'un côté, on trouve une vaste gamme d'expressions signalant qu'une relation doit être établie entre deux unités adjacentes ou proches. Cette classe comprend elle-même deux grandes sous-classes selon que la relation indiquée met en jeu des référents (anaphores) ou des énoncés avec leur valeur illocutionnaire et leur contenu propositionnel (connecteurs). (1997 : 2)

Le deuxième type de relation n'établit pas de véritable relation entre deux items dans le texte mais désigne un processus d'indexation.

> D'un autre côté, on a toute une série d'expressions marquant que plusieurs unités doivent être traitées de la même manière relativement à un critère (plus ou moins) spécifié par ces expressions. (*Ibid.* : 3)

À cette seconde classe de relations appartiennent les expressions introductrices de *cadres* dont, par exemple, certains adverbiaux en position « pré-sujet ». Une propriété essentielle de ces expressions introductrices est le fait de ne pas porter seulement sur la proposition à laquelle elles sont adjointes, mais de pouvoir intégrer sous leur « influence » une ou plusieurs propositions qui forment un tout sémantiquement et discursivement homogène. Cette fonction intégrative, c'est-à-dire la capacité qu'elles ont d'englober dans leur portée sémantique la ou les proposition(s) qui suivent, leur sert à structurer (à répartir et à subdiviser) l'information au fil du discours.

Une autre piste très intéressante pour étudier l'organisation intra et interphrastique et le fonctionnement interne des textes est évidemment fournie par tout ce qui touche en général à la structure informationnelle des propositions : l'identification des thèmes/topiques et surtout les progressions thématiques établies entres les propositions. Nous nous attarderons en détail sur ce modèle de l'organisation textuelle dans le chapitre 2 de ce mémoire, ainsi que sur les liens assez complexes entre les adverbiaux cadratifs et la notion de topique ou de thème.

Les apports de la structure informationnelle dans une perspective didactique m'apparaissent toujours particulièrement dignes d'intérêt. Dans mes travaux antérieurs sur l'apprentissage de l'anglais langue étrangère par des étudiants francophones, ainsi que dans certains travaux plus récents sur la structuration des articles de recherche en anglais écrits par des chercheurs non anglophones, j'ai été frappée par le nombre de problèmes de « style » imputables à des considérations liées à ces questions. Si la structuration des propositions individuelles et/ou les progressions thématiques dans lesquelles ces propositions rentrent ne correspondent pas aux attentes des récepteurs, le texte risque de ne pas être suivi ou compris. Très souvent un texte, dépourvu en surface d'erreurs perceptibles, peut être considéré comme « défaillant » parce que sa structuration thématique ne répond pas aux attentes de son lecteur. Ces aspects seront abordés dans le chapitre consacré aux aspects plus « applicatifs » de ma recherche (chapitre 7)

1.1.3 Cohésion et cohérence : pour une approche intégrée

Dans l'analyse de toutes ces formes de connexion et de cohésion, il est toutefois très difficile de faire abstraction du monde extérieur au texte. Il est vrai que la taxinomie de la cohésion de Halliday et Hasan souligne la nature purement formelle des liens établis (*« When we consider cohesion, we are investigating the linguistic means whereby a text is enabled to function as a single meaningful unit »* ; 1976 : 30). Les données extralinguistiques (situationnelles ou pragmatiques) ne sont pas du tout prises en compte dans cette définition. Mais comme de nombreux chercheurs l'ont démontré, une séparation entre ces deux plans est très délicate à maintenir (par exemple, Adam 1990 ; Charolles 1993 ; 1995 ; Péry-Woodley 1993 ; 2000). Un texte peut exhiber des signes de cohésion et de connexion parfaites sans pour autant être cohérent ou parlant pour son ou ses destinataires. La présence d'un lien cohésif ne peut être confirmée sans que soient pris en compte d'autres facteurs extérieurs à cette manifestation formelle. Autrement dit, l'interprétation des procédés de cohésion dépendrait elle-même du modèle interprétatif global en cours d'élaboration.

En tant qu'unité fonctionnelle, le texte implique la participation des deux membres de la communication : producteur et récepteur. M.-P. Péry-Woodley l'explique de la façon suivante :

> Cette notion de cohérence réunit en fait deux facettes distinctes du fonctionnement des textes : la construction par le(s) scripteur(s) d'un discours cohérent dont le texte est la trace ; l'élaboration par le lecteur, à partir de cette trace, d'un modèle interprétatif cohérent. (2001 : 28)

L'interprétation d'un texte reste une procédure active. Si un récepteur interprète un texte comme cohérent, il aura trouvé une interprétation qui correspond à sa vision du monde, car la cohérence n'est pas strictement dans le texte mais résulte de l'interaction avec un récepteur potentiel. M. Charolles (1978 ; 1988, 1995) pose un principe premier de cohérence, un principe qui orientera le lecteur dans une lecture cohésive des marques dans le texte :

> La cohérence n'est pas une propriété des textes […] Le besoin de cohérence est, par contre, une sorte de forme *a priori* de la réception discursive. (1988 : 55)

Ce « besoin » de cohérence auquel se réfère Charolles semble important à souligner. Ayant présupposé une cohérence au texte en question, le récepteur va tout mettre en œuvre pour établir cette cohérence Même face à un texte très dense ou décousu le lecteur hésitera à juger un texte comme « incohérent ». Selon Charolles encore :

> Tout se passe comme si le « récepteur ignorant » faisait crédit de cohérence à l'émetteur, admettait qu'il a ses raisons (supérieures aux siennes) et s'efforçait précisément de les retrouver afin de reconstruire le suivi de son discours. (1978 : 38)

Vu sous cette lumière, la notion de cohérence apparaît très intuitive et étroitement dépendante de chaque interprétant. Cependant, cela ne signifie pas qu'elle soit un phénomène exclusivement subjectif. Un texte produit par un locuteur non natif et qui de ce fait contient de nombreuses fautes lexico-grammaticales ou orthographiques peut sérieusement entraver la perception de la cohérence chez tout lecteur. Le contexte et le genre de discours peuvent eux aussi fortement influencer l'interprétation d'un texte. Dans une situation discursive précise, au sein d'un genre spécifique, le lecteur aura des attentes particulières qui vont l'influencer dans l'interprétation de la cohérence du texte (cf.

chapitre 6 *infra*). De plus, à chaque genre sont associées des stratégies interprétatives qui se manifestent par des instructions intrinsèques. La construction d'une interprétation cohérente passera par la facilité à suivre ces instructions.

Dans mes travaux je défends l'hypothèse qu'il existe un lien fort entre stratégies organisationnelles et genre discursif. Un article de recherche primaire destiné à un lectorat averti n'adoptera pas les mêmes stratégies organisationnelles qu'un article de vulgarisation destiné à un vaste public. On peut s'attendre à différentes exigences selon les différents genres discursifs.

De la même façon, des cultures linguistiques différentes peuvent aussi présenter des variations importantes quant au degré ou au type de stratégie organisationnelle mise en œuvre au sein d'un genre particulier. Il est souvent dit par exemple que les auteurs anglophones d'articles de recherche ont tendance à employer plus d'indications metatextuelles pour guider l'interprétation des lecteurs que dans certaines autres cultures linguistiques. C'est la responsabilité du rédacteur ou de l'émetteur du message de rendre explicite son argumentation. Dans la culture française en revanche, la responsabilité incombe davantage au lecteur/récepteur et on peut s'attendre à une signalisation moins explicite. (Voir par exemple, Dahl 2004b ; Flottum *et al.* 2007.)

Le mode de transmission du texte – oral ou écrit – va aussi avoir un impact important sur les types de stratégies informationnelles mises en œuvre et sur la nature des formes de cohésion et de connexité susceptibles d'être employées. La communication écrite, par rapport à la communication orale spontanée, implique des contraintes très spécifiques en rapport avec la nature de cette interaction : absence d'un interlocuteur « en direct », distance dans le temps et dans l'espace, autonomie contextuelle. Cette absence d'un contexte partagé en temps réel implique dans bien des cas une exigence d'explicitation beaucoup plus forte que dans le cas d'un texte oral.

La cohérence d'un texte ne peut être ainsi établie de façon absolue. Elle sera relative au degré de facilité éprouvé par le récepteur à l'interpréter dans une situation particulière. En effet, les stratégies et moyens pour exprimer cette cohérence sont différents selon la culture linguistique, selon le genre de texte concerné et selon son mode de transmission. Seule une analyse fine de la situation de communication permettra de juger de l'efficacité de ces moyens.

Si la cohérence n'est pas directement dépendante de la cohésion, devrions-nous alors affirmer que le contraire est vrai ? La cohésion dépend-elle de la cohérence ? Si la réponse à cette question est certainement affirmative – afin de reconnaître un lien cohésif, le récepteur a besoin de prendre en compte la cohérence globale – elle mériterait cependant d'être nuancée. Tout est question de perspective. Cohérence et cohésion entrent dans un rapport dialectique. Le lecteur construit la cohérence textuelle en partie grâce à certains signaux, tels que des liens de cohésion à la surface textuelle ; cependant, son interprétation de ces signaux reste en même temps dépendante de sa perception de la cohérence du texte en question – une perception qui reste fortement contrainte par des considérations extralinguistiques. Il y a un mouvement constant entre ces plans.

1.2 Du texte au discours

Dans la perspective adoptée ici, un jugement de cohérence passe donc par une articulation de l'unité texte avec le contexte socio-pragmatique de l'interaction : autrement dit par la mise en relation du texte avec le discours qui l'a engendré.

Dans les paragraphes suivants, je m'efforcerai de décrire la distinction souvent faite entres les termes *texte* et *discours*. Je préciserai ensuite la manière dont j'ai envisagé les relations entre les deux termes dans mes propres travaux.

1.2.1 Distinction texte /discours

Si dans la langue courante les deux termes sont distingués sur la base d'une opposition entre l'écrit et l'oral, en linguistique leur emploi est tout autre. La

linguistique ne réduit pas la distinction à la question du support. Réduire la distinction au support occulte le fait qu'un texte peut faire intervenir plusieurs sémiotiques à la fois (du verbal et du non verbal) : un article de recherche dans un périodique contient très souvent des images et des schémas ; l'enregistrement vidéo d'un cours magistral contiendra des informations gestuelles. Tout comme le terme *discours* n'est pas forcément associé avec la notion d'oralité, le texte ne sera donc pas obligatoirement un texte écrit

La distinction texte/discours telle qu'elle est traditionnellement comprise en linguistique repose en grande partie sur la question du contexte. Selon J.M. Adam une première distinction qu'il déclare « assez communément admise aujourd'hui » se résume de la façon suivante :

> DISCOURS = Texte + Conditions de production
>
> TEXTE = Discours - Conditions de production. (1990 : 23)

Autrement dit, le discours n'est pas seulement caractérisé par ses propriétés textuelles, mais également par son existence dans une situation de communication particulière. En revanche, le texte serait un objet plus abstrait obtenu par la soustraction du contexte au discours concret. Cette définition fait écho à la définition proposée par D. Slakta quelques années auparavant lorsqu'il distinguait le texte, « objet formel abstrait » du discours, « pratique sociale concrète » (1975 : 31).

Toutefois, cette distinction a tendance à s'estomper de nos jours. Beaucoup de linguistes récusent cette séparation stricte entre les domaines du texte et du discours. Comme nous l'avons vu dans la partie précédente (1.1.4), même si on choisit de se focaliser sur les propriétés internes de l'unité texte, il est difficile en même temps de faire abstraction du monde extérieur. Selon F. Rastier par exemple :

> Il n'existe pas de texte (ni même d'énoncé) qui puisse être produit par le seul système fonctionnel de la langue (au sens restreint de mise en linguistique). (1989 : 37)

Une approche purement linguistique de l'étude du texte où ne seraient pas prises en compte les pratiques discursives qui l'entourent serait inutilement

restrictive. Un type de langage approprié et donc efficace dans une situation donnée ne le sera pas forcément dans une autre. Si, par exemple, nous trouvons plus de nominalisations et de passivations dans un article de recherche scientifique que dans un anglais littéraire, c'est en rapport direct avec les phénomènes analysés et avec leurs buts communicatifs très différents. Alors qu'une description événementielle impliquera l'emploi de phrases chronologiques avec formes verbales, dans l'énoncé d'une règle ou d'un théorème la forme nominale sera préférée. Le lien avec le discours est ce qui motive dans bien des cas le choix de telle construction plutôt que de telle autre.

Pour cette raison, je préfère considérer le texte comme *« the linguistic product of a discourse process »* (Seidlhofer & Widdowson 1999 : 206) ou comme « la trace d'un discours ancré dans un contexte » (Péry-Woodley 2000) – trace verbale, ou bien trace non verbale, signal prosodique, iconographique ou gestuel. C'est cette trace, une fois mise en relation avec les données contextuelles et pragmatiques qui l'entourent, qui permet la reconstruction du discours.

1.2.2 Des notions complémentaires

La séparation entre discours et texte n'est pas de l'ordre d'une véritable opposition mais d'une différence de conceptualisation. Texte et discours en tant que « objet formel » et « pratique sociale » se chevauchent. En parlant du *texte*, on met l'accent sur le produit linguistique et sur les choix de formulation, tandis qu'avec le terme *discours* on privilégie plus les facteurs situationnels et interpersonnels. Pour cette raison, je préfère une utilisation des deux termes en complémentarité.

Dans mes travaux, je fais un va-et-vient constant entre les deux plans. Il s'agit d'identifier et d'expliquer les choix de formulation au ras du texte dans la perspective de leur fonctionnement en situation de communication. Je souhaite examiner l'impact de ces choix sur le discours en cours d'élaboration et également évaluer l'influence des facteurs contextuels sur les choix qui sont faits.

L'actualisation du texte en discours est une activité coopérative et dynamique, dans laquelle les partenaires de discours exploitent conjointement les traces et indices laissés à la surface textuelle en convoquant un contexte approprié afin de construire du discours (cf. Cornish 2006). Toutefois, comme Cornish le souligne, afin d'arriver au message voulu par l'énonciateur il ne s'agit nullement « d'un simple décodage du texte » (*ibid.* : 3). Le récepteur s'approprie le texte et restructure l'information reçue en fonction de la situation extralinguistique et à la lumière de ses propres savoirs et connaissances. Il est vrai que dans ce modèle dynamique de la construction de discours, le texte peut être considéré comme incomplet et indéterminé. C.S. Butler l'exprime de la façon suivante :

> *Text (…) can be seen as a very imperfect and at times unreliable trace of what happens during such a discourse process.* (2003 : 304)

La tâche principale du linguiste travaillant dans cette perspective consiste justement à définir une base solide pour étudier la mise en relation du texte au discours.

1.3 Dangers du « tout pragmatique »

En mettant en avant le rôle des apports externes au texte dans l'interprétation, la pragmatique a tendance à réduire l'importance des choix de formulation. Toutefois, je tiens à me distancier des approches privilégiant le « tout pragmatique » qui minimalisent le rôle des marques formelles.

Par un mouvement de pendule on est passé ces dernières années d'une approche microlinguistique privilégiant les phénomènes phrastiques à une approche très globalisante et cognitiviste qui aurait tendance à gommer les traces laissées par le rédacteur dans la construction du texte (Reboul et Moeschler 1998). Si un intérêt exclusif pour l'intérieur du texte et ses régularités syntaxiques, lexicales et phoniques ne suffit pas à expliquer son fonctionnement – comme l'écrivent R. de Beaugrande et W. Dressler, « *those artifacts are inherently incomplete when isolated from the processing operations performed upon them* » (1981 : 35) – la tendance inverse est également à éviter. Les mêmes auteurs, en effet, mettent en garde contre une tendance à exagérer l'importance du récepteur et son rôle dans l'interprétation du texte et à dévaloriser ainsi le texte lui-même :

> *On the other hand we must guard against allowing the text to vanish away behind mental processes. Recent debates over the role of the reader point up the dangers of assuming that text receivers can do whatever they like with a presentation. If that notion were accurate, textual communication would be quite unreliable, perhaps even solipsistic. There must be definitive, though not absolute controls on the variations of utilising a text by different receivers.* (*Ibid.* : 35)

Comme le dit Charolles :

> Le point étant d'arriver à comprendre comment chaque langue a sécrété des moyens plus ou moins originaux, plus ou moins spécialisés, plus ou moins fiables, plus ou moins économiques et fiables, pour résoudre (au mieux) les problèmes d'interprétation que son usage pouvait engendrer. (1988 : 62)

La nature de ces « moyens de contrôle » ou « prises », c'est-à-dire le détail de ces indications guidant l'interprétation du récepteur, est au centre de mes travaux et va donc être au cœur de ma réflexion dans les chapitres suivants.

1.3.1 Prises sur le texte

L'interprétation de la cohérence d'un texte est, comme nous l'avons vu, un processus largement idiosyncrasique. Le rôle du récepteur dans ce processus est primordial. Néanmoins, il est possible d'anticiper un certain degré d'uniformité en ce qui concerne à la fois la forme et le fond du texte[2], en prenant en compte par exemple les contraintes relatives au genre textuel en question. L'organisation globale du texte selon des procédés rhétoriques spécifiques au genre permet de formuler certains jugements généraux quant à l'efficacité de ce texte. Dans le cas de textes spécialisés scientifiques, les normes et contraintes sont très fortes et peuvent avoir des implications importantes pour la mise en mots [THOM 08d]. Dans les chapitres suivants (cf. en particulier chapitre 6) je vais détailler certaines des contraintes de « genre » qui président à l'interprétation des textes scientifiques (textes oraux et textes écrits).

Une autre « prise » sur le texte et sur l'interprétation de sa cohérence est fournie par une étude de sa structuration informationnelle. Dans une telle approche la phrase est analysée, non seulement comme une structure

[2] Ceci dépend, en partie, du genre en question – certains genres textuels ayant des contraintes plus rigoureuses que d'autres (cf. chapitre 3).

syntaxico-sémantique, mais également comme une structure porteuse d'information au sein d'une certaine dynamique textuelle. L'étude de la construction textuelle, par le biais d'une analyse de sa structuration informationnelle en thème et rhème (ou topique et commentaire[3]) permet de comprendre comment les rédacteurs maîtrisent le flux d'information et arrivent ainsi à guider l'interprétation du récepteur.

Une analyse informationnelle présente également l'avantage de nous permettre d'associer en même temps des considérations sur le genre de texte. L'organisation informationnelle et le genre textuel sont en effet étroitement imbriqués. À un niveau phrastique, ce que l'émetteur choisit de mettre en thème sera en relation étroite avec le sujet général de son discours et du genre textuel dans lequel il s'inscrit. Dans un rapport technique, par exemple, on s'attend à ce que l'émetteur nous parle de procédés ou d'objets inanimés appartenant au monde réel. Dans un tel texte il serait inhabituel de trouver des sujets thématiques personnalisés. Quand nous parlons de la progression thématique d'un texte, nous pouvons également envisager cette progression par rapport au type de texte en question. Certains genres textuels favorisent un certain type de progression plutôt qu'un autre (cf. section 7.1 *infra*).

Une évaluation de la cohérence ne peut pas se faire en vase clos. C'est pour toutes ces raisons que j'ai récemment proposé un enseignement de la cohérence axé sur le genre [THOM 08d]. Les différentes prises sur la cohérence fournies par les indices de cohésion (explicites) ou par les différentes configurations de structuration informationnelle peuvent varier en fonction des besoins discursifs. Une prise en compte du genre aide à stabiliser ce contexte discursif.

1.3.2 Articulation phrase/texte/discours

Comme déjà préconisé par M. Charolles et B. Combettes (1999 :110), je pense qu'il serait utile « de repenser les relations entre les approches locales… et les

[3] Voir les chapitres suivants pour l'approfondissement d'un certain nombre de questions relatives aux notions de la structure informationnelle et pour des explications sur la terminologie employée.

approches globales (types de textes, genres) ». Aucun texte n'est écrit seulement « dans une langue ». Tout en tenant évidemment compte des règles de la langue, un texte est écrit dans un genre au sein d'un discours

Les idées évoquées dans ce chapitre ont cherché à défendre le point de vue que de la phrase au texte et du texte au discours, il y un continuum, avec le texte comme pivot central. On peut aller en amont (le discours), tout comme on peut aller en aval en examinant le détail des phrases et propositions au ras du texte :

Discours ←→Texte←→ Phrase (Proposition)

Afin d'être efficace, une analyse linguistique du texte doit constamment naviguer entre les approches ascendantes et les approches descendantes. Ces principes ont guidé l'ensemble de mes travaux.

Chapitre 2 – Analyser la structure informationnelle

Résumé

Dans le deuxième chapitre je fais état de trois théories de la structure informationnelle1 qui m'ont profondément influencée : l'approche de l'Ecole de Prague ; l'approche de M. Halliday et de la linguistique systémique ; l'approche de K. Lambrecht. Il ne s'agit pas de faire une synthèse complète de chaque approche mais d'identifier les aspects les plus orignaux et les plus pertinents.

Je commence par l'approche pragoise (2.1). Sur le plan chronologique, ce sont les linguistes de l'École de Prague qui ont été les pionniers dans ce domaine en isolant un niveau d'analyse portant sur la distribution de l'information dans la phrase – une analyse articulée autour de la division de la phrase en partie thématique et partie rhématique. Un aspect important de l'approche pragoise concerne la notion de dynamisme communicatif proposée par J. Firbas (1964 ; 1992). Toutefois, je considère que l'apport le plus intéressant dans cette approche concerne la notion de progression thématique. En abordant explicitement la question des enchaînements interphrastiques, les travaux de F. Daneš (1974) permettent d'articuler les niveaux locaux et les niveaux globaux de l'organisation textuelle.

L'approche adoptée par Halliday et la linguistique systémique fonctionnelle (2.2), tout en prenant appui sur l'approche pragoise, établit une distinction nette entre deux plans de structuration, le couple thème/rhème et le couple donné/nouveau. Les autres aspects auxquels je porte une attention toute particulière dans cette présentation de l'approche de Halliday concernent le critère de position, la notion de thème marqué et la notion de thèmes multiples.

L'approche de Lambrecht (2.3) permet surtout d'affiner les notions de « donné » et de « nouveau », en proposant un traitement fouillé des questions relatives au statut cognitif des référents, et de leur degré d'identifiabilité et d'accessibilité chez le récepteur. Je présente également la définition pragmatique du topique proposée par Lambrecht, avant de terminer par une présentation de sa notion de focus et de sa taxonomie de structures focales.

Dans la dernière partie de ce chapitre (2.4), je résume ce que je considère comme les principaux enjeux de la problématique informationnelle et je précise mon propre angle d'attaque.

[1] Dans ce mémoire, le terme *structure informationnelle* sera employé comme un titre global, à la place de *structure thématique* ou de *structure communicative*.

Chapitre 2 – Analyser la structure informationnelle

Les problèmes traités sous le label de structure informationnelle concernant la distribution de l'information dans la phrase et les modalités de la progression textuelle, font depuis longtemps l'objet de descriptions très différentes. Pour cette raison, bon nombre de linguistes restent très méfiants à l'égard de tout ce qui concerne cette dimension de l'organisation linguistique. En effet, les différentes tendances linguistiques européennes, anglo-saxonnes et scandinaves ont chacune adopté une approche et une terminologie spécifiques. Si certains se contentent d'une définition purement linguistique, d'autres préfèrent une définition fondée presque exclusivement sur des critères psychologiques. Il est vrai aussi que, si les termes de *thème*, *topique* et *focus* sont intuitivement accessible à tous, il existe une certaine confusion terminologique et conceptuelle à leur propos (Fradin & Cadiot 1988 ; Mondada 1994 ; Prévost 1998) [THOM 00]. En parlant du « thème » M. Galmiche, par exemple, fait part de :

> …l'extrême inconfort que l'on éprouve devant une notion (un concept ?) aux dénominations multiples, aux caractéristiques variées, souvent équivoques, voire contradictoires… (1992 : 3)

Une autre raison de cette réticence concerne une certaine subjectivité inhérente à ces notions. Même au sein d'une même approche, la délimitation d'une zone thématique et rhématique n'est pas toujours aisée (Lundquist 1980). Comme la structure informationnelle doit prendre en compte l'organisation pragmatique de l'énoncé, il est en effet difficile d'avoir le même degré de certitude ou d'objectivité pour les phénomènes relatifs à ses manifestations que pour les phénomènes relevant seulement par exemple de l'organisation syntaxique d'un texte. L'appréhension de l'organisation informationnelle, comme je l'ai souligné dans mes travaux précédents [THOM 00 : 51-52], dépendra toujours à un certain degré des réactions de chaque récepteur.

L'étude de la structure informationnelle implique nécessairement le traitement simultané de plusieurs dimensions linguistiques et pragmatiques. Elle porte à la fois sur l'organisation pragmatique du contenu dénotationnel d'un énoncé et sur sa codification du point de vue linguistique – une codification qui résulte d'une interaction complexe entre les indications données par le lexique, par la morphosyntaxe, par la prosodie et par la ponctuation. Comme Lambrecht le constate, l'imbrication de ces différentes perspectives peut mettre l'analyste dans une situation peu confortable :

> *Information structure research neither offers the comfort which many syntacticians find in the idea of studying an autonomous formal object or provides the possibility enjoyed by sociologists of putting aside issues of formal structure for the sake of capturing the function of language in social interaction.* (Lambrecht 1994 :1)

La complexité de la notion de structure informationnelle peut certainement en partie expliquer les *« negative and defeatist views »* (*ibid.*) qui continuent à circuler à l'égard de la notion. Toutefois, le foisonnement d'études consacrées ces dernières années aux notions de topique ou de thème témoigne en même temps de l'importance de cette notion en linguistique (cf. Grobet 2002 ; Pohriel 2005 *inter alia*). Comme B. Fradin et P. Cadiot le remarquent, en faisant état de la multiplication d'approches de la notion de topique ou de thème :

> *S'il est vrai que chacune de ces approches reconnaît la nécessité de recourir explicitement à une notion de thème, n'y a-t-il pas fort à parier que ce qu'elle vise à saisir n'est pas un artefact théorique, mais correspond bien à une réalité dont il va falloir rendre compte ?* (1988 : 4)

Les critiques portant sur la fluctuation de la notion portent finalement plus sur l'utilisation plus ou moins rigoureuse qui en est faite que sur cette notion en elle-même. Comme je l'ai démontré dans mes travaux précédents, une fois clairement explicitée l'approche adoptée [THOM 00], l'étude de la structuration informationnelle peut fournir un appareil puissant pour aborder un grand nombre d'aspects relatifs à la cohérence textuelle, sur le plan purement descriptif, mais aussi sur un plan plus applicatif [THOM 04b et d].

Dans les paragraphes suivants je propose de revenir sur les spécificités de trois approches : l'approche de l'École de Prague, l'approche de Halliday et la linguistique systémique, et l'approche de Lambrecht. Je propose d'identifier certains points de divergence et de convergence entre ces trois théories majeures, afin d'arriver à une terminologie stable et à un cadre d'analyse opérationnel[2].

2.1 L'approche de l'École de Prague

D'un point de vue chronologique, ce sont les linguistes de l'École de Prague, d'abord V. Mathesius dans les années 1920, puis Daneš et Firbas à partir des années 1960, qui ont joué en quelque sorte le rôle de pionniers dans le développement des analyses informationnelles. Néanmoins, si l'on remonte au XIXe siècle, l'ouvrage de H. Weil (1879/1991), *De l'ordre de mots dans les langues anciennes comparées aux langues modernes,* pourrait être considéré comme précurseur de certaines idées des linguistes pragois. Weil établit une distinction entre « mouvement des idées » et « mouvement syntaxique » ; dans les langues romanes et germaniques il estime que l'ordre de mots peut être considéré comme un reflet de l'ordre des idées. Weil pose que :

> *il y a un point de départ, une notion initiale, qui est également présente et à celui qui parle et à celui qui écoute, qui forme comme le lieu où les deux intelligences se rencontrent ; et une autre partie du discours, qui forme l'énonciation proprement dite. Cette division se retrouve dans presque tout ce que nous disons.* (1879/1991 : 29)

Il s'agit ainsi d'une conception fonctionnelle et non pas logique du langage où le rôle des interlocuteurs dans l'acte de communication est mis en avant. Cette conception est développée par l'Ecole de Prague sous le nom de *Functional Sentence Perspective* (FSP).

2.1.1 Délimitation d'un niveau informationnel

Dans le contexte du FSP, trois niveaux d'analyse phrastique sont identifiés : un niveau grammatical, un niveau sémantique et un niveau informationnel. Ce

[2] Si je reprends dans ce chapitre certaines idées déjà présentes dans ma thèse et dans mon ouvrage de 2000, les nouvelles orientations proposées ici doivent beaucoup aussi aux discussions avec Sophie Prévost et les autres membres d'opération SIOMST (Structure Informationnelle, Ordre des Mots et éléments Spatiaux et Temporels) de l'équipe Lattice pendant nos réunions de la période 2006-2008.

dernier niveau concerne la répartition de l'information sur la ligne de la phrase, avec sa division en parties thématiques et rhèmatiques.

Sans aborder pour l'instant en détail la définition des notions de thème et de rhème, il me semble important d'insister sur le fait que ce niveau d'analyse est différent de celui qui traite essentiellement des questions de syntaxe ou de sémantique. Il s'agit vraiment d'un autre niveau d'analyse avec sa structuration propre et ses règles de fonctionnement. Un des principaux reproches faits par les linguistes de l'École de Prague aux descriptions syntaxiques traditionnelles et générativistes, et plus précisément à N. Chomsky, portait sur cette confusion des trois niveaux. Selon Daneš :

> *Like many others, Chomsky, apparently does not respect the difference between the grammatical and semantic level in syntax.* (1964 : 227)

Il continue en disant, à propos du troisième niveau d'analyse :

> *…the organisation of the utterance disposes of special means of systemic character, which have been wrongly classed with grammar or stylistics.* (*Ibid.* : 228)

Le fait que d'autres linguistes aient choisi de dénommer cette troisième division différemment (C. Hagège – 1982 – parle par exemple d'un « point de vue énonciatif-hiérarchique ») ou que d'autres encore, notamment dans l'approche systémique, aient choisi de diviser cette composante informationnelle en deux parties (cf. *infra*), me paraît de moindre importance, par rapport à la reconnaissance même de l'existence d'une stratégie énonciative ou informationnelle de la part de l'énonciateur. En distinguant ce niveau « informationnel », les linguistes pragois ont construit une solide base de départ, non seulement pour approfondir les questions relatives aux analyses informationnelles, mais pour toute analyse textuelle.

Toutefois, la reconnaissance de la spécificité de ce niveau de structuration n'est pas forcément acquis dans toutes les sphères de la linguistique. Dans certains manuels de grammaire employés aujourd'hui dans les cours d'anglais langue étrangère, le niveau informationnel n'est souvent pas formellement explicité. Souvent, dans les explications sur l'utilisation du passif, comme par exemple

dans l'ouvrage de G.M. Spankie (1983), l'apprenant a l'impression que l'emploi d'un passif correspond surtout à une simple manipulation syntaxique (c'est le niveau syntaxique qui prime), ou bien à une construction qui permet de ne pas mentionner l'*agent* de l'opération prédicative (c'est le niveau sémantique qui prime). On ne dit pas, ou au moins explicitement, que l'utilisation première de la voix passive est de permettre une réorganisation du poids informationnel de la phrase. Ce troisième niveau d'analyse semble être camouflé, sinon perdu, dans les préoccupations concernant les niveaux syntaxique et sémantique.

2.1.2 – Noyau thème-rhème

Les définitions de thème et de rhème proposées par Mathesius[3] et reprises dans une certaine mesure par Daneš (1970 ; 1974) et Firbas (1964) recouvrent deux perspectives[4] : celle liée au concept de *« aboutness »* et celle liée au concept de « donné-nouveau ». Mathesius a défini le thème (*« theme »*) comme étant ce qui est connu et qui forme le point de départ pour l'émetteur :

> *…that which is known or at least obvious in the given situation, and from which the speaker proceeds.*

Et il définit le rhème (*« rheme »*) comme ce que l'émetteur affirme sur le thème :

> *…that which the speaker states about, in regard to, the theme of utterance* (Mathesius cité par Firbas, 1964).

Tout d'abord, en définissant ainsi le thème comme « le point de départ pour l'émetteur » et le rhème comme « ce qu'on en dit », les linguistes de Prague confèrent au couple thème-rhème un rapport relationnel direct

Cette première partie de la définition du couple thème-rhème, correspondant de près à celles d'un sujet et d'un prédicat psychologiques (Ducrot et Todorov 1972), trouve un large consensus dans la communauté linguistique. Cependant, même si un grand nombre de linguistes s'accordent sur une définition de ce type, le critère de *« aboutness »* reste assez flou. Il semble s'appliquer à la fois à la

[3] Les références aux travaux de Matheseius sont ceux rapportées par Daneš, (1966, 1970; 1974), Firbas (1964, 1992) et Sgall, (1974).
[4] C'est ce que Fries (1994) appelle « *the combined approach* ».

définition de *thème* et à celle de *rhème*. Le thème, c'est ce dont on parle (*« what we're speaking about »*) ; le rhème c'est ce que l'on dit du thème *(« what we're saying about the theme »*). Ceci m'amène à distinguer deux aspects de ce rapport relationnel. Le rhème semblerait se définir par rapport au thème de la phrase, donc à un niveau « local ». Le thème, en revanche, semblerait défini non seulement par rapport à la phrase, mais aussi par rapport au sujet global du texte (cf. 2.3.2 *infra* dans la section consacrée à l'approche de Lambrecht pour plus de discussions sur la notion de *« aboutness »*). La notion de « point de départ », telle qu'elle a été initialement formulée par Mathesius pose également des problèmes par rapport à la question de la linéarité. Le point de départ est-il positionnel, ou bien psychologique (cf. section 2.2.1) ?

C'est cependant le deuxième aspect de la définition, celle ayant trait à la distinction « donnée-nouvelle » qui est sans doute le plus problématique. Les définitions de thème et de rhème proposées par Mathesius consistaient à considérer les deux notions de thème et de rhème non seulement dans un sens relationnel, mais également par rapport au degré de nouveauté des éléments d'information transmis. Alors que le thème regroupait essentiellement les éléments « connus », le rhème était censé regrouper les données « nouvelles » :

> *An utterance may usually be divided into two portions: the theme conveying the known (given) elements, and the rheme conveying the unknown (not given) elements of the utterance. (Daneš 1964 : 228)*

Or cette deuxième conception du thème et du rhème a été très critiquée en raison surtout des trop grandes simplifications qu'elle entraînait. Que veut-on dire par informations « connues » ou « données ? Faut-il distinguer entre les informations contextuellement données et celles cognitivement connues au moment de l'énonciation (Chafe 1976) ? Il y a également une confusion quant à l'objet que l'on qualifie de donné ou de nouveau. S'agit-il de l'information elle-même ou du référent ? Très souvent dans l'approche pragoise, ce sont en fait les référents qu'on qualifie de donné ou de nouveau, et non pas la relation établie et il est donc difficile de parler dans ce cas en termes d'informativité *per se* (cf. l'approche de Lambrecht, section 2.3).

L'établissement d'un rapport direct entre le couple thème-rhème et la présence des informations « connues » ou « nouvelles » pose également problème au sein de l'Ecole pragoise elle-même. Pour F. Travnicek (1962), véhiculer des informations connues n'est pas la fonction essentielle du thème. Que faire dans le cas des énoncés du type existentiel ? Travnicek préfère considérer le thème comme l'élément de la phrase qui assure la liaison avec l'objet de la pensée (cf. la présentation de Combettes 1983). Firbas, pour sa part, refuse également de restreindre le thème à la notion de donné ou de connu. Firbas considère néanmoins l'approche de Travnicek comme trop psychologique et propose à la place d'introduire la notion de « dynamisme communicatif ».

2.1.3 – Dynamisme communicatif

L'approche proposée par Firbas (1964 ; 1992) est fondée sur l'idée d'un continuum informatif. Au lieu d'opérer une division plutôt dichotomique entre les notions de donné et de nouveau, il propose la notion de « dynamisme communicatif ». Chaque élément de la phrase diffère quant au degré avec lequel il contribue au développement de la communication :

> *By the degree of CD carried by the sentence element we understand the extent to which the sentence element contributes to the development of the communication, to which it pushes the communication forward, as it were.* (1964 : 270)

La répartition des différents degrés de dynamisme communicatif dans la phrase est influencée par trois facteurs : leur dépendance par rapport au contexte, leur structure sémantique et leur disposition linéaire sur la ligne de la phrase. Quand les critères relatifs à la dépendance par rapport au contexte et à la structure sémantique ne s'appliquent pas, on pose alors que le dynamisme communicatif augmente par degrés du début à la fin de la phrase. De cette manière, le problème des phrases dont aucun élément n'est « donné » ou « connu » est résolu. De telles phrases peuvent néanmoins présenter une structure thématique car, parmi tous les constituants de la phrase, certains ont moins d'importance que d'autres dans la progression de l'information :

> *It is obvious that elements conveying new unknown information show higher degrees of CD than elements conveying known information. But even within a sentence made up entirely of elements conveying new information, the degrees of CD are not the same (homogeneous).* (*Ibid.* : 270)

Dans cette perspective, il est ainsi tout à fait envisageable d'avoir un thème qui puisse se référer à un fait qui n'est pas contextuellement ou situationnellement « connu » par le récepteur. Cet élément dont l'émetteur a choisi de parler reste néanmoins un thème si le récepteur arrive à l'identifier et peut ainsi assimiler ce que l'émetteur dit à son propos.

Si Firbas, dans ses premières présentations du dynamisme communicatif, refuse de restreindre le thème à la notion de « connu », il semble néanmoins considérer la présence d'un thème dans chaque phrase comme obligatoire. Dans cette optique, même les énoncés existentiels auraient, ne serait-ce que par défaut, une partie thématique :

> *The theme is constituted by the sentence element (or elements) carrying the lowest degrees of CD within the sentence.* (*Ibid.* : 272)

Nous retrouvons les mêmes arguments chez Daneš qui considère la division bipartite de la phrase comme une norme :

> *Analysing the organisation of the utterance, we state as a rule its bipartite structure.* (1970 : 134)

Toutefois, quelques années plus tard, Firbas modifie cette position en déclarant que :

> *A sentence may be themeless, but it cannot be rhemeless [...] The Th(eme) may remain linguistically unexpressed.* (1992 : 92)

Une telle position me semble plus logique vis-à-vis des propositions présentatives ou existentielles comportant un sujet syntaxique sémantiquement vide. Dans la phrase suivante, par exemple, tirée d'un rapport technique :

> *(I) There are a lot of constraints of interference between received spots.*

Le présentatif *there*, tout en fournissant un ancrage pour l'expression des nouvelles informations, n'a pas d'identité propre. À mon avis, un tel élément sans référent ne peut fournir une base de départ pour un enchaînement phrastique ultérieur et ne joue donc pas de rôle actif dans la progression thématique. Le fait que, selon la norme écrite, l'expression d'un sujet

grammatical reste obligatoire ne change rien à la nature communicativement redondante d'un *there* ou d'un *it* impersonnel dans les énoncés existentiels[5]. Pour ces raisons je préfère ne pas conférer d'office de statut thématique à ces éléments impersonnels et je souscris à l'opinion de M.M.J. Fernandez lorsqu'elle écrit :

> Pour ce qui est de la structure de base de l'énoncé, un seul constituant, à fonction de rhème au niveau énonciatif, est nécessaire pour former un énoncé communicatif minimal. (1994 : 199)

2.1.4 – Du thème de la phrase au texte : progression thématique

Une des grandes contributions de l'École de Prague, et particulièrement de Daneš (1974), est d'avoir proposé une analyse de la structuration informationnelle au sein de séquences discursives plus larges. Daneš présente explicitement les enjeux (et les difficultés) de l'intégration de l'analyse structure informationnelle phrastique dans le cadre du texte. Selon lui, tout énonciateur se trouve confronté au choix du thème à prendre chaque fois pour base de l'énoncé suivant – un choix qui n'est pas effectué au hasard mais en fonction de l'information introduite dans le discours antérieur :

> *(…) the choice of distribution of themes in a in the text reveal a certain patterning;* (1974 : 09)
>
> *We must not be content with a statement that certain sentence elements convey the known information (in contrast to others conveying the new one), but we ought to find out the principles according to which this and not another portion of the mass of known information has been selected. In other words, we have to inquire into the principles underlying thematic choice and thematic progression.* (*Ibid.* :112)

Comment suivre l'évolution des thèmes et rhèmes et les enchaînements inter-phrastiques à un niveau plus global ? C'est ce que fait Daneš en cherchant à établir un nombre de schémas représentatifs de l'organisation thématique sur toute une portion du texte. Le terme de *progression thématique* désigne l'ensemble des relations thématiques dans le texte : la concaténation et la connexion des thèmes, leur ordre et la hiérarchie qui les unit, dans leurs relations aux paragraphes et à l'ensemble du texte ainsi qu'à la situation de

[5] Pour le rôle joué par les énoncés existentiels dans la structuration thématique, voir 2.2.3.

communication. Selon Daneš, les relations thématiques qui unissent les phrases successives dans la chaîne du texte contribuent à assurer la connexité de l'information dans le texte :

> *Our basic assumption is that text connexity is represented, inter alia, by thematic progression (TP).* (*Ibid.* : 114)

C'est le cadre qui permet à l'ensemble de prendre forme ou, pour reprendre Daneš, « *Thematic progression might be viewed as the skeleton of the plot* » (*ibid.*).

Les analyses de Daneš (1970, 1974, 1978) portant sur des textes d'ordre scientifique et professionnel en tchèque, allemand et anglais lui ont permis d'identifier trois grands types de progression thématique : la progression linéaire, la progression à thème constant et la progression à thèmes dérivés.

Dans la progression linéaire le rhème (Rh) d'une première phrase (ou une partie du rhème) devient le thème (Th) de la phrase suivante. Ainsi Rh1 devient le Th2, auquel est joint un nouveau rhème (Rh2), qui devient à son tour thème de la phrase suivante (Th3). Soit :

```
Th1→Rh1
    ↓
     →Th2 (Rh1)→Rh2
                ↓
                Th3 (Rh2)
```

Dans la formule « Th→Rh », l'ordre des symboles ne correspond pas forcément à l'ordre concret, réalisé dans le texte ; ce dernier dépend, comme Daneš le rappelle, des moyens mis en œuvre dans le cadre de la FSP. Toutefois, l'ordre canonique est thème avant rhème. Daneš illustre ce type d'enchaînement par l'exemple suivant :

> *The first of the antibiotics was discovered by Sir Alexander Flemming. <u>He</u> (Th2) was busy at the time investigating a certain species of germ <u>which</u> (Th3) is responsible for boils and other troubles.* (Daneš 1974 : 118)

Le deuxième type de progression est caractérisé par la présence d'un thème constant, auquel s'adjoignent différents rhèmes successifs :

Th1 → Rh1

Th1 → Rh2

Th1 → Rh3

The Rousseauist (Th1) especially feels an inner kinship with Prometheus and other Titans. He (Th2) is facilitated by any forms of insurgency… He (Th3) must show an elementary energy in his explosion against the established order and at the same time a boundless sympathy for the victims of it. Further the Rousseauist (Th4) is ever ready to discover beauty of soul in anyone who is under the reprobation of society. (*Ibid.* : 119)

Ici encore, c'est la continuité d'un même thème de phrase en phrase qui est primordiale, et non pas la position en début de phrase. Les thèmes eux aussi ne seront pas nécessairement exprimés de façon strictement identique – différents moyens linguistiques (grammaticaux et lexicaux) existent qui permettent de reprendre un même thème.

Dans la progression à thème dérivé, les thèmes de chacune des phrases sont issus d'un hyper-thème ou d'un hyper-rhème. Dans l'exemple donné par Daneš, c'est un hyper-thème, qui pourrait être considéré comme le titre du paragraphe, qui fournit les thèmes des différentes propositions successives. Il s'agit en l'occurrence de l'énumération des différentes données géographiques d'un pays (New Jersey) qui est ainsi l'hyper-thème :

T (hyper-thème)

Th1→ Rh1↗ ↑

Th2→Rh2 ↖

Th3→Rh3

New Jersey (T) is flat along the coast and southern portion ; the north-western region (Th1) is mountaineous. The coastal climate (Th2) is mild (…). Summers (Th3) are fairly hot. (*Ibid.* : 120)

Selon Daneš, ces trois principaux types de progression se combinent d'ordinaire dans les textes. Il arrive fréquemment que les schémas « 1 » et « 2 » s'enchaînent. Daneš envisage également des sous-divisions de ces types principaux déjà énumérés. Dans le cas de la progression linéaire, par exemple, la reprise thématique peut ne s'appliquer dans une première instance qu'à une

partie du rhème, ce que Daneš appelle *« split rheme ».* Il illustre ce phénomène avec l'exemple suivant :

> All substances can be divided into two classes: elementary substances (*Rh1*) and compounds (*Rh2*).
>
> An elementary substance is a substance *which consists of atoms of only one kind.*
>
> A compound is a substance *which consists of atoms of two or more different kinds*. (*Ibid.* : 121)

Ici, le rhème de la première phrase contient deux parties. Les nominaux *elementary substance* et *compounds* deviennent tous les deux les thèmes d'une phrase séparée.

Comme Daneš le reconnaît lui-même, il peut y avoir non seulement des croisements mais des ruptures dans le système qu'il propose. Il peut s'agir de rajouts, de notes explicatives ou même d'omissions. Dans le cas des omissions, on peut parler des liens omis, ou bien d'un *« thematic jump »* ou d'une rupture thématique. En faisant état de cette possibilité d'un *saut thématique,* la représentation de Daneš, qui pourrait apparaître comme essentiellement théorique, commence à prendre une tournure plus pratique. L'absence de lien entre des phrases existe dans le monde réel. Certaines phrases, tout en comportant une division en parties thématiques et rhématiques, ne contiennent pas d'élément « connu », déjà mentionné ou suggéré dans le texte. De plus, même si l'élément a été déjà mentionné dans le texte, il ne s'agit pas forcément de la reprise d'un élément contigu. Certaines étapes sont sautées. On passe par exemple du rhème d'une seconde phrase au thème d'une quatrième, mais le lecteur est censé pouvoir rétablir la chaîne complète. Le thème initial, pour diverses raisons, ne sera pas forcément lié *directement* aux éléments précédents.

Dans cette première présentation de 1974 de la progression thématique, Daneš insiste sur le fait que l'étude de la structure thématique d'un texte ne permet pas uniquement d'obtenir des modèles descriptifs mais peut avoir des répercussions pratiques, non seulement dans le domaine de la didactique mais aussi dans le traitement automatique des textes. Il attire notre attention sur le

fait que la langue met à la disposition des rédacteurs des expressions spécifiquement destinées à illustrer un certain type de progression thématique. Dans le cas du *« split rheme »* (progression à rhème éclaté), Daneš signale l'existence (en anglais) de diverses tournures qui permettent de conférer le statut d'équivalence à deux ou plusieurs éléments de la phrase tels que :

both and,

on the one hand on the other hand,

in the first instance in the second instance, etc. (*ibid.* : 121)

Si je persiste à faire une présentation détaillée de la notion de progression thématique, c'est parce que le concept reste selon moi essentiel à une compréhension de l'organisation du texte. Si les travaux de Daneš sur ce sujet ne sont pas très connus, nous pouvons attribuer cette ignorance à une difficulté d'accès (jusqu'à récemment en tout cas) aux travaux des linguistes tchèques, ou bien peut-être à une question de mode. En tout cas, depuis les années 1970, les travaux de Daneš font référence en la matière pour les initiés et à ma connaissance aucune autre théorie complète n'a été élaborée qui permettrait d'éclipser leur importance dans ce domaine.

Les idées de Daneš m'apparaissent toujours d'actualité aujourd'hui. Au lieu de se focaliser uniquement sur la division thème-rhème dans les phrases isolées, la notion de progression thématique permet de rendre compte des enchaînements thématiques interphrastiques et fournit un outil d'analyse puissant pour aborder la cohérence textuelle sur des portions de texte assez étendues. Le modèle de Daneš fournit des pistes d'application diverses. Les travaux qui ont été faits dans le sillage des recherches pragoises ont en effet démontré l'importance de la notion de progression thématique dans l'enseignement (de la langue maternelle ou des langues étrangères) [THOM 98 ; THOM 99b et c ; THOM 00 ; THOM 04b] (Combettes et Tomassone 1988 ; Downing 2001) ; en traductologie, dans le milieu de l'édition scientifique (Ventola 1994, Ventola et Mauranen 1991) ainsi que dans la typologie textuelle

[THOM 99a] (McCabe 1999 ; Nwogu 1990). Certaines de ces applications potentielles seront décrites dans le chapitre 7 de cet ouvrage.

2.2 L'approche de Halliday

2.2.1 – Séparation de thème et « donné »

La position adoptée par Halliday (1967, 1970, 1994) et les systémistes se rapproche dans un sens des théories élaborées par Firbas dans la mesure où ces auteurs soulignent le fait que les notions de donné/nouveau ne coïncident pas obligatoirement avec le couple thème-rhème. Cependant, Halliday va plus loin. Il considère que, même si dans de nombreux cas il y a coïncidence entre les deux, il s'agit néanmoins de deux plans de réalité différents.

Le niveau thématique, *Thematic structure*, présente l'information du point de vue de l'émetteur. Le thème est ce qui est choisi par l'émetteur comme point de départ du message. Cette fonction est réalisée en anglais (comme en français) par sa position en début de phrase (cf. *infra*). Le niveau informationnel, *Information structure*, organise les informations en fonction du récepteur, en fonction de ce qui est connu ou non de lui. Le connu est ce que le locuteur présente comme étant déjà disponible ou « récupérable » chez son interlocuteur. Le nouveau est ce qu'il présente comme n'étant pas disponible, ou ce qui est « non récupérable » (Halliday 1994 : 299-302 ; Butler 2005). On voit ainsi que ce n'est pas la nouveauté contextuelle qui est déterminante. Le « *given* » ne correspond pas forcément aux informations déjà mentionnées, mais aussi aux informations disponibles dans la situation d'énonciation et/ou qui font partie du « terrain commun » des interlocuteurs (Banks 2005). Le « *new* », de même, est dit « focal » non pas forcément parce qu'il n'a pas été mentionné auparavant (bien que ce soit souvent le cas), mais parce que le locuteur le présente comme étant une information nouvelle : soit parce qu'il souhaite la contraster avec d'autres informations, soit tout simplement pour attirer

l'attention du récepteur sur quelque chose de déjà connu [THOM 00, 02 & 09a] (Banks 2005).

La séparation entre les deux niveaux d'organisation textuelle est cependant ce qui distingue principalement l'approche de Halliday de celle des linguistes de l'École de Prague. Pour Halliday, l'organisation textuelle repose sur deux structures distinctes :

> *The two are independently variable, and derive from different sources; given-new is a discourse feature, while theme-rheme is not. The difference may be summed up in the observation that, in dialogue, "given" means 'what you were talking about' while "theme" means 'what I am talking about'; and, as is well known, the two do not necessarily coincide.* (1976 : 179-180)

Ce que P.H. Fries (1983) a appelé « *the separating approach* » de Halliday, par rapport à *« the combined approach »* de l'Ecole de Prague », a suscité beaucoup de controverses. Si le thème devait effectivement être, comme Halliday le prétend, *« independent of what has gone before »* (ibid. : 178), l'organisation thématique semblerait être sans rapport avec l'organisation du texte. Or, comme Daneš l'explique, une telle conclusion paraît peu probable, car le choix du thème n'est pas un choix fortuit, sans rapport structurel avec le texte :

> *Such a conclusion appears very doubtful in the light of the act that the choices of themes of particular utterances can hardly be fortuitous, unmotivated and without any structural connection to the text.* (1974 : 109)

Pour Daneš, tout énonciateur se trouve confronté à la question du thème à prendre chaque fois comme base de l'énoncé suivant.

Une autre particularité de la présentation de Halliday à propos de la distinction établie entre ces deux plans d'organisation textuelle concerne les unités auxquelles ces notions s'appliquent. Le niveau thématique s'applique à une proposition grammaticale. En revanche, le niveau informationnel s'applique à une unité d'intonation ou à un groupe de souffle[6] (et cette « unité » ne correspond pas forcément à une division grammaticale). Dans ce système, les

[6] *Tone unit.*

informations nouvelles sont identifiées par la prosodie, avec le focus des informations nouvelles étant signalé par une proéminence intonative (*« tonic prominence »*).

> *Information focus reflects the speaker's decision as to where the main burden of the message lies. (…) Information focus is one kind of emphasis, that whereby the speaker marks out a part (which may be the whole) of a message block one as that which he wishes to be interpreted as informative.* (Halliday 1967 : 204)

Du fait que la division *donné/nouveau* de Halliday reposait sur des effets prosodiques (ce qui est « nouveau » dans la phrase serait plus accentué), cette notion n'est pas facilement transférable au texte écrit. Dans la langue écrite, c'est le lecteur qui doit reconstituer mentalement la position hypothétique de l'accent tonique et, à partir de cela, la structure informationnelle. Pour le linguiste travaillant sur les textes écrits, l'accès à ces informations est donc problématique. L'analyste peut demander à des volontaires de lire les textes à analyser à haute voix (cf. Banks 1999) [THOM 02, THOM 09a], mais de telles solutions restent compliquées à mettre en oeuvre. Étant entendu qu'il existe une certaine flexibilité des interprétations selon les lecteurs, la fiabilité des résultats est aussi très difficile à assurer. Si le lecteur n'a que peu de familiarité avec le domaine concerné (dans le cas par exemple d'un article de recherche très spécialisé), il risque d'interpréter le texte avec plus d'informations focalisées que le spécialiste du domaine.

Dans toute réflexion sur la structure informationnelle, il me semble en effet réducteur de limiter l'analyse aux seules marques prosodiques. La formulation spécifique choisie – les choix lexico-grammaticaux, les configurations syntaxiques, l'ordre des mots – permettent aussi d'acquérir des informations précises sur la structuration informationnelle des textes.

Comme l'admettent les linguistes systémiques T. et M. Bloor, il est aussi très difficile dans la langue écrite de séparer le plan thématique du plan informationnel :

> *In written English prose, it is more difficult to vary the relationship of Theme and Given. In the vast majority of English written sentences, the two functions are realized by the same*

constituent, and similarly the Rheme and the New information are realized by the rest of the clause. (1995 : 80)

Ces remarques trouvent un écho dans les commentaires de Fries. Dans un article (1994) où il s'intéresse surtout à la nature de la partie rhématique de la phrase, il étudie les moyens disponibles pour signaler qu'une information est « nouvelle » – ou *newsworthy* (« digne d'attention »), pour reprendre son expression – **dans la langue écrite**. Il est possible, constate-t-il, de la souligner ou de l'écrire en gras (voir ci-dessus !), mais le moyen le plus courant est la manipulation de l'ordre des mots dans la phrase. Selon Fries, l'information nouvelle est généralement placée à l'endroit où l'accent tonique serait apparu dans l'énoncé oral, c'est-à-dire vers la fin de la phrase. Quant au thème, Fries le considère comme moins important au niveau informationnel (*« less informationally important »*), sa fonction principale étant de permettre au récepteur de s'orienter vers le reste du message :

Writers use position at the end of the clause to indicate the newsworthy information to their readers, and […] the beginnings of their clauses to orient their readers to the message that will come in the rest of the clause. (Fries, 1994 : 234)

Ses commentaires renforcent ainsi la corrélation entre les éléments rhématiques et les éléments nouveaux d'une part, et les éléments thématiques et connus d'autre part.

2.2.2 Critère de position

La question d'une association entre la fonction de thème et la position phrastique est souvent considérée comme caractéristique de l'approche systémique. Il est vrai que Halliday compte parmi les auteurs qui accordent un statut particulier aux constituants initiaux de la phrase. Il est important néanmoins de souligner qu'il ne présente pas la position initiale comme un critère définitionnel. Sa définition de *thème* est fonctionnelle :

The theme is the element is which serves as the point of departures of the message, the part in which the clause is concerned. The remainder of the message, the part in which the Theme is developed is called in Prague school terminology the rheme. (1994 : 37)

Dans le cas particulier de l'anglais (et par extension dans d'autres langues de la famille indo-européenne), la fonction de thème est indiquée par la position initiale. Dans d'autres langues, comme par exemple le japonais (cf. Klingler 2003), le thème peut être signalé par un morphème spécial.

> *In some languages (…), the theme is announced but means of a particle (…). In other languages of which English is one, the theme is indicated by position in a clause. In speaking or writing English we signal that an item has thematic status by putting it first.* (*Ibid*. : 37)

Toutefois, même si Halliday ne prétend pas que le thème est universellement porté par les constituants initiaux, sa définition du thème en tant que *« point of departure »* reste ambiguë. Dans de nombreux travaux systémiques entrepris dans la continuation de ceux de Halliday, le terme *« theme »* reçoit une interprétation explicitement positionnelle (Banks 1999 ; 2002 ; 2005 ; Bloor et Bloor 1995 ; Thompson 1996).

2.2.3 Thèmes marqués

Selon Halliday, tout constituant principal est susceptible de jouer le rôle de thème. Dans la phrase déclarative en anglais, le thème sera très souvent réalisé par un sujet sur le plan syntaxique. Dans ce cas, Halliday parle d'un thème « non marqué » :

Cuthbert	*read his breviary in the monastery garden*
Thème non marqué	Rhème

Quand un autre constituant majeur – le complément du verbe ou un groupe circonstanciel – précède le sujet, il est considéré comme un thème marqué :

In the monastery garden,	*Cuthbert read his breviary*
Thème marqué	Rhème

His breviary	*Cuthbert read in the monastery garden*[7]
Thème marqué	Rhème

[7] Ces trois exemples viennent de Banks (2005).

On constate que dans ces deux derniers exemples comportant un thème marqué, le sujet syntaxique (« Cuthbert ») fait partie du rhème. En effet, quand plusieurs éléments à portée « idéationnelle » apparaissent en position initiale, seul le premier sera considéré comme thème (cf. *infra*).

Si la distinction établie entre thème marqué et thème non marqué semble importante, j'estime cependant qu'il est difficile de considérer les différents types de thèmes comme étant tous « marqués » sur le même plan. Halliday lui-même suggère l'existence d'une échelle de thématicité. Plus un élément est susceptible de figurer en thème, moins il est marqué. Dans le cas d'un objet antéposé, son statut en tant que thème marqué semble très clair :

> *We are aware of our responsability to our cities. We are aware of our responsability to the author (…). <u>This responsibility</u> we accept wholly.* (Halliday 1994 : 45)

Dans cet exemple cité par Halliday, la thématisation du complément dans la troisième proposition apparaît comme un choix délibéré et fort de la part du locuteur. En revanche, l'analyse semble beaucoup moins évidente lorsqu'il s'agit d'un thème marqué composé d'un adverbe ou groupe prépositionnel circonstanciel suivi d'un GN sujet – une configuration relativement fréquente en anglais comme en français. Dans l'exemple suivant :

> *<u>Yesterday,</u> John left before lunch.* (Fries 1995 :14)

Il est difficile d'envisager un contexte où le circonstant temporel serait considéré comme l'unique point de départ de la proposition, et *« that with which the clause is concerned »* (Halliday 1994 : 37) ou le topique sémantique. En revanche, il est plus facile d'imaginer un contexte où cette fonction serait réalisée par le groupe adverbial et le sujet syntaxique ensemble : la proposition concerne la description des activités du protagoniste « John » pendant la journée d'hier. Pour cette raison, M. Berry (1996) propose par exemple d'étendre le groupe thématique jusqu'au verbe afin d'inclure également dans ce cas le GN sujet. Je reviendrai sur ce problème du rapport entre un élément circonstanciel initial et le sujet syntaxique dans la section consacrée aux adverbiaux cadratifs (3.2.2).

2.2.4 Thèmes composites

Au-delà de la définition thème marqué/thème non marqué, Halliday distingue différents types de composantes thématiques. Si le thème va obligatoirement jusqu'au premier constituant majeur – actant, procès, ou circonstance – le potentiel thématique n'est pas épuisé pour autant. D'autres composants thématiques (ceux-ci facultatifs) peuvent apparaître avant le thème principal. L'identité de ces différents composants est étroitement imbriquée avec le système de « métafonctions » (une des notions clés du modèle systémique) : les métafonctions idéationnelles, interpersonnelles et textuelles (Halliday 1994 : 35 ; Banks 2005 : 37).

À partir de ces métafonctions, Halliday distingue trois types de composantes thématiques : thèmes idéationnels, thèmes interpersonnels et thèmes textuels (Halliday 1994 : 52-54). Les thèmes idéationnels correspondent aux constituants principaux de la proposition (sujet, complément ou circonstance), examinés ci-dessus (2.2.3). Ce type de thème est obligatoire. Le premier type de thème facultatif, le thème interpersonnel, donne des informations sur la relation entre le locuteur et le récepteur ou sur l'attitude du locuteur envers le contenu de sa proposition. Le thème interpersonnel est réalisé par des éléments modaux (*peut-être*, *certainement*, *sans doute* etc.) ou des vocatifs. Le deuxième type de thème non obligatoire est le thème textuel. Les thèmes textuels créent des liens (explicites) avec le co-texte. Il s'agit de marques de conjonction (*cependant, néanmoins, de plus*) et dans le discours oral des éléments de liaison discursive (*Eh bien*, *donc* etc.).

Il est possible de trouver des thèmes interpersonnels et des thèmes textuels dans la même proposition. Dans ce cas, l'ordre normal d'apparition de ces différents types de thèmes est le suivant : thème textuel, thème interpersonnel, thème idéationnel (ou thème « topical »[8]) :

[8] Les termes de thème « topical » et de thème « idéationnel » sont employés tous les deux dans la tradition systémique pour décrire l'élément obligatoire du matériel thématique.

Vraiment,	*c'*	*est une fille bizarre*[9]
Thème interpersonnel	Thème topical	Rhème

mais,	*probablement*	*toutes les filles*	*ont des idées pas ordinaires*
Thème textuel	Thème interpersonnel	Thème topical	Rhème

Il est également possible d'avoir deux ou plusieurs thèmes textuels (ou interpersonnels) au sein de la même proposition, à condition qu'ils se situent à gauche du thème idéationnel. Le fait que ces différents types de thèmes apparaissent souvent à l'intérieur d'une même phrase conduit Halliday à proposer la notion de « *multiple theme* » (1994 : 52). En revanche, il n'y peut y avoir qu'un seul thème idéationnel. Le thème idéationnel marque la fin de la zone thématique :

> *(…) the Theme extends from the beginning of the clause up to (and including) the first element that has a function in transitivity. (…) the theme of a clause consists of the topical Theme together with anything else that comes before it.* (Halliday 1994 : 53)

Le fait que Halliday reconnaît explicitement la nature potentiellement composite du groupe thématique me parait très important. Mes analyses sur corpus m'ont appris que la partie thématique d'une phrase est souvent très complexe et qu'il est effet difficile en conséquence de réserver l'étiquette de thème à un seul constituant phrastique. En revanche, la décision de Halliday de restreindre la portée thématique au seul premier élément « idéationnel » me paraît problématique. Cet aspect de l'approche de Halliday a d'ailleurs fait l'objet de beaucoup de controverses au sein même de la communauté systémique (Downing 1991 ; Berry 1996 ; Mauranen 1996 ; Taboada 1995). Comme on l'a vu ci-dessus avec l'exemple de Fries (*Yesterday, John left before lunch)*, si nous voulons appliquer une définition sémantique du thème[10], il est difficile de le restreindre au groupe adverbial initial en isolation. Est-ce que cela signifie

[9] Ces exemples viennent de Caffarel (2000).

[10] On note toutefois que certains chercheurs dans la tradition systémique refusent toute allusion à la notion d' « à-propos » dans leur utilisation du concept de thème (Taboada 1995 ; Banks 2005). Pour citer Banks : « On a pu, par le passé, prétendre que le thème était ce dont on parlait dans la proposition, mais cette description s'avère ambiguë, et il est préférable de le considérer strictement comme le point de départ de l'énoncé »(2005 : 69).

qu'il faudrait accepter la possibilité d'avoir deux thèmes idéationnels dans la même proposition : un thème marqué (groupe circonstanciel) et un thème non marqué (sujet) ? Afin de répondre à cette question avec précision, il faudrait un recours au contexte complet (autrement dit, une vraie analyse textuelle). Si un circonstant en position initiale peut sous certaines conditions remplir à lui seul une fonction topicale (cf. Charolles 2003 et section 3.2 *infra*), ceci est loin d'être toujours le cas. À mon point de vue, la zone thématique peut souvent s'étendre par défaut à toute la zone préverbale comme Berry (1996) le propose. Je reviendrai sur cette épineuse question de l'étendue de la zone thématique dans le chapitre 3.

2.3 L'approche de Lambrecht

2.3.1 Distinction référents du discours et apport « informatif »

Un reproche souvent fait à la présentation de la structure informationnelle par l'École de Prague porte sur un certain flou dans le traitement des notions d'*information* donnée et d'*information* nouvelle (cf. 2.1.2 *supra*). Le thème est souvent présenté comme véhiculant des *informations* données et le rhème des *informations* nouvelles. Or comme Lambrecht l'explique, une telle présentation est trompeuse, voire erronée, puisque l'information en tant que concept ne peut pas être segmentée et répartie parmi les constituants de l'énoncé :

> *(…) the information conveyed by a proposition cannot be factored out and matched with individual sentence constituents. In particular, the difference between old and new information cannot be equated with the difference between "old" and "new" referents.* (1994 : 49)

Autrement dit, il y a confusion entre le statut de l'information et le statut des référents dans l'univers du discours. A ma connaissance, Lambrecht est le premier linguiste à établir cette distinction explicite entre le statut cognitif des référents et les relations qu'ils entretiennent avec les structures dans lesquelles ils sont insérés au sein de la proposition. C'est la mise en rapport de l'information « connue » (appelée dans son approche la *présupposition*) avec l'information nouvelle (appelée l'*assertion*) qui devient « informative ».

Toutefois, l'indépendance de ces deux domaines n'est pas totale. Comme le topique constitue la préoccupation courante (*« matter of current concern »*), le référent choisi en tant que topique doit appartenir à la présupposition pragmatique. En règle générale, on ne peut pas parler à propos d'un référent si ce référent n'est pas déjà disponible et identifiable dans l'univers du discours :

> *(…) for a proposition to be construable as about a topic referent this referent must be part of the pragmatic presupposion, i.e. it already be under discussion or otherwise available from the context.* (*Ibid.* : 150)

En revanche, l'inverse n'est pas vrai. Ce n'est pas parce qu'un réfèrent est disponible qu'il sera forcément choisi comme topique.

Lambrecht propose également un traitement fouillé des questions relatives au statut cognitif des référents. En s'inspirant de W.L. Chafe (1976 ; 1987) il distingue d'abord entre deux critères, celui d'identifiabilité et celui d'activation. Dans cette perspective, les connaissances prêtées au récepteur ne sont donc pas déterminantes. Ce qui importe est le statut du référent dans la conscience du récepteur au moment de la réception du message. Comme l'explique Lambrecht :

> *Knowing something and thinking of something are different mental states. (…), the conveying of information in natural language on only involves knowledge but also consciousness.* (*Ibid.* : 93)

Un référent sera considéré comme identifiable dès lors qu'il en existe une représentation dans l'esprit du récepteur au moment de l'énonciation. En revanche, un référent sera considéré comme non identifiable tant que cette représentation n'existe pas chez le récepteur. Pour préciser la situation des référents non identifiables, Lambrecht adopte la notion d'ancrage de E.F. Prince (1981). Un référent apparaîtra soit comme totalement nouveau *(« brand new anchored »*) mais ancré dans la mesure où il est mentionné par rapport à un autre référent du discours identifiable, soit comme non ancré *(« brand new unanchored »*) quand il n'existe aucun lien avec les autres référents du discours.

Le critère d'activation permet de préciser le statut précis des référents identifiables : inactif, accessible (ou semi-actif) ou actif. Dans l'exemple suivant de Chafe (1976 : 30) : « *I saw your father yesterday »*, *your father* serait

analysé comme « inactif », car l'émetteur a jugé que le destinataire ne pensait pas à son père au moment de l'énonciation. En revanche, le fait qu'un référent soit actif signifie qu'il est présent et activé dans l'esprit des interlocuteurs et se traduit en général par la possibilité de reprise pronominale. Le statut intermédiaire de semi-actif (ou accessible) signifie que le référent est dans la conscience périphérique ou en arrière-plan. Trois sous-catégories sont distinguées : accessible textuellement (il s'agit d'un référent déjà utilisé qu'il faut réactiver), accessible situationnellement et accessible par inférence.

Lambrecht propose une échelle de correspondance, associant en général ces différents degrés d'activation de référents à leur propension à fonctionner en tant que topiques (fig. 1) :

active	*most acceptable*
accessible	
unused	↓
brand-new anchored	
brand-new unachored	*least acceptable*

Fig. 1. *The Topic acceptablity scale* (1994 :109).

En choisissant un topique, l'émetteur doit faire quelques hypothèses quant au degré d'accessibilité de ce référent pour son interlocuteur. Au niveau de l'interprétation, plus un référent est actif, moins la charge cognitive est lourde pour le récepteur. Les référents qui sont le plus facilement traités par les interlocuteurs sont ceux qui figurent en haut du classement, c'est-à-dire ceux qui sont les plus « actifs » dans le discours. En français, comme en anglais, il s'agit souvent de pronoms et de groupes nominaux précédés d'un article défini, qui se réfèrent donc aux entités qui sont déjà « activées » dans le discours. En revanche, un référent non indentifiable sera difficile à interpréter comme topique et l'énoncé risque d'apparaître comme mal formé :

> *From the requirements that topic referents must have a degree of pragmatic accessibility it follows that sentences with insufficiently accessible topic referents must pose certain difficulties of interpretation, hence will tend to be perceived as ill-formed.* (*Ibid.* : 164)

Dans mes travaux portant sur l'évaluation de la cohésion textuelle dans les devoirs écrits en anglais par des étudiants francophones, j'ai constaté que bon nombre de maladresses et de problèmes d'interprétation sont liés aux difficultés d'identification des thèmes phrastiques [THOM 98, THOM 99c, THOM 00]. Souvent, les référents choisis comme thèmes par les apprenants ne sont pas suffisamment actifs dans le monde du discours. Soit l'idée que se fait l'émetteur des connaissances de son interlocuteur est erronée, soit l'émetteur oublie de fournir certaines informations. Le résultat est le même : le lecteur n'arrive pas à trouver un référent à sa portée (dans le texte ou en dehors du texte) qui lui permette d'arriver à une interprétation logique.

2.3.2 Un topique pragmatique

Chez Lambrecht, le topique[11] est surtout un topique pragmatique. Il le définit en fonction du critère de « ce dont on parle » :

> *The topic of a sentence is the thing which the proposition expressed by the sentence is about.* (1994 : 118)

Même si cette définition classique du topique reste très apparentée à celle du sujet d'Aristote, pour Lambrecht il n'existe pas de correspondance directe entre le topique et le sujet grammatical de la syntaxe moderne. Le topique ne sera pas forcément un sujet syntaxique et les sujets syntaxiques ne sont pas forcément topiques. Des constituants autres que le sujet peuvent jouer le rôle du topique (« le chocolat, je l'adore ») et les sujets sur le plan grammatical ne sont pas forcément des topiques. Dans les énoncés thétiques ou existentiels du type : « Il y a trois types de produits », le sujet grammatical n'est pas topique. Lambrecht refuse aussi d'établir une association systématique entre le topique et les constituants initiaux de la phrase. Sa définition du topique est avant tout pragmatique : *« Topic is a pragmatically construed sentence relation »* (*ibid :* 127).

[11] Lambrecht emploie le terme *topique* et non pas *thème*. Cependant, il ne fait pas de distinction entres les deux termes.

Un autre élément central à la définition du topique proposée par Lambrecht est sa nature nécessairement référentielle. Selon lui, le topique doit exister dans l'univers du discours. Il prend bien soin de distinguer ici entre le topique et son expression linguistique. Pour lui, le topique ne sera pas forcément codé par une *topic expression.* Il peut être implicite. Il définit les deux termes de la manière suivante :

> *TOPIC: a referent is interpreted as the topic of a proposition if in a given situation the proposition is construed as being about this referent, i.e. as expressing information which is relevant to and which increases the addressee's knowledge of this referent*
>
> *TOPIC EXPRESSION: a constituent is a topic expression if the proposition expressed by the clause with which it is associated is pragmatically construed as being about the referent of this constituent.* (Lambrecht 1994 : 131)

Cette reconnaissance de la nature parfois implicite du topique était également présente chez Firbas (section 2.1.3), et me semble être un élément essentiel pour l'application de la structure informationnelle aux textes réels.

La conception du topique proposée par Lambrecht s'étend aussi aux topiques scéniques (*« scene-setting »*) ou aux *« Chinese style topics »* (Li & Thompson 1976). Il s'agit ainsi des SN détachés à gauche – essentiellement des adverbiaux spatio-temporels ou de groupes prépositionnels adverbaux. À la différence de SC Dik (1997), qui établit une distinction entre les thèmes (les éléments détachés à gauche) et les topiques (un constituant de la prédication propre), Lambrecht assimile les deux sous le terme de *topique*.

2.3.3 Taxinomie de structures focales

Le focus, par des aspects fondamentaux, peut être considéré comme le complément du topique, tout comme le rhème est le complément du thème dans l'approche pragoise. Toutefois, dans l'approche de Lambrecht, le focus n'est pas toujours ni seulement le complément d'un topique. Tout d'abord, si le topique peut rester implicite (cf. *supra*), le focus est toujours exprimé. Il n'est en revanche pas nécessaire que le référent du focus (à l'inverse du topique) existe dans l'univers du discours. Dans le système de Lambrecht, le focus reçoit une définition indépendante.

The focus of the proposition expressed by a sentence in a given utterance contest is seen as the element of information whereby the presupposition and the assertion differ from each other. The focus is that portion of a proposition which cannot be taken for granted at the time of speech. It is the unpredictable or pragmatically non-recoverable elements in an utterance. The focus is what makes an utterance into an assertion. (1994 : 211)

Autrement dit, il s'agit d'un élément d'information par lequel l'assertion diffère de la présupposition. Le focus n'est pas nécessairement un *denotatum* (ou référent) nouveau, mais peut signaler uniquement une relation nouvelle entre le *denotatum* et la proposition (*ibid.* : 217). Il est intéressant de noter à cet effet que Lamabrecht utilise le terme de *non-recoverable*, tout comme Halliday (cf. *supra*) afin d'écarter toute assimilation abusive entre nouveau et focus.

Lambrecht (*ibid.* : 223) distingue trois principaux types de structures morpho-syntaxiques permettant d'actualiser la structure informationnelle, à savoir : focalisation du prédicat, focalisation d'un argument et focalisation d'une phrase. Il illustre ces trois types avec des exemples en quatre langues (anglais, italien, français et japonais) :

1. PREDICATE-FOCUS STRUCTURE

What happened to your car?

a. My car/It broke DOWN.

b. (La mia macchina) si è ROTTA.

c. (Ma voiture) elle est en PANNE.

d. (Kuruma wa) KOSHOO-shi-da.

2. ARGUMENT-FOCUS STRUCTURE

I heard your motorcycle broke down?

a. My CAR broke down.

b. Si è rotta la mia MACCHINA. / E la mia MACCHINA che si è rotta.

c. C'est ma VOITURE qui est en panne.

d. KURUMA ga koshoo-shi-da.

3. SENTENCE-FOCUS STRUCTURE

What happened?

a. My CAR BROKE DOWN.

b. Mi si è rotta (ROTTA) la MACCHINA.

c. J'ai ma VOITURE qui est en PANNE.

d. KURUMA ga KOSHOO-shi-da.k.

C'est seulement dans le premier type de structure (Focus sur prédicat) que nous retrouvons la confrontation classique Topique/Comment. On affirme quelque chose à propos d'un référent de discours qu'on présume topical *(car, voiture, macchina)*. Dans la structure « Focus sur argument », il s'agit d'un focus étroit qui permet d'identifier un argument par rapport à une proposition évoquée précédemment (« le *quelque chose* est en panne »). Tandis que dans le troisième type de structure, le focus porte sur la phrase entière.

Cette classification de Lambrecht permet de faire des comparaisons intéressantes entre les manières dont ces structures focales sont réalisées à travers différentes langues. À titre d'illustration, dans les exemples d'« Argument Focus » présentés ci-dessus, nous pouvons remarquer un recours à la prosodie plus important en anglais que dans la phrase française. Tandis que le terme focalisé *CAR* est uniquement marqué par la prosodie, le terme *voiture* est marqué au moyen d'un procédé syntaxique spécifique : la phrase clivée. Je reviendrai dans le chapitre 4 sur les emplois des phrases clivées et sur une étude que j'ai consacrée aux différences dans l'emploi des clivées en anglais et français [THOM 09a].

2.4 Pour une approche opérationnelle

Dans ma thèse de 1998 et les travaux entrepris dans son sillage, j'ai adopté la description suivante du couple thème-rhème :

> …la description de thème et de rhème sera faite dans une perspective double :
>
> I Thème et rhème seront considérés dans un rapport relationnel : le thème correspond à ce dont on parle, et le rhème à ce qu'on en dit.
>
> II Thème et rhème seront considérés par rapport au statut du référent : le rhème est généralement « nouveau » et le thème est généralement « connu ».
>
> Nous ajouterons deux explications supplémentaires à ce dernier critère :
>
> – si le rhème n'est pas « nouveau » pour le locuteur (il peut s'agir d'un élément déjà mentionné dans le texte ou familier au récepteur), son rapport au thème reste néanmoins nouveau ;
>
> – certaines propositions seront considérées comme étant entièrement rhématiques (véhiculant seulement des informations nouvelles). (2000 : 71)

Ces indications, essentiellement inspirées de l'approche pragoise, étaient proposées dans un but surtout *applicatif,* et comprennent un certain nombre de simplifications. Il s'avère souvent nécessaire de faire des généralisations afin de proposer un modèle opérationnel. J'ai souhaité établir un cadre d'analyse solide pour l'étude de certaines séquences de progression thématiques dans les travaux écrits de mes étudiants.

Dix années plus tard, mon désir de focaliser sur le niveau applicatif de la structure informationnelle reste intact. Je souhaiterais néanmoins un peu affiner mes propos par rapport à certains aspects de ces définitions. Si le critère sémantique de *« aboutness »* reste pour moi central, il reste trop délicat à manier à lui seul pour fournir une prise solide. Dans une approche opérationnelle du thème ou du topique, la relation de *« aboutness »* doit aussi être associée à une réflexion sur les caractéristiques d'identifiabilité et d'activation des référents « topicaux ». Toutefois, j'aimerais apporter quelques précisions à mes propos précédents en ce qui concerne la relation entre le thème et le statut du référent.

Comme Lambrecht l'a clairement démontré (cf. section 2.3.1), une distinction entre information et référent s'avère essentielle. L'accent mis dans l'approche pragoise sur la notion de « donneité » contextuelle avait pour résultat un amalgame souvent trop rapide entre les notions de donné et de thème. Un référent contextuellement donné ne sera pas forcément le thème, tout comme un référent nouveau ne sera pas forcément rhématique. Dans l'analyse informationnelle que je préconise, les expressions linguistiques seules ont moins d'importance que les relations qui unissent les groupes. Qu'un élément soit « nouveau » ou « connu » compte moins que les relations qu'il entretient avec les structures dans lesquelles il est inséré. C'est cette mise en rapport qui devient informative et qui crée l'information « nouvelle ». Il est certain que l'approche d'Halliday, en proposant une distinction entre le plan informationnel et le plan thématique, permet aussi d'éviter une association trop précipitée des deux plans. Toutefois, comme nous l'avons vu, le système informationnel systémique est fondé sur les indices prosodiques et en tant que tel est moins facilement transférable aux textes écrits. Le système proposé par Lambrecht me parait aussi plus fin dans son traitement des référents discursifs et de leur degré d'accessibilité cognitif, que la simple division « récupérable » ou « non récupérable » de l'approche systémique (cf. [THOM 09a] pour une présentation plus détaillée de la notion de « récupérable » chez Halliday).

Un autre affinement que j'estime nécessaire d'introduire concerne la notion de « thèmes multiples » de Halliday (2.2.4 *supra*). Il me semble particulièrement important de souligner la nature potentiellement composite du groupe thématique. Très souvent le thème phrastique et le sujet syntaxique coïncident (cf. chapitre 3 *infra*), mais il est nécessaire aussi de prendre en compte dans l'analyse informationnelle les autres éléments susceptibles d'apparaître en position « pré-sujet ». L'approche de Halliday nous fournit une solide base de départ pour le classement et l'analyse de ces autres éléments figurant dans la zone thématique. Dans le prochain chapitre, je propose de rediscuter de ce classement lors de la présentation des théories sur les adverbiaux cadratifs de M. Charolles (Charolles 1997 ; 2003)

Si ma conception de la structuration informationnelle reste essentiellement une conception phrastique, elle accorde néanmoins une grand importance au contexte textuel et situationnel. Il ne me semble guère envisageable de défendre une approche de la structure informationnelle qui serait coupée de tout contexte. Dans toutes les approches de la structure informationnelle présentées dans ce chapitre, le recours au contexte est essentiel. On a besoin du contexte afin de déterminer ce qui constitue un réfèrent identifiable (Lambrecht), le plus faible degré de dynamisme communicatif (Firbas) ou bien l'à-propos de l'énoncé.

Nombre de mes travaux portent précisément sur les interactions entre contexte et structure informationnelle. Certaines situations énonciatives et certains genres de textes favorisent l'utilisation de stratégies informationnelles spécifiques. Les éléments linguistiques susceptibles d'apparaître en tant que thèmes, ainsi que différentes configurations de progression thématique employées peuvent varier selon le genre de texte abordé [THOM 99a, 04d). Certaines structures dites « spécialisées », telles que les phrases clivées, l'extraposition, la diathèse passive, les énoncés existentiels, sont aussi plus exploitées dans certains genres textuels que dans d'autres [THOM 01, 04c et 05d]. Je reviendrai sur toutes ces questions dans les chapitres 7 à 9 de ce mémoire.

Partie B

De la théorie à la pratique

Chapitre 3 – Marqueurs de la structure informationnelle

Résumé

Mon troisième chapitre porte plus spécifiquement sur certaines réalisations linguistiques et formelles de la structure informationnelle.

J'examine dans un premier temps les liens privilégiés qui semblent exister – au moins dans les langues à sujet (Li et Thompson 1976) – entre la fonction de thème et la fonction de sujet grammatical (3.1). Cet examen est suivi par une réflexion sur les liens entre la fonction de thème et la position initiale de la phrase (3.2). Après avoir distingué entre différents types de détachement à gauche, je fais état des rapports complexes entre les thèmes et les adverbiaux cadratifs, tels qu'ils sont présentés par M. Charolles (2003) et S. Prévost (2003a).

Dans la troisième section du chapitre (3.3), j'examine un certain nombre de questions relatives à la problématique de l'ordre des mots. J'aborde ces questions dans une perspective typologique, en mettant en regard les langues qui ont un ordre de mots relativement rigide ou au contraire un ordre plus souple (3.3.1). Les études sur l'ordre des constituants et les permutations de cet ordre ont fait l'objet de différentes positions théoriques. Certains remettent en question le fait que toutes les langues ont un ordre syntaxique de base en envisageant la possibilité d'un ordre pragmatique dominant ; ce serait en particulier le cas dans les langues de type polysynthétique (cf. Mithun 1992). Après avoir discuté de ces différents enjeux, j'examine le cas d'une langue hautement polysynthétique à ordre « souple » : l'inuit (3.3.2). En faisant référence à mes propres études [THOM 05c ; THOM 04a], je présente ce que peuvent être les réalisations de la structure informationnelle dans cette langue.

Chapitre 3. Marqueurs de la structure informationnelle

Ce chapitre porte sur certaines réalisations linguistiques de la structure informationnelle. Alors que des langues comme le japonais ont des marques morphologiques dédiées à l'expression du thème (cf. D. Klingler 2003), l'anglais et le français ne possèdent pas de marqueur spécifique. Toutefois, il est généralement admis que les thèmes phrastiques entretiennent des liens privilégiés avec certaines des fonctions syntaxiques, tout comme avec certaines positions dans la phrase. Je commence par une brève réflexion sur les liens privilégiés qui semblent exister entre la fonction de thème et le sujet grammatical.

3.1 Thème et sujet grammatical

Nous avons déjà vu qu'une distinction nette entre niveau syntaxique et niveau thématique n'est pas facile à opérer et que les notions de thème et de sujet grammatical ont souvent été rapprochées ou même assimilées. Pour un générativiste comme J. Katz par exemple, le thème (*« topic »*) serait analogue au sujet d'une prédication :

> *The notion of a discourse topic is that of a common theme […] the topic carried from sentence to sentence as the subject of their predication.* (1980 : 26)

Cette assimilation de deux points de vue essentiellement différents semble être enracinée dans la tradition linguistique.

La tradition linguistique tend, en effet, à privilégier la catégorie du sujet syntaxique et à considérer que celle-ci structure la phrase. Cependant, la réalité est plus compliquée. Tout d'abord, toutes les langues ne sont pas des langues à sujet (*« subject prominent languages »*). De plus, même dans les langues à sujet obligatoire, le thème ne correspond pas forcément au sujet grammatical.

Des études typologiques, telle celle de C.N. Li et S.A. Thompson (1976), apportent un démenti à ces généralisations sur l'universalité du sujet. Li et Thompson isolent deux grandes classes de langues, les langues à sujet (*« subject prominent languages »*) et celle des langues à topique (*«* topic

prominent languages »). Cependant, il ne s'agit pas forcément d'une distinction absolue. On notera que dans le premier de ces deux cas de figure au moins, il s'agit seulement d'une préférence pour l'une des formes. En anglais (comme en français), les deux types de phrases coexistent :

John	*hit Mary*
Subject	Predicate
As for education	*John prefers Russell's ideas*
Topic	Commentary

Li et Thompson (1976 : 459)

Toutefois, dans les langues à sujets, l'expression linguistique du sujet syntaxique reste obligatoire : d'où la présence de sujets impersonnels, *il* en français, *it* en anglais, vides de toute information référentielle. Si le locuteur désigne un autre élément comme thème, le sujet doit aussi être exprimé. Dans un exemple avec un thème détaché, c'est le sujet qui doit s'accorder avec le verbe. Dans l'exemple suivant de C. Hagège (1978 : 16) : « les enfants, ça se met à la crèche »*,* le sujet syntaxique « ça » contribue à façonner l'autonomie du thème en disloquant, dans la procédure de thématisation, l'accord du pluriel avec le verbe. Le thème, en revanche, est indépendant du verbe et autonome. Comme Li et Thompson le soulignent :

> *Topic selection is independent of the verb. (…) within the constraints of the discourse, the speaker still has considerable freedom in choosing a topic […] regardless of what the verb is.* (1976 : 463)

Le thème n'est pas indifférent et répond à un choix. La syntaxe, la prosodie et également, à un certain degré, la sémantique, offrent des latitudes qui permettent de choisir l'élément que l'on souhaite mettre en position de thème. On trouve ainsi en anglais et en français, de nombreuses constructions détachées, ainsi que (mais plus rarement) des énoncés construits autours des objets syntaxiques préposés (cf. *infra*). Grâce à ces différents moyens, le locuteur est libre de choisir d'adopter la structuration informationnelle la plus appropriée à ces besoins. Malgré cela, l'existence d'un rapport étroit entre l'ordre syntaxique de base des langues à sujet obligatoire et le positionnement

du thème et du rhème est indéniable. C'est le sujet grammatical (qui apparaît aussi souvent en position initiale) qui est couramment associé avec le thème et les éléments en position d'objet syntaxique font souvent partie du rhème. Toutefois, il est préférable de considérer cette coïncidence de rôles non pas comme une règle absolue mais plutôt comme une configuration par défaut. Pour reprendre l'explication d' A. Davison :

> *In the absence of stronger indicators of salient and topic-indicating constituents, the topic is defined by subject position and morphology (…) The interpretation of subjects as topics may be "cancelled" (…) by stronger topic-indicating factors.* (1984 : 809)

3.2 Thème et position initiale

La question d'une association entre la fonction de thème et la position phrastique est liée aux questions soulevées dans les paragraphes précédents sur les rapports entre la fonction thématique et la fonction syntaxique. Dans les langues à sujet, si le positionnement de l'objet par rapport au verbe est variable, le sujet syntaxique est généralement antéposé au verbe[1]. C'est pour cette raison que la corrélation entre thème, sujet syntaxique et position initiale est répandue. Cependant, si comme nous l'avons vu, le thème et le sujet syntaxique ne coïncident pas constamment, le thème doit-il toujours apparaître en position initiale ?

Dans l'approche systémique, la position initiale constitue une marque de thématisation très répandue (cf. chapitre 2 *supra*). De la même façon, G. Brown et G. Yule établissent une identification catégorique entre thème et position initiale :

> *We shall use the term theme to refer to a formal category, the leftmost constituent of the sentence.* (1983 : 126)

Cependant, ce que Brown et Yule considèrent comme *« theme »* n'est pas le thème tel que nous l'avons défini dans le chapitre précédent. Ils font une distinction entre *« theme »* et *« topic entity »*. C'est la *topic entity* qui englobe tous les attributs de *« aboutness »* (ce dont on parle). Si les deux termes ne

[1] Cf. Fuchs(2006) pour un traitement détaillé de la place du sujet nominal en français contemporain.

sont pas représentés par la même unité linguistique dans la phrase, les auteurs gardent le terme de *« theme »* seulement pour l'élément en position initiale.

Pour J. Firbas, la position initiale ne peut pas être systématiquement associée au thème. L'ordre des mots n'est qu'un des paramètres dans la FSP :

> *The system of FSP is hierarchically superior to that of word order. Word order is only one of the means of FSP and acts as such only under certain conditions.* (1992 : 125)

Les deux autres facteurs qui jouent dans la FSP, la dépendance contextuelle (les éléments dépendant du contexte – les éléments donnés – ont un faible degré de dynamisme communicatif) et la structure sémantique (certaines relations sémantiques possèdent un plus haut degré de dynamisme communicatif que d'autres) peuvent contrebalancer les effets de la linéarité. En conséquence, ce qui sera interprété comme thème ne sera pas forcément en début de phrase :

> *These interpretative arrangements, however, need not coincide with the actual linear arrangements.* (*Ibid.* : 139)

Comme K. Lambrecht le précise également, si l'on souhaite adopter une conception de topique/thème fondée sur la notion d'à-propos, et en l'absence d'une marque morphologique spécifique, on ne peut pas se fier seulement à la position phrastique :

> *The definition of topic in terms of the pragmatic concepts of aboutness and relevance explains that it is sometimes not possible to determine the topic of a sentence on the basis on the syntactic structure of that sentence alone, at least in languages like English in which neither grammatical relation nor linear constituent order are reliable topic indicators.* (1994 : 120)

Je préfère m'aligner sur de telles conclusions. Tout comme pour le sujet grammatical, je ne crois pas qu'il faut associer systématiquement la position initiale et le thème. Une analyse du contexte est nécessaire afin de pouvoir confirmer le statut thématique de tout élément. Un élément en position initiale peut aussi assumer d'autres fonctions. Pour citer Lambrecht de nouveau :

> *Given that sentence initial position is cognitively speaking an eminently salient position, it would be surprising if the prominence associated with this position could be exploited for a single function, such as the marking of the topical relation.* (*Ibid.* : 201)

Les théories sur le statut spécial de la position initiale sont nombreuses (Enkvist 1989, Givón 1988 ; Halliday 1994 ; Virtanen 2004 etc.). Pour T. Givón, la position initiale correspond à la position la plus saillante sur le plan cognitif. Selon N.E. Enkvist (1985 ; 1989), le rôle de la position initiale serait de nature discursive : on met en première position l'information la plus pertinente pour la bonne interprétation du texte (le principe du CIF : Crucial Information First). T. Virtanen souligne l'importance de la position initiale à la fois comme point d'ancrage et comme point de départ. Manifestement, la position initiale reste tellement significative d'un point de vue pragmatique qu'elle mérite une attention particulière dans toute analyse textuelle (cf. la discussion sur les adverbiaux cadratifs ci-dessous). Toutefois, une première distinction importante s'impose. Quelle est la nature grammaticale de l'élément en position initiale ? Lorsque l'élément initial n'est pas le sujet, mais un SP adverbial, il peut entretenir avec le sujet toute sorte de relations adjuvantes ou concurrentes (Fuchs et Fournier 2003). Il convient en particulier de distinguer entre la mobilité positionnelle des adverbiaux et la « thématisation » des groupes nominaux, autrement dit le détachement à gauche.

3.2.1 Détachement à gauche

Sans établir de correspondance stricte, je considère que la position de détachée à gauche, tout comme la position du sujet, fait partie des positions syntaxiques privilégiées, propres à loger les thèmes.

Il y a une importante littérature sur les différentes types de structures détachées (Fradin 1990 ; Lambrecht 2001 ; Prévost 2003b), mais on s'accorde généralement à distinguer entre deux types principaux : les détachements avec reprise et les détachements sans reprise (ou non liés). Dans les deux cas, la position initiale détachée se définit par sa position extra-propositionnelle – une position qui sera normalement appuyée par une intonation spécifique à l'oral ou une virgule à l'écrit.

(1) ***Now the wizard,*** *he lives in Africa.*

(2) ***The typical family today,*** *the husband and wife both work.* (Lambrecht 1994)

Devant de tels exemples on s'accorde, malgré quelques différences terminologiques entre les approches théoriques[2], à reconnaître l'élément GN détaché comme un thème (ou topique), sur lequel il est pertinent de dire quelque chose. On est en présence d'une construction syntaxique spécialisée [THOM 00], où le statut topical de l'élément détaché est volontairement explicité.

Je reviendrai, dans le prochain chapitre, sur l'emploi de ces deux types de détachement à gauche, ainsi que sur d'autres constructions syntaxiques spécialisées destinées à expliciter la structure informationnelle d'une phrase. Toutefois, dans une discussion sur la position initiale détachée, il est également important de bien distinguer entre la nature des éléments détachés. Dans les deux constructions détachées ci-dessus, les éléments détachés sont des GN référentiels. Or dans cette position de détachée à l'initiale, sont également susceptibles d'apparaître divers types d'adverbiaux (adverbes, SN, SP) avec des valeurs diverses. Malgré le fait que les adverbiaux initiaux présentent certaines similarités formelles avec les éléments détachés, leur fonction dans le texte est souvent très différente.

3.2.2 Adverbiaux cadratifs

J'ai déjà abordé dans la discussion sur le modèle systémique certains problèmes liés à l'analyse informationnelle des groupes circonstanciels initiaux (cf. 2.2.3). De tels éléments – des thèmes topicaux marqués dans la terminologie de M.A.K. Halliday – ne constituent pas en général l'à-propos de la phrase en tête de laquelle ils sont détachés. Dans la séquence :

(3) In England, most young women go out to work

They often work in poorly paid sectors.

l'expression adverbiale de lieu ne constitue pas à elle seule le topique/thème de la phrase. Pour cette raison un certain nombre de chercheurs choisissent de désigner les adverbiaux spatio-temporels initiaux de ce type par le terme

[2] Dans le cas des détachements avec reprise, Prévost préfère parler de « topiques en construction » (2003b : 120).

spécifique de *« scene-setting topics »* (Chafe 1976 ; Lambrecht 1994) ou *topique scénique* (Ertershik-Shir 1999).

Dans mes travaux antérieurs, j'ai choisi d'analyser les circonstanciels antéposés de ce type comme faisant partie inhérente d'un groupe thématique composite. Dans la première phrase de la séquence (3) :

> *In England, most young women go out to work.*

ce n'est pas à propos de l'expression adverbiale spatiale que l'on dit quelque chose, mais à propos de ce que font les femmes en Angleterre. Autrement dit, l'adverbial fixe un critère pour l'interprétation de la proposition et contribue à pousser ainsi le thème sujet. Comme l'explique P. Le Goffic :

> *Même s'il n'apparait pas, psychologiquement, comme ce dont l'énoncé parle, le circonstant initial de temps ou de lieu sert de repère, permettant de localiser ou de dater un événement ou une situation dans son ensemble.* (1993 : 463)

Toutefois, cette analyse ne permet pas de vraiment rendre compte du rôle discursif de ces adverbiaux.

L'hypothèse linguistique de l'encadrement de discours (Charolles 1997) permet, en revanche, de tenir compte du fort potentiel « organisateur » des adverbiaux en position initiale. Selon cette hypothèse, certaines expressions, en général des adverbiaux, de par leur position spécifique à l'initiale de la phrase, possèdent une capacité d'étendre un critère d'interprétation au-delà de la phrase dans laquelle ils apparaissent. Dans la séquence (3), par exemple l'adverbial spatial *« In England »* étend sa portée sur les deux propositions qui le suivent.

> *(3) In England, most young women go out to work.*
>
> *They often work in poorly paid sectors.*

Les adverbiaux en tête de phrase cumuleraient ainsi deux fonctions : une fonction représentationnelle et une fonction procédurale. La fonction représentationnelle est liée au fait que l'expression introductrice pose un critère linguistique ou métalinguistique suivant lequel les propositions suivantes seront interprétées. La fonction procédurale vient du fait que l'adverbial indexe non

seulement la proposition d'accueil, mais aussi potentiellement un certain nombre de propositions subséquentes. Dans les termes des métafonctions de la linguistique systémique (métafonctions idéationnelles, interpersonnelles et textuelles), les circonstanciels initiaux fonctionneraient sur le plan idéationnel et textuel, alors que les circonstanciels intraprédicatifs ne fonctionneraient que sur le plan idéationnel.

Cette fonction d'indexation ne vaut pas que pour les circonstants scéniques mais peut s'étendre à toute une gamme de syntagmes prépositionnels : organisationnels ; méditatifs, énonciatifs etc. (cf. Charolles et Péry Woodley 2005). Elle vaut aussi pour les subordonnées adverbiales. Comme j'ai pu le constater dans mes travaux sur les propositions en *« if »* et « si » dans le discours médical spécialisé, la postposition de la phrase conditionnelle favorise une lecture intraprédicative [THOM 07 ; THOM 08a] :

(4) The regimen was repeated after 28 days if the patient had recovered from all toxic effects.

La décision thérapeutique est présentée avant les définitions opérationnelles détaillant les critères applicables pour cette décision. L'étendue de cette phrase conditionnelle semble ainsi se limiter à la proposition principale antéposée, et ne fournit pas un point de départ pour la suite. En revanche, l'antéposition de la phrase conditionnelle signale une lecture extraprédicative :

(5) If we want to optimise patient accrual, we need to better understand the resources issue required to properly conduct clinical trials.

La phrase conditionnelle ouvre un cadre qui ne s'applique pas seulement pour la prédication suivante, mais qui reste disponible pour d'autres prédications dans la suite du discours. Ce qui semble être commun à tous les introducteurs de cadres, SP et proposions subordonnées, c'est le pouvoir conféré par la position initiale à exercer une influence structurante au-delà des contours de la phrase.

Toutefois, le lien entre les adverbiaux cadratifs et la notion de thème/topique est complexe. Selon Charolles et Péry-Woodley (2005), les expressions

cadratives (ou potentiellement cadratives) ne sont pas habituellement aussi des thèmes phrastiques :

> Ils se contentent d'indexer des contenus exprimés par des phrases qui sont à propos d'autres choses, le plus souvent à propos des référents dénotés par leur sujet. (*Ibid.* : 4)

Toutefois, si les expressions cadratives (ou potentiellement cadratives) ne sont pas *a priori* destinées à être des thèmes phrastiques, il n'est cependant pas exclu que dans certains cas elles peuvent devenir des thèmes par la suite. Tout dépend du texte qui suit. Dans la phrase, *In England, most young women go out to work,* il est impossible sans plus de contexte, de décider du statut thématique de *most young women*. On pourra au mieux attribuer un statut de thème par « défaut » au GN, en considérant qu'en tant que sujet, il est le constituant qui possède le plus de chance d'être désigné comme thème/topique. En revanche, si le locuteur poursuit son discours en énumérant les activités d'autres catégories de la population en Angleterre, on peut estimer que c'est l'adverbial qui a un statut de thème/topique. Comme le précise Charolles :

> Il n'est pas possible de statuer sur le topique (*aboutness*) d'une phrase isolée. Par défaut de contexte ultérieur, les adverbiaux détachés en tête de phrase, n'indiquent pas ce à propos de quoi est la phrase. Mais rien n'empêche que la suite oblige à leur restituer ce statut. Ce constat s'il est bien fondé, milite en faveur d'approches dynamiques des phénomènes de topicalisation. (2003 : 44)

Autrement dit, si le contexte s'y prête, un adverbial initial peut avoir à la fois une fonction cadrative et une fonction de thème phrastique.

Bien que le modèle des « cadres de discours » de Charolles se soit révélé extrêmement efficace pour rendre compte du potentiel d'indexation des adverbiaux initiaux en français, à ma connaissance, ce modèle n'a pas été largement testé sur d'autres langues. À partir d'un (petit) corpus de textes « expositifs » en anglais, P. Crompton (2006) prétend par exemple, ne pas trouver de réelle spécificité aux adverbiaux en position initiale en anglais. Selon cet auteur, les adverbiaux en position intraprédicative ont autant de portée sémantique que ceux détachés à l'initiale. Cependant, à mon avis, la question de la portée reste largement tributaire du type et du genre de texte examiné.

Pour cette raison, il serait important d'examiner ces questions non seulement sur d'autres langues, mais aussi sur des genres textuels divers.

Dans les travaux entrepris dans le cadre du projet ANR « Spatial Framing adverbials[3] » (2006-2009) de l'équipe LaTTiCe, nous avons créé un corpus bilingue (anglais/français), précisément afin de mettre à l'épreuve l'hypothèse linguistique de l'encadrement de discours sur divers types de texte [THOM 08c]. Ce corpus contient à la fois des articles de presse généraliste, des articles tirés des journaux *Le Monde* et *The Guardian,* et un corpus de vulgarisation scientifique sur la vulcanologie dans les deux langues. Notre étude était centrée sur l'analyse de deux types d'adverbiaux initiaux, ceux introduits par *Sur* en français et ceux introduits par *On* en anglais. Parmi les résultats que l'on peut retenir de cette étude figure la mise en évidence d'une large utilisation du potentiel des adverbiaux cadratifs en anglais, plus importante que dans le corpus comparable en français.

Un autre aspect important de ces travaux porte sur les interactions entre certains types de topiques et les introducteurs de cadres. Prévost (2003a) propose de considérer les notions de topique et d'introducteur de cadre dans un continuum. Il me semble aussi que la frontière entre les deux types de fonctions n'est pas toujours nette, en particulier entre les topiques non liés (cf. *supra*) et les cadres thématiques. Dans ces travaux récents entrepris avec L. Sarda, nous avons essayé de mieux caractériser les interactions entre ces deux mécanismes de structuration textuelle. Pour cette raison nous avons établi une distinction nette entre structuration informationnelle et structuration textuelle. Nous avons analysé d'une part le statut informationnel des SP adverbiaux en précisant leurs rôles de « s*etting* » (ou « décor ») pour la phrase ou de topique. Nous avons considéré d'autre part la capacité des SP adverbiaux à jouer, à un autre niveau, un rôle d'introducteur de cadre. Cette dissociation des concepts

[3]: Le projet ANR *Spatial Framing Adverbials,* dirigé par M. Charolles et L. Sarda, a pour objectif de décrire les adverbiaux introduits par des prépositions spatiales et leur changement de sens selon leur position et leur fonction (cf. Sarda 2008). Les SP en « *Sur »* font partie des syntagmes prépositionnels étudiés.

informationnels et textuels a permis de décrire l'interaction et le partage du travail de ces deux dimensions [THOM 08c].

3.3 L'ordre des mots

Que l'on se situe du côté de l'émetteur ou du récepteur d'un texte, l'ordre des mots reste un facteur essentiel pour en encoder ou pour en décoder la structure informationnelle. Comme on vient de le voir, la position initiale dans la phrase est particulièrement significative et peut être exploitée pour signaler différentes interprétations pragmatiques. Toutefois, la liberté de disposer des constituants selon son gré n'est pas totale. Pour Enkvist (1976 ; 1985 ; 1991), les contraintes pesant sur l'ordre de mots sont très diverses. Il identifie en plus de la syntaxe et de la structure sémantique, des facteurs tels que l'iconicité expérientielle, la lourdeur des constituants et les contraintes liées au traitement de l'information en temps réel. Le poids accordé à tous ces différents facteurs et leurs interactions vont varier selon les situations de communication (oral/écrit) et selon les structures des langues concernées.

Une première grande distinction s'impose entre les langues soumises à de fortes contraintes syntaxiques et les langues avec un ordre de mots souple.

3.3.1 Ordre syntaxique *vs* Ordre pragmatique

Bon nombre de travaux sur la typologie syntaxique posent comme hypothèse de base que toutes les langues disposent d'un ordre de mots de base (Comrie 1989 ; Croft 2003). Tout en reconnaissant la possibilité de réagencement pragmatique, cet ordre de base est considéré comme un trait fondamental qui permet de prévoir d'autres caractéristiques de cette langue.

Pour beaucoup de langues, l'existence de cet ordre de mots de base semble en effet une évidence. Il y a un ordre qui prédomine sur le plan statistique ; d'autres ordres sont non seulement rares mais aussi très marqués sur le plan pragmatique.

En anglais, comme en français, il existe un ordre de base reconnu[4]. Si les fonctions non essentielles, telles que les adverbiaux circonstanciels, jouissent d'une assez grande mobilité (cf. *supra*), les fonctions essentielles (Le Goffic 1993) suivent un ordre fixe. La phrase anglaise adopte typiquement l'ordre SV(O/C) : Sujet-Verbe-Objet/ complément essentiel. Cet ordre de base est fortement régi par la syntaxe. Les deux phrases *« John hit Paul » et « Paul hit John »* se distinguent ainsi par la place respective des deux GN.

Dans une interprétation pragmatique non marquée, cet ordre canonique est habituellement corrélé à un ordre thème/rhème, allant de la gauche vers la droite. L'élément qui joue le rôle du sujet sera normalement le thème/topique et le verbe et le complément éventuel seront rhématiques. Toutefois, cette configuration n'est pas toujours adaptée aux besoins de la communication. Le locuteur peut choisir d'expliciter au moyen de la prosodie ou avec l'aide de certaines structures syntaxiques spécialisées (cf. chapitre 4) une autre interprétation. Certaines permutations dans cet ordre de base peuvent aussi être exploitées pour indiquer une structuration informationnelle particulière :

(6) We painted these blocks –SVO

(7) These blocks we PAINTED – OSV

(8) On this axis is electric FIELD[5] *– CVS-* [THOM 04d ; THOM 05d]

Un ordre de mots inhabituel peut ainsi être employé comme indice, servant à guider le récepteur vers une interprétation marquée.

Comment fonctionnent alors les langues avec un ordre de mots plus « souple » ? Si, dans le cas de l'anglais et du français, l'existence d'un ordre de mots de base régi par la grammaire est reconnue, ce n'est pas le cas de toutes les langues. Dans les langues slaves, telles que le russe ou le tchèque, l'agencement des constituants reste beaucoup moins contraint par la grammaire. Les facteurs pragmatiques jouent également un rôle très important. Mais il n'en reste pas moins que l'on peut toujours identifier un ordre de mots

[4] Lambrecht (1987) a constaté que les dislocations à gauche avec reprises pronominales (Pierre, il …) sont plus fréquentes que les phrases traditionnelles SV(O) avec sujets lexicaux en français parlé.

[5] La fonction de l'inversion dans les communications scientifiques orales est traitée, par exemple, en [THOM 01 et THOM 04d].

de base. Un locuteur tchèque saura distinguer un ordre pragmatique neutre d'un ordre plus marqué. Toutefois, dans d'autres langues encore, l'identification d'un ordre de mots de base reste très discutable (Mithun 1992). Sur le plan grammatical, tous les ordres sont théoriquement envisageables : SVO ; SOV ; OVS ; OSV ; VOS ; VSO.

Dans une étude sur trois langues de ce type : le ngandi (une langue australienne), et le coos et le cayuga (langues américaines et canadiennes), M. Mithun (*ibid.*) décrit les difficultés à identifier un ordre de mots de base à partir des critères habituels de fréquence statistique, de simplicité syntaxique ou de neutralité pragmatique. L'opposition entre ordre des mots marqué et ordre des mots neutre n'est pas pertinente car il n'existe pas un ordre de mots de base. Les ordres les plus fréquents ne sont pas plus « neutres » que les ordres alternatifs[6]. Néanmoins, pour Mithun, il est impensable de considérer toutes les permutations constatées dans ces langues comme étant purement aléatoires.

> *Surely speakers would not randomly vary a cue as salient as word order. If order does not signal the syntactic or semantic roles of constituents, perhaps it has a pragmatic function.* (1992 : 26)

Mithun conclut en revanche à l'importance cruciale de la pragmatique dans l'agencement des constituants.

Dans ces langues dites *« pragmatically based languages »* (*ibid.* : 40), c'est l'importance relative de chaque constituant qui est déterminante. Toutefois, cet ordre pragmatique ne correspond pas à une organisation thème-rhème traditionnelle, où le dynamisme communicatif augmente par degrés du début à la fin de la phrase (Firbas 1964 ; 1992 et chapitre 2 *supra*). Dans le ngandi, le coos et le cayuga, l'ordre adopté est du type nouveau-donné. Les mots porteurs des informations revêtant le plus d'intérêt sont placés au début de la proposition ; suivent les mots transmettant des informations à caractère plus prévisible, ou d'une importance moindre. Si les constituants sont sur un pied d'égalité vis-à-vis de leurs statuts d'activation dans le discours, ils sont classés selon leur degré

[6] Les situations pragmatiques inhabituelles sont signalées par des constructions spéciales : en ngandi, par exemple, le préfixe « ga » est un préfixe focalisant (Mithun 1992 :44).

d'importance ou leur degré de *« newsworthiness »* (1992 : 31). On assisterait en quelque sorte à un principe de dynamisme communicatif inversé.

Il semblerait ainsi que l'organisation thème-rhème ne représente pas une tendance universelle. Mais on peut se demander si ce type d'organisation pragmatique décrite par Mithun est aussi « efficace » qu'une organisation du type thème-rhème. Pour beaucoup de linguistes comme nous l'avons vu (chapitre 2 *supra*), le thème/topique se définit en termes d'à-propos. Le thème, c'est ce dont on parle. On considère aussi très souvent que le thème établit un point de départ ou une orientation. Selon Halliday, c'est ce rôle organisateur joué par le thème qui explique sa position habituelle en début d'énoncé. Il s'agit de : « *the peg on which the message is hung* » (Halliday 1970 : 161). Ainsi peut-on légitimement se demander, comme le fait Mithun, comment dans ces langues on peut laisser la mention de ce « point de départ » jusqu'à la fin de l'énoncé :

> *If it assumed that a progression from theme to rheme is inherently more logical cognitively, are Cayuga, Ngandi and Coos somehow less logical or natural? (…) If themes do indeed provide such a point of departure, how can (…) speakers leave them until the end?* (1992 : 42)

La réponse est que les locuteurs de ces langues n'attendent pas la fin de la phrase pour établir son orientation. Si les thèmes « constants » sont simplement rappelés par les affixes sur le groupe verbal, l'introduction d'un nouveau référent thématique en revanche se fera plus tôt dans l'énoncé. On trouve aussi en début d'énoncé des adverbiaux et des particules pour signaler les indications d'espace et de temps. L'organisation des énoncés n'est donc pas moins logique ou plus logique, mais elle est exprimée de manière différente.

Les trois langues « pragmatiques » décrites par Mithun sont très polysynthétiques et possèdent une morphologie verbale très riche. Mithun semble suggérer qu'il existe un lien entre ces traits morpho-syntaxiques particuliers et un ordre de mots strictement pragmatique. Mais dans quelle mesure pouvons-nous établir réellement des corrélations entre la structure morphologique d'une langue et l'expression de sa structuration informationnelle ? Dans la prochaine section, je propose de décrire une étude faite sur une autre

langue très polysynthétique : le tunumiisut, un dialecte inuit du Groenland de l'Est. Cette étude entreprise avec une collègue spécialiste du domaine inuit (N. Tersis) vise justement à corréler les variations constatées dans l'organisation de la phrase inuite avec une étude de sa structuration informationnelle.

3.3.2 Le cas d'une langue polysynthétique : l'inuit

Les langues du continuum inuit sont caractérisées par un haut degré de polysynthèse au sens où derrière une base lexicale unique nominale, verbale ou verbo-nominale se synthétisent un grand nombre de morphèmes exprimant des valeurs quantitatives, qualitatives, impressives, aspectuelles, actancielles. La morphologie verbale est particulièrement développée : le radical verbal est accompagné d'un morphème défini traditionnellement comme un mode, lui même suivi d'indice(s) de personne et de nombre renvoyant à l'agent et au patient de l'opération prédicative. Le radical nominal est suivi d'une marque indiquant sa fonction dans la phrase et désigné comme un cas. Tous ces traits semblent de nature à favoriser un ordre de mots très souple avec un ordre de mots déterminé par la pragmatique.

Cependant malgré cette flexibilité potentielle, les spécialistes de la langue inuite s'accordent pour considérer l'ordre SOV comme un ordre de base neutre et très fréquent dans la plupart des dialectes (Fortescue 1993). Dans notre étude [THOM 04a et THOM 05c], nous avons donc voulu examiner le degré de stabilité de cette configuration SOV dans le tunumiisut. Notre propos était d'analyser les différentes variations constatées par rapport à cet ordre de base en mettant en relation différents paramètres syntaxiques, sémantiques et pragmatiques.

Il nous a paru fondamental d'envisager cette analyse à partir d'un corpus (oral) entièrement contextualisé, et non pas seulement à partir d'énoncés isolés de leur contexte. Cela nous permet d'étudier les enchaînements inter-phrastiques, et de façon plus large, d'envisager la cohérence textuelle dans une perspective pragmatique. Nous avons analysé la transcription d'un discours narratif d'un locuteur du tunumiisut, qui portait sur une description de sa ville natale et de

ses coutumes. (Rappelons que le tunumiisut est une langue exclusivement orale.) L'analyse détaillée des cent trois propositions principales contenues dans ce petit compte rendu nous a permis d'étudier les stratégies informationnelles adoptées par le locuteur et la progression de son récit en fonction d'un contexte et d'un co-texte précis.

Nos résultats semblent confirmer l'existence d'un ordre de mots de base SOV en tunumiisut. Face aux possibilités théoriques de permutation, seul un faible pourcentage d'énoncés présentent un changement par rapport à l'ordre canonique. Sur les cent trois propositions analysées, il y a seulement treize propositions avec un ordre de mots différent de l'ordre SOV attendu. Dans chaque cas, les permutations effectuées signalent une interprétation pragmatique particulière. Il s'agit très souvent d'un élément postposé au verbe (VS ou VO) – une permutation qui signale une focalisation particulière sur cet élément :

(9) *aamma* *mamaat-kaayu-u-pa-qput*
and find good-intensive-be-indic-1pl.3sg
V
Topic

kataatti-it-tu-ta *// mattak* *//*

Greenlanders-be.pl-contemporative-1pl whale skin
DV O
Comment
"*and what we find really good, we who are Greenlanders, is whale skin."* [THOM 05c : 495]

Dans cet exemple, les éléments les plus informatifs sont focalisés en position finale : l'inverse de ce que Mithun a trouvé dans ses trois langues « pragmatiques (cf. *supra*). Dans le texte antérieur, notre narrateur vient de décrire quelques spécialités culinaires de la région. La peau de narval n'a pas encore été mentionnée : il s'agit ainsi d'un référent entièrement nouveau. Le positionnement de ce GN à la fin de l'énoncé (après le groupe verbal dépendant déjà focalisé) souligne son importance comme le plat préféré des Groenlandais.

En considérant ces résultats d'ensemble, il est important de tenir compte de deux facteurs. Tout d'abord, comme en inuit le groupe verbal contient toujours des indices actanciels, l'expression lexicale du sujet et de l'objet reste facultative sur le plan syntaxique. Un groupe verbal peut suffire à faire un énoncé complet. Un quart

des énoncés dans le texte analysé (ving-cinq propositions sur cent trois) rentre dans ce cas de figure et ne permet pas par conséquent une analyse en termes d'ordre de mots. Dans toute réflexion sur l'ordre des mots en inuit, il est également très important de prendre en compte le type de structure verbale choisi par le locuteur. Dans les dialectes inuits, il existe deux principales constructions actancielles (Tersis 2000). Soit le verbe est monoactanciel et intègre un participant unique (agent) ; soit il est biactanciel et intègre deux participants (agent/patient) sous la forme d'indices en finale. L'orientation du verbe et son degré de transitivité sont fondamentaux pour l'ordre de mots puisqu'ils impliquent une construction directe ou indirecte de l'objet – un choix qui sera aussi déterminant pour l'interprétation de sa structure informationnelle.

Dans le premier cas où le verbe est monoactanciel (ou intransitif), le sujet n'a pas de marque spécifique (Ø absolutif). C'est la construction dite anti-passive. L'adjonction d'un deuxième participant (patient) dans la mesure où il n'est pas requis par le verbe sera forcément péripherisée. Il est dénoté par le morphème *-mi,* un morphème qui sert également par ailleurs à signaler la fonction instrumentale, et peut être considéré comme un objet indirect. Cette structure semble surtout utilisée quand le patient est indéfini, partitif ou générique Des raisons pragmatiques viennent s'ajouter à cette interprétation car dans ces énoncés, les noms définis à l'absolutif reçoivent habituellement une interprétation thématique et les objets indirects sont rhématiques. Il s'ensuit que nous constatons une interaction très nette entre un ordre SiOV[7] et ordre thème-rhème, comme dans l'exemple suivant :

(10) *asimiuqaqpiit qani-niqsa-a* *ikkatti-mi*

surrounding village.pl be close-most-3sg Ikkateq
S iO

ati-qaq-pu-q
name-exist-indic-3sg
V
"The nearest of the surrounding villages is called Ikkateq."

[7] SiOV : *Subject, indirect Object, Verb.*

Dans l'exemple ci-dessus, extrait de notre corpus, nous percevons la progression thème-rhème, de la gauche vers la droite. Le locuteur vient de décrire son village natal et poursuit avec une description des villages environnants. C'est le groupe sujet qui joue le rôle de thème et le groupe verbal, *ati-qaq-pu-q* (qui signifie littéralement d'avoir comme nom*)* et son complément indirect (le nom du village le plus proche) qui constituent l'apport d'information ou le rhème.

Toutefois, dans la deuxième construction actancielle de l'inuit, l'organisation informationnelle est habituellement inverse. Avec un verbe biactanciel (ou transitif) l'énoncé s'oriente sur le patient de l'opération prédicative. Dans cette construction, dite ergative, c'est l'agent nominal du prédicat verbal à deux participants qui reçoit une marque du cas ergatif *-p*. Le choix de la structure à marquage ergatif est privilégié si le patient, à l'absolutif, est spécifique et défini. Il s'ensuit que dans l'énoncé à marquage ergatif, c'est souvent le patient qui semble avoir le rôle de thème et l'agent marqué qui apporte l'information nouvelle. Si les deux participants sont réalisés lexicalement ; on aura ainsi un ordre rhème-thème. Au pluriel la marque d'ergatif n'est pas représentée sur l'agent :

(11a) tikkaa-p	ittiq	aaqqissii-nikuu-wa-a
man-erg	house.abs	build-past-indic-3sg.3sg
S	O	V

« La maison, l'homme l'a construite »

(11b) *amiiqikkaqmiit mittaqpik*		*aaqqissii-nikuu-wa-at*
American.pl	airport.abs	build-past-indic-3pl.3sg
S	O	V
Rhème	Thème	

« L'aéroport, les Américains l'ont construit »

Dans l'exemple (11b), extrait du notre corpus, l'aéroport (*mittaqpik)* a déjà joué le rôle de thème dans le texte antérieur. Ce sont les Américains, en tant que constructeurs de l'aéroport, qui apportent un complément d'information par rapport à ce thème déjà établi. Le choix entre la construction transitive ou intransitive semblerait ainsi avoir un impact crucial sur l'interprétation pragmatique de l'énoncé. Les « instructions » associées à l'ordre de mots sont

dans ce cas moins importantes que celles rattachées au choix de la construction verbale. Autrement dit, l'orientation de la phrase autour de la fonction objet l'emporte sur le positionnement phrastique. Les constructions biactancielles sont de manière générale assez rare en tunumiisut. Dans notre corpus seulement vingt-deux propositions (sur cent trois) sont « transitives ». Du fait de cette moindre fréquence, l'emploi même de cette construction peut aussi signaler au récepteur un emploi « marqué » de la langue.

Notre analyse de l'ordre de mots en tunumiisut révèle ainsi une interaction complexe entre les niveaux morpho-syntaxiques, sémantiques et pragmatiques. Le tunumiisut n'est pas une langue exclusivement « pragmatique » du type décrit par Mithun. Au sein de la proposition intransitive, il y a une corrélation entre un ordre SOV de base (ou plus précisément SOiV) et un ordre pragmatique « neutre » thème-rhème. Un premier type de réorganisation pragmatique est ensuite lié à l'emploi d'une construction transitive – ce qui peut se traduire sur le plan informationnel par un ordre rhème-thème. Un deuxième type de réorganisation pragmatique est fourni par les possibilités de permutations entre constituants au sein de ces deux constructions.

Il est évident que ce travail sur un corpus restreint ne nous permet pas de faire de grandes généralisations. Néanmoins, l'étude d'une langue très différente de l'anglais et du français m'a donné un aperçu des rapports complexes entre l'ordre de mots et le concept d'un ordre pragmatique « neutre ». Dans les langues avec un ordre de mots (théoriquement) souple (hongrois, inuit, ancien français etc.), l'opposition entre une phrase neutre et une phrase marquée ne se posera pas forcément de la même façon que dans les langues avec un ordre de mots rigide régi par la grammaire (anglais, français moderne). L'impact relatif des effets sémantiques, morphologiques, prosodiques syntaxiques et pragmatiques qui en résultent peut être très différent.

Les moyens mis en œuvre pour signaler une interprétation pragmatique « marquée » varient selon la langue concernée. Toutefois, même à l'intérieur d'un même système, ces moyens peuvent aussi varier selon le mode (écrit/oral)

et le genre du texte concerné. Par exemple, un agencement syntaxique qui sera perçu comme un ordre « neutre » dans un anglais écrit « soigné », pourra apparaître comme très marqué dans un récit oral spontané, et vice versa (cf. [THOM 05d). Autrement dit, selon les besoins de la communication, les valeurs communicatives peuvent prendre le devant sur les relations syntaxiques.

Comme les travaux de Mithun l'ont démontré, l'ordre thème-rhème n'est pas un ordre « universel ». Certaines langues s'organisent selon un ordre pragmatique d'importance décroissante (cf. *supra*), mais ces langues ne sont pas pour autant moins « claires » que celles qui s'organisent selon un ordre thème-rhème. Toutefois, il me semble qu'au lieu d'établir une distinction tranchée entre deux types de structuration pragmatique, on peut aussi envisager un type de continuum. Tout comme les langues du monde peuvent varier par rapport au degré de contrôle de la pragmatique sur l'ordre des mots, il me semble logique aussi de concevoir des variations en fonction du type de structuration pragmatique souhaitée. Différentes situations communicatives nécessitent des solutions particulières.

Chapitre 4 – Une syntaxe au service du texte

Résumé

Dans ce chapitre je cherche à démontrer le rôle important que peut jouer la syntaxe dans la signalisation de la structuration informationnelle. Dans les langues d'ordre fixe, telles que l'anglais ou le français, il existe un certain nombre de structures syntaxiques, dites structures spécialisées [THOM 00], qui permettent de « jouer » ou de jongler avec l'ordre canonique des constituants dans la phrase afin de signaler une interprétation pragmatique précise (4.1). Parmi ces structures syntaxiques ou configurations spécialisées on trouve :

– les structures détachées (gauche et droite) ;
– la diathèse passive ;
– les phrases clivées ;
– les phrases pseudo-clivées ;
– les énoncés existentiels ;
– l'extraposition ;
– les structures hypotaxiques.

Dans la deuxième section de ce chapitre (4.2), je décris une étude que j'ai faite spécifiquement sur l'emploi des phrases clivées en anglais et français [THOM 02, THOM 09a]. On a parfois considéré les phrases clivées comme la structure spécialisée par excellence – un moyen pour compenser la rigidité d'un ordre de mots fixe dans les deux langues. Or les phrases clivées sont nettement plus fréquentes en français qu'en anglais. Je propose un certain nombre d'hypothèses pour expliquer pourquoi l'anglais a moins fréquemment recours à cette structure.

Chapitre 4. Une syntaxe au service du texte

Dans ce chapitre, je passe en revue un certain nombre de constructions syntaxiques spécialisées destinées à expliciter la structure informationnelle d'une phrase. Je préfère employer le terme de *syntaxe « spécialisée »* [THOM 00 : 76], plutôt que de *syntaxe marquée*, afin de réserver les termes de *marqué* et de *non marqué* pour l'expression de la structure informationnelle. Il est important de souligner que dans ma conception de la linguistique les constructions spécialisées ne sont pas des structures « dérivées ». Il n'y a pas de transformation par rapport à une structure de base. Ce sont des structures qui sont générées directement par les locuteurs en fonction d'un besoin communicatif spécifique.

J'estime que la signalisation de la structuration informationnelle n'est pas quelque chose d'aléatoire. Elle répond à un choix. La syntaxe, la prosodie et la sémantique, offrent des latitudes qui permettent de choisir l'élément que l'on souhaite mettre en position de thème. En ce qui concerne la syntaxe, en anglais comme en français, l'ordre canonique dans la phrase est habituellement l'ordre SVO (sujet, verbe, objet), au moins dans un écrit « formel »[1]. Souvent, comme nous venons de le voir, dans les phrases canoniques de ce type, c'est le sujet grammatical qui est couramment associé avec le thème et le verbe et les éléments en position d'objet syntaxique avec le rhème. Toutefois, afin de respecter la structure thématique neutre ou non marquée (thème-rhème), ou au contraire afin de signaler une interprétation marquée (rhème-thème), le rédacteur aura besoin d'employer une syntaxe plus spécialisée. Il s'agit là de structures syntaxiques qui permettent de manipuler l'ordre des éléments dans la phrase. Grâce à ces différents moyens le locuteur est libre de choisir la structuration informationnelle la plus appropriée à ces besoins.

A. Davison (1984) propose pour l'anglais une échelle des marquages syntaxiques du thème (ou *« topic »* chez elle) qui va de la structure active

[1] Il est vrai qu'à l'oral l'ordre est moins rigide (cf. Lambrecht 1987).

canonique aux dislocations et topicalisations, c'est-à-dire des structures les plus usuelles aux plus inhabituelles. Je ne crois pas pour ma part qu'une classification rigide soit utile, car les structure syntaxiques employées sont également très liées à la situation de communication et au genre de texte concerné. Les dislocations à gauche, très courantes par exemple dans le français oral spontané, sont rares à l'écrit. De la même manière, les passifs, fréquents dans les textes scientifiques formels, ne sont guère employés dans une conversation informelle.

Dans un certain nombre de mes travaux antérieurs, je me suis attachée à décrire et expliquer les usages faits de ces structure syntaxiques spécialisées en fonction des impératifs communicatifs [THOM 01 ; THOM 04c et d ; THOM 05d]. J'ai entrepris toute une série d'études et de comparaisons afin de cerner de plus près l'impact précis de la situation d'énonciation (par exemple, le mode de transmission, les exigences au niveau de la rhétorique et des différentes relations interpersonnelles) sur les stratégies informationnelles adoptées. Je me suis penchée en particulier sur les marqueurs de la structuration informationnelle dans différents genres de productions scientifiques : articles de recherche primaire, articles de vulgarisation scientifique, éditoriaux spécialisés, communications de colloque. Je reviendrai sur le détail de ces travaux dans le chapitre 7. Je propose maintenant de simplement faire un bref inventaire des structures spécialisées concernées.

4.1 Les structures dites spécialisées

4.1.1 Passif(s)

Le passif peut jouer un rôle très important dans la signalisation de la structuration informationnelle. Une fonction majeure de la voix passive, bien que celle-ci figure rarement dans les livres de grammaire, réside dans le fait que le passif permet de faire correspondre la structure thématique et la structure syntaxique. Autrement dit, en employant un passif l'émetteur peut thématiser ce qui serait apparu en position d'objet syntaxique et par conséquent

aurait été normalement rhématique dans une phrase active. La phrase suivante qui vient de mon corpus de copies d'étudiants peut servir d'illustration :

> *(1) In the picture… there is an abundance of straight lines, particularly horizontal, so that frequencies are higher in the vertical domain. Coding techniques can take into account these characteristics.* [THOM 00 : 78]

Dans la première phrase le thème est *« the picture »* et la partie rhématique de la phrase apporte des informations sur la composition de cette image. La deuxième phrase s'enchaîne sur un nouveau thème *« coding techniques »*. Par intuition nous sentons que le vrai thème de cette deuxième phrase, ce sont en fait les caractéristiques de l'image déjà évoquée. L'information pertinente dans ce cas est le fait que *« coding techniques »* permette de se rendre compte de ces divers traits. L'emploi de la diathèse passive aurait permis au scripteur de thématiser l'objet grammatical de cette phrase – *« these characteristics »* (ce dont il parle). Ce qui est communicativement le plus important (et nouveau à cet instant) – *« coding techniques »* – aurait ainsi figuré en position de rhème :

> *(2) These characteristics **can be taken into account** by the use of coding techniques.*

Le titre du paragraphe suivant est effectivement *« Principal coding techniques »* et la réorganisation du poids « informationnel » de la phrase aurait aidé à mieux suivre le fil conducteur du discours.

Néanmoins, malgré ce qui semblerait être l'importance de son rôle dans la structuration informationnelle, le passif est souvent considéré comme lourd et complexe et, dans les manuels de grammaire, son importance sur le plan communicatif est passée sous silence. Dans beaucoup de manuels scolaires et de grammaires destinés aux étudiants en anglais langue étrangère[2], les auteurs ne soulignent que les aspects sémantiques du passif – le fait qu'il permet de ne pas mentionner l'auteur de l'action prédicative – ou, en revanche, ils se concentrent essentiellement sur les aspects formels – comment transformer une phrase active en phrase passive et vice versa. Peu d'ouvrages[3] insistent

[2] Je fais allusion ici à plusieurs grammaires scolaires et manuels toujours populaires en France dont Turk et Kirkman (1982), Spankie (1983), Berland-Délepine (1989) .

[3] Même les grammaires issues du courant énonciatif en France, tout en refusant de voir le passif comme une simple transformation, n'ont accordé que peu d'intérêt à son rôle sur le plan thématique (cf. Lapaire et Rotgé 1992).

sur le rôle communicatif du passif et son importance dans la structuration thématique.

Il est vrai que le passif est relativement peu fréquent par rapport à l'actif. T. Givón (1979) a trouvé entre 4% et 18% de passifs selon le type de texte. Toutefois ce pourcentage, qui n'est déjà pas négligeable, est beaucoup plus élevé dans certains types de textes. Dans les textes scientifiques spécialisés par exemple, le passif peut représenter, selon C.L. Barber (1962) 30% des verbes. Mes propres travaux corroborent ces pourcentages [THOM 01 ; THOM 04d ; THOM 05d]. Néanmoins, dans les textes scientifiques, l'emploi du passif est déconseillé par les manuels spécialisés (cf. Day 1995). L'importance de son rôle dans la structuration thématique ne semble pas avoir été comprise.

4.1.2 Énoncés existentiels

Les énoncés existentiels, dont la structure avec *« there is/there are »* est la plus répandue en anglais, permettent à un rédacteur de présenter des informations nouvelles. Ce sont des structures fréquentes au début d'un paragraphe dans l'anglais formel de type expositoire :

(3) There are two fundamental classes of traffic: guaranteed traffic and best-effort traffic.

Dans cet exemple, le thème *there* est sémantiquement vide, mais sa présence permet d'introduire de nouvelles informations dans la partie rhématique de la phrase, et donc de respecter l'ordre donné-nouveau, l'ordre non-marqué d'un point de vue thématique. Une fois que le rédacteur a présenté cette nouvelle information, il est ainsi libre de la choisir comme thème dans les phrases ou paragraphes suivants.

Toutefois, l'utilité de la structure existentielle sur le plan thématique n'est pas évoquée dans les grammaires et manuels scolaires. Dans l'enseignement de l'anglais langue étrangère, les exemples donnés avec *« there is/ there are »* sont généralement caractéristiques de l'oral. La structure est souvent associée à l'emploi d'un locatif pour décrire les alentours matériels (*« There is a tree in the left-hand corner of the field. It is an apple tree »*). La fonction discursive de

la structure est passée sous silence. Toutefois, comme mes propres travaux le démontrent, cette structure peut jouer un rôle très important, non seulement dans la présentation de nouveaux référents mais aussi dans le maintien de la continuité textuelle [THOM 01 : 27-28].

4.1.3 L'extraposition

Davison (*op. cit.*) n'inclut pas l'extraposition dans sa liste des structures syntaxiques marquées. Si je tiens à l'inclure dans cette inventaire des structures syntaxiques spécialisées, c'est parce que l'emploi de l'extraposition [THOM 01 ; THOM 04d] semble jouer un rôle particulièrement important dans la rédaction des articles de recherche scientifique.

Dans l'extraposition, la phrase commence par une proposition avec *it* qui joue un rôle cataphorique par rapport à la totalité de la proposition qu'il précède :

(4) It is evident that the greatest variation occurs amongst occytes.

Sur le plan thématique, le thème de la proposition principale (*« it »*) est ainsi initialement sémantiquement vide, son contenu se trouvant inclus dans la position extraposée. Cette structure permet ainsi de pousser les éléments plus complexes (et nouveaux pour le lecteur) vers la fin de la phrase, respectant ainsi le principe de *« end-weight »* (Quirk *et al.* 1985). En revanche, une version non « extraposée » de cet exemple semblerait beaucoup moins « naturelle » car cela obligerait le locuteur à surcharger la portion thématique de la phrase :

(5) That the greatest variation occurs amongst occytes is evident[4]*.*

Cette utilité des structures expositives sur le plan informationnel est rarement mise en évidence dans les grammaires.

[4] Exemple pris de Banks 1995.

4.1.4 Structures détachées

J'ai déjà évoqué le cas des structures détachées dans le chapitre précédent. Les GN détachés peuvent être considérés comme des thèmes/topiques privilégiés, dans la mesure où ils sont le résultat d'une opération de topicalistion très explicite. Dans des travaux précédents j'ai identifié trois grandes catégories de structures détachées : les détachements avec reprise pronominale, les détachements sans reprise pronominale et les détachements avec terme introducteur [THOM 04d] :

(6) The beam, (…) it's still slower.

(7) The repetitions rates, we've been up to a kilohertz…

(8) Moving on to the carbon fibre cathodes, these are… [THOM 04d : 157-158]

Je reviendrai sur la fonction précise des ces différentes types de détachement dans le cadre des communications scientifiques de colloque dans le chapitre 7, mais je propose ici de faire quelques brèves remarques sur la spécificité de ces structures par rapport à une structure SVO canonique.

En français comme en anglais, un thème continu est généralement marqué par un pronom clitique. Les structures détachées, en revanche, permettent de changer de thème ou de réactiver un référent topical devenu inactif sous une forme lexicale. Comme le souligne K. Lambrecht, il est difficile d'introduire un référent et d'en parler en même temps. Il s'agit du principe de *« separation of reference and role »* (1994 : 185). Pour le récepteur, le décodage est également plus facile si l'identification même du topique est indépendante de l'opération de prédication établie à son propos. Les structures détachées permettent justement cette séparation des domaines de la référence et de la relation, en établissant le thème/topique hors de la proposition elle-même. Il est vrai que cette contrainte sur le cumul des fonctions présentatives et prédicatives et moins forte à l'écrit. Un lecteur peut se permettre de ralentir sa lecture ou d'effectuer des retours en arrière. Pour cette raison ces structures sont beaucoup plus fréquentes à l'oral qu'à l'écrit (cf. chapitre 7 *infra*).

Le cas des structures détachées à droite n'a pas encore été évoqué (« Il est fou, Jean »). Tout en faisant partie de la panoplie de structures spécialisées à la disposition des locuteurs, la fonction des détachées à droite est un peu différente de celle des détachées à gauche. Alors que les détachements à gauche correspondent à l'établissement d'un topique, les détachements à droite signalent au contraire le maintien ou la confirmation d'une relation topicale déjà établie. Autrement dit, ils ont un rôle de confirmation tandis que les détachements à gauche permettent une identification anticipée du référent. On peut ainsi considérer que ces deux constructions contribuent d'une manière différente, mais complémentaire, à la clarification du message.

4.1.5 Structures hypotaxiques

Toute question relative à la notion de hiérarchisation des propositions dans la phrase est également particulièrement importante dans une analyse de la structure informationnelle. Il convient de distinguer entre les phrases composites, c'est-à-dire des phrases contenant une ou plusieurs propositions coordonnées, et les phrases complexes contenant des propositions subordonnées. Si, d'un point de vue syntaxique, nous avons affaire dans les deux cas à une unité grammaticale – la phrase, du point de vue thématique l'analyse ne sera pas la même.

Dans le cas des énoncés composites, l'analyse thématique permet de dégager soit deux ou plusieurs axes thèmes-rhèmes « simples » :

> *(9) Goethe wrote the second part of Faust after eighty, and V Hugo astounded the world with Torquemada at eighty.* (Daneš 1974 : 117)

Soit un seul axe thème-rhème avec un thème ou un rhème multiple, comme en (10) et (11) respectivement :

> *(10) The melting of solid ice and the formation from ice of liquid water exemplify physical changes.*
>
> *(11) It is further postulated that the activated amino acids are joined together… and that the long chains are molded in a specific manner.* (*Ibid.*)

En revanche, dans le cas des énoncés complexes, des éléments subordonnés parenthétiques sont assimilés à une partie de la phrase de base. Une proposition est enchâssée dans une autre phrase, soit dans une partie thématique, soit dans une partie rhématique :

> *(12) This dark-coloured liquid, known as crude petroleum or crude oil, is obtained from wells of different depths.*
>
> *(13) The amino acids are required for making proteins, consisting of long chains of these units.* (*Ibid.*)

En fin de compte, même s'il est possible de décomposer la phrase complexe et d'identifier plusieurs liens thème-rhème à l'intérieur de la phrase, les énoncés complexes n'ont qu'un seul noyau T-R principal. La progression thématique sera établie à partir du thème de ce noyau. En revanche, les énoncés composites peuvent avoir plus qu'un seul centre thème-rhème.

Comme je l'ai constaté dans des travaux précédents [THOM 98, THOM 99b et c ; THOM 00], la simplification syntaxique ne va pas forcément de pair avec une structure thématique lisible. Une série de propositions simples ou une série de phrases comportant plusieurs propositions coordonnées ne permettent pas la création d'une structure hiérarchisée de référents (Lautamatti 1987). Une multiplication de thèmes potentiels peut même nuire à la compréhension d'un texte donné. En revanche, sur le plan de la structure thématique, l'hypotaxe permet de créer une hiérarchie de thèmes.

4.1.6 Constructions clivées

On présente souvent la phrase clivée comme la structure spécialisée par excellence, dans la mesure où elle permet de pallier un ordre de mots contraint par la grammaire. Elle est aussi associée par un bon nombre de travaux en linguistique textuelle à une opération de rhématisation. Le particule *c'est* en français (ou *it is/they are* en anglais) sert à extraire un constituant de la phrase en lui donnant « la garantie et l'exclusivité du premier plan » (Le Goffic 1993 : 221). Ce constituant ainsi mis en valeur dans le rhème de la

première partie de la phrase peut ensuite remplir le rôle de thème dans la proposition qui suit :

(14) C'est Obélix... qui est tombé dans la potion magique (Péry-Woodley 1993 : 71)

↑ ↑ ↑
Rh. Th.

Le rôle de ces structures sur le plan thématique est ainsi très important car, comme les énoncés existentiels, elles permettent l'introduction de nouvelles informations dans la première partie de la phrase, lieu habituellement réservé aux informations connues. Les informations suivant la copule sont généralement connues du récepteur. Selon W.J. Vande Kopple (1986) :

> *When writers use this form, the urgency to convey the new information justifies expressing the new before the given, especially since the given is usually very nearly apparent.* (1986 : 60)

Toutefois, l'usage des clivées dans le discours peut s'avérer complexe. Il est important aussi de distinguer entre les différentes sortes de clivées. Dans la tradition linguistique anglo-américaine, on distingue généralement trois types de phrases clivées : les phrases clivées simples (*« it-cleft »*), les phrases pseudo-clivées (Wh-clefts) et les pseudo-clivées inversées (*« reverse pseudo-clefts »*). Ces trois types de clivées peuvent remplir différentes fonctions sur le plan discursif [THOM 04d]. Je reviendrai dans le chapitre 7 (section 7.2) sur le cas spécifique des pseudo-clivées en *« wh »* dans l'anglais oral, et dans les communications de colloque en particulier. Il existe aussi de subtiles différences dans l'exploitation des phrases clivées en anglais et en français que je détaillerai dans la section suivante (4.2)

4.2 Le cas des phrases clivées : comparaison anglais/français

Dans la dernière partie de ce chapitre consacrée au chassé-croisé de la syntaxe et de la structure informationnelle, je propose de m'attarder sur mes travaux sur le fonctionnement des phrases clivées « simples » en anglais et en français [THOM 02, THOM 09a]. Bien que l'anglais et le français soient très proches sur le plan typologique, les structures spécialisées ne sont pas forcément identiques sur le plan formel et elles ne sont pas nécessairement employées non plus dans les mêmes circonstances. Pour peu qu'on ait quelque

connaissance des deux langues, on s'aperçoit que les phrases clivées ne fonctionnent pas de la même façon. Les phrases clivées en « c'est » sont nettement plus fréquentes en français que ne le sont les *« it-clefts »* en anglais, à l'écrit comme à l'oral.

Une de mes motivations dans la comparaison des clivées en anglais et en français a été de cerner plus précisément leur rôle dans la structuration informationnelle, d'un point de vue didactique. Les difficultés éprouvées par bon nombre de mes étudiants francophones dans la création d'un discours proche de celui d'un natif, résident en grande partie au niveau informationnel. Même s'ils ont une bonne connaissance de l'anglais (et font peu d'erreurs graves au niveau grammatical), les étudiants ont tendance à reproduire certains schémas syntactiques du français, au lieu d'employer des stratégies syntactiques plus adaptée au contexte communicatif du point de vue de la structure informationnelle dans la langue cible. J'ai relu récemment un exposé philosophique très dense, écrit en anglais par un collègue universitaire français. Dans son article de vingt pages, il se trouvait pas moins de treize *« it-clefts »*, dont il ne restait plus que cinq après un travail sérieux de réécriture du texte. Ce sont ces types d'erreurs qui contribuent souvent aux maladresses et au ton « non natif » du discours d'apprenants d'anglais avancés.

Afin de vraiment comparer le rôle des phrases clivées dans la structuration informationnelle dans les deux langues, il m'a semblé important de fonder l'analyse sur un corpus translangagier le plus homogène possible (en ce qui concerne le contenu propositionnel). Pour cette raison j'ai décidé d'orienter ma recherche vers un corpus de traductions.

4.2.1 Corpus et méthodologie

Ce corpus se composait d'une série d'articles tirés du journal *Le Monde* qui était ensuite repris et traduits en anglais dans le journal hebdomadaire *The Guardian Weekly*. Les articles d'origine et les traductions étaient ainsi quasiment identiques sur le plan de leur contenu propositionnel. Dans le

dépouillement de ce corpus j'ai pris soin de distinguer entre les phrases clivées en « c'est » et quelques constructions superficiellement semblables telles les propositions relatives à fonction restrictive. Dans le cas des propositions subordonnées, l'accord morphologique peut parfois jouer un rôle discriminant en français, car les accords sont rares dans la clivée (Molinier 1996 : 86) :

(15) C'est une élégie que j'ai écrit la semaine dernière (phrase clivée)

(16) C'est une élégie que j'ai écrite la semaine dernière (proposition relative).

Cependant, dans la plupart des cas, l'identification positive d'une clivée nécessite la prise en compte de l'environnement co-textuel complet.

À la suite de ce processus j'ai pu identifier trente-six phrases clivées en français. Ensuite j'ai analysé les traductions anglaises de ces articles afin de voir dans quelle mesure la structure clivée était maintenue dans la traduction. Sur les trente-six clivées d'origine, seulement dix-sept (47%) étaient également traduites par une clivée dans la version anglaise. Dans tous les autres cas, le traducteur a préféré employer une forme sans « clivée ». Il est certain que la question de l'interférence doit jouer un rôle. Le traducteur risque peut-être d'être influencé par le texte source. Mais, l'inverse peut aussi être vrai. Dans une tentative de ne pas reproduire la forme du texte source, le traducteur peut aussi décider de réduire l'usage de la clivée dans sa traduction[5]. Malgré ces restrictions et malgré la taille assez réduite de mon corpus, je crois néanmoins que mes résultats permettent une réflexion utile sur la distribution des phrases clivées dans les deux langues.

Mon étude était entreprise dans le cadre de la linguistique systémique. Dans la tradition systémique, l'analyse des clivées est étroitement imbriquée avec la notion de thématisation. Dans cette approche, le « *cleft sentence* » est renommé justement « *predicated theme* » (Halliday 1994) ou « *enhanced theme* » (Fawcett et Huang 1995). Selon M.A.K. Halliday (1994), la prédication thématique permet au locuteur d'extraire et de mettre en relief tout constituant à

[5] Il serait ainsi très intéressant de comparer les résultats obtenus ici avec un corpus de traduction anglais/français pour voir si l'on trouve l'image inversée (cf. Johansson 2001).

valeur idéationnelle. Il attribue à cette construction une structure thématique double : un niveau « local » et un niveau propositionnel. Au niveau local, l'élément extrait reçoit une proéminence rhématique sur les plans syntaxiques et prosodiques[6]. Toutefois, au niveau propositionnel, la phrase principale constitue le thème et la proposition subordonnée constitue le rhème :

	It	*was his teacher*	*who*	*persuaded him to continue*
a) *local thematic structure*	*Theme*	*Rheme*	*Theme*	*Rheme*
b) *thematic structure of a predicated theme*	*Theme*		*Rheme*	

Figure 1. Structure thématique d'une phrase avec un thème prédiqué.
D'après Halliday (1994 : 60).

Comme on l'a vu dans le chapitre 2, l'analyse informationnelle dans l'approche systémique est divisée en deux partie distinctes : d'une part une analyse en termes de « thème et de rhème », d'autre part une analyse en termes de « donné et de nouveau » (cf. sections 2.2 *supra*). Si en règle générale la coïncidence entre « thème » et « donné » peut être considérée comme le cas par défaut, avec un thème prédiqué la réalité est tout autre. Halliday affirme même qu'une fonction de cette construction est précisément de faire coïncider le focus avec le thème

> *It is precisely the function of this system to align Theme + Rheme with Given + New in such a way that focus falls on the Theme; this makes the Theme New and the Rheme Given.* (1994 : 302)

L'emploi de la construction permet ainsi une « déviation » par rapport à l'ordre non marqué. Les « nouvelles » informations apparaissent avant les informations données.

Toutefois, l'analyse informationnelle des clivées et du type de « nouveauté » en question est complexe. Comme de nombreux linguistes l'ont fait remarquer (par exemple, Banks 1999 ; Collins 1991 ; Lambrecht 2001), le clivage peut porter sur des pronoms : des éléments nécessairement contextuellement donnés. Dans sa

[6] La proéminence tonique se trouvant en règle générale à la fin du l'unité informationnelle locale (Halliday 1994 : 298).

présentation de l'opposition donné/nouveau, Halliday insiste néanmoins sur le fait que le « nouveau[7] » n'est pas forcément nouveau dans un sens littéral. En revanche, le nouveau est ce qui est présenté comme n'étant pas disponible ou « non récupérable », soit afin de le mettre en contraste avec d'autres informations, soit tout simplement afin d'attirer l'attention du récepteur sur quelque chose de déjà connu (Halliday 1994 : 299-302 ; Banks 2005). Halliday reconnaît ainsi la possibilité pour les informations données ayant déjà figuré dans le discours antérieur d'être présentes en tant que « focus ». Si, par exemple, un pronom anaphorique est focalisé, il sera considéré comme « nouveau ». La décision du locuteur quant à la présentation d'un élément en tant que « donné » ou « nouveau » prime ainsi sur la « recouvrabilité » de la référence.

Ce qui importe finalement dans un examen des constructions clivées, n'est pas le degré de nouveauté d'un référent mais celui de la nouvelle répartition de la structuration informationnelle, provoquée par la réorganisation de la phrase. L'élément qui est clivé est mis en relief (bien qu'il ne soit pas forcément le seul élément dans la phrase à bénéficier d'une emphase), ce qui a des répercussions sur toute l'interprétation de la phrase en question. L'emploi d'un clivée permet au locuteur de diriger l'attention du récepteur vers une interprétation particulière. Voici un exemple tiré de mon corpus d'articles journalistiques qui permet d'expliciter certains points que je viens d'évoquer. Il s'agit d'une phrase clivée en français (17) qui était également traduite par une clivée dans la traduction anglaise (18) :

> (17) À la veille de la révolution bolchevique, cette vraie Russie et son intelligentsia libérale rejetaient l'antisémitisme ; le pouvoir tsariste était lui-même sur le point, affirme Soljenitsyne, de lever toutes les restrictions civiques, sociales, éducatives qui pesaient encore sur la population juive.
>
> Et <u>c'est à ce moment-là</u> qu'une déflagration mit en pièces le système politique et social de la Russie.
>
> *(18) And <u>it was at that point</u> that an explosion destroyed Russia's political and social system.*

[7] Dans la présentation systémique il me semble cependant que l'emploi du terme *nouveau* prête à confusion.et qu'il est préférable d'employer l'appellation de *newsworthty* (Fries 1994) ou de *focalisé* (Banks 2004 ; 2005) pour cette fonction.

Cet extrait est issu d'un entretien avec A. Soljenitsyne sur les rapports entre les Juifs et les Russes à la veille de la révolution bolchevique. Le point dans le temps auquel se réfère l'auteur (« Et c'est à ce moment-là ») a déjà été défini dans les phrases précédant immédiatement la clivée**.** Il s'agit d'une information qui est donc facilement « récupérable » pour le lecteur. Toutefois, en plaçant l'expression adverbiale de temps au sein de la clivée, cet élément est signalé comme en étant particulièrement digne de l'attention du lecteur. C'était à ce moment dans le temps et non pas à un autre moment (ce qui est un peu ironique, étant donné que l'antisémitisme n'était pas répandu) qu'a eu lieu la déflagration ou révolution. Il n'y a pas de contraste explicite. En revanche, il existe un contraste implicite avec ce qu'on aurait pu attendre. Comme l'explique Halliday : *« The meaning is attend to this: this is news. »* (1994 : 228)

Tout comme Halliday, Lambrecht souligne la distinction entre le concept de nouveauté référentielle et l'indication de focus. Pour lui aussi le focus est de nature imprévisible et « non récupérable ». Je reprends ici une partie de sa définition du focus déjà présentée dans le chapitre 2 :

> *The focus is that portion of a proposition which cannot be taken for granted at the time of speech. It is the unpredictable or pragmatically non-recoverable elements in an utterance.* (1994 : 211)

Cependant, Lambrecht met aussi l'accent sur la nature relationnelle du concept de focus. L'information ne peut pas être sectionnée en compartiments individuels. Ainsi il ne s'agit pas d'un trait intrinsèque à un signifié mais d'une relation entre un signifié et une proposition (Lambrecht 2004). Le vrai focus de la phrase (17) ci-dessus ne réside pas dans l'élément clivé lui-même mais dans les rapports créés entre cet élément et les informations véhiculées par la proposition subordonnée. Toutefois, sans la présence de la clivée une lecture focalisée de l'expression adverbiale est difficile à indiquer.

Dans une version non clivée de cet exemple dans les deux langues (17b et 18b), l'expression adverbiale initiale serait normalement considérée comme une information « donnée » constituant l'arrière-plan :

?(17b) A la veille de la révolution bolchevique, cette vraie Russie et son intelligentsia libérale rejetaient l'antisémitisme ; le pouvoir tsariste était lui-même sur le point, affirme Soljenitsyne, de lever toutes les restrictions civiques, sociales, éducatives qui pesaient encore sur la population juive.

Et à ce moment-là une déflagration mit en pièces le système politique et social de la Russie.

?(18b) And at that point an explosion destroyed Russia's political and social system.

Dans la version non clivée en anglais (18b), un lecteur vigilant pourrait éventuellement conférer une valeur de focus et (de contraste implicite) à l'expression adverbiale. En revanche, dans la version non clivée en français, la locution adverbiale semble seulement constituer la toile de fond devant laquelle l'action (la révolution) s'est passée. Sans la clivée, les rapports entre l'adverbe temporel et le reste de la proposition ne sont pas explicites.

Dans une autre variante sans clivée, les locutions adverbiales sont placées après le verbe, et dans cette position feront normalement partie du rhème (17c et 18c) :

?(17c) A la veille de la révolution bolchevique, cette vraie Russie et son intelligentsia libérale rejetaient l'antisémitisme ; le pouvoir tsariste était lui-même sur le point, affirme Soljenitsyne, de lever toutes les restrictions civiques, sociales, éducatives qui pesaient encore sur la population juive.

Et une déflagration mit en pièces le système politique et social de la Russie à ce moment-là.

?(18c) And an explosion destroyed Russia's political and social system at that point.

En théorie, il serait envisageable d'attribuer ainsi une valeur de focus et de contraste à l'adverbial en fin de phrase. Toutefois, l'accumulation des groupes rhématiques rend l'interprétation de la phrase difficile : la hiérarchie des informations ne peut se dégager facilement**.** Cette variante est très difficile à lire et à interpréter dans sa version française et anglaise. L'utilisation de la structure clivée dans les formulations d'origine semblerait dans ce cas être le seul moyen de fournir des consignes claires au lecteur quant à la délimitation précise des différentes composantes informationnelles de la proposition.

Dans ces travaux sur les phrases clivées, je défends la thèse que la motivation principale pour l'emploi de la forme clivée est celle de désambiguïsation. En choisissant d'employer une phrase clivée, un locuteur fournit des consignes de

traitement plus précises que dans une version équivalente non clivée. Les effets sémantiques couramment associés avec l'emploi de clivées, tels la spécification ou la contrastivité, sont dans cette perspective plutôt un résultat du clivage que sa raison d'être. Mais comment alors expliquer le plus grand recours aux phrases clivées en français qu'en anglais ?

Dans les deux langues la phrase clivée permet d'échapper d'une certaine manière aux contraintes d'un ordre de mots relativement rigide (cf. section 3.3 *supra*). O. Jespersen a même proposé la généralisation suivante :

> *…this construction may be considered one of the means by which the disadvantages of having a comparatively rigid grammatical word order (SVO) can be obviated. This explains why similar constructions are not found, or are not used extensively, in languages in which the word order is considerably less rigid than in English, French, or the Scandinavian languages.* (1937 : 85)

Toutefois, il existe d'autres moyens de « compensation » – des moyens syntaxiques, prosodiques ou morphologiques. Afin de vraiment comparer le fonctionnement des clivées dans les deux langues, il est donc important d'évaluer dans quelles mesure les avantages apportés par l'emploi de la clivée peuvent aussi être obtenus grâce à d'autres moyens linguistiques.

4.2.2. Traduction des phrases clivées en anglais

Sur les trente-six clivées du *Monde*, seules dix-sept (47%) étaient également traduites par une clivée dans la version anglaise du *Guardian*. La proportion de clivées maintenues dans la traduction anglaise semblerait aussi être directement liée à la catégorie grammaticale de l'élément clivé. Bien que tout élément idéationnel puisse en principe faire l'objet d'une thématisation ciblée à l'aide d'une clivée, Collins (1991), dans son étude de la clivée anglaise, a trouvé que dans la pratique la procédure de clivage concerne essentiellement les sujets syntaxiques et les adverbiaux (75% des cas). Dans sa thèse sur la phrase clivée française, S. Katz (1997) a trouvé des résultats semblables. Cette préférence peut selon P. Collins être considérée comme une conséquence de la question de proéminence (ou saillance) thématique :

…from a thematic perspective, the role of the cleft construction is to add to an already thematic item the further prominence deriving from predication in the case of the cleft construction. (1991 : 64)

Comme les GN sujets et les groupes adverbiaux sont les thèmes les plus « naturels » dans les deux langues, il n'est donc pas étonnant de les retrouver comme objets de clivage. Or d'après mes travaux, dans la distribution précise de ces deux catégories de clivées, l'anglais et le français ne se comportent pas de la même façon. Le tableau ci-dessous reprend ces données contrastives :

	Phrase clivée d'origine	Maintien de la clivée en anglais	Autre formulation en anglais
Sujet	15	4	11
« Adjunct »[8]	19	12	7
Objet	2	1	1
Total	36	17	19

Tableau 1. Traduction des clivées et catégorie grammaticale [De Thom 09a].

Dans la rubrique *« adjuncts »*, la majorité des phrases clivées portant sur des expressions adverbiales (en particulier des expressions adverbiales temporelles) dans la version anglaise d'origine, sont également maintenues dans la traduction anglaise. Le français et l'anglais sembleraient toutes les deux souvent avoir besoin d'employer une clivée dans ce cas. En revanche, si l'on ne retient que la catégorie grammaticale du sujet, on constate que les clivées françaises dans leur quasi-majorité ne sont pas maintenues dans la version anglaise. Le traducteur anglais emploie une autre formulation.

Dans la plupart des clivées portant sur les sujets grammaticaux, l'élément clivé dans le texte d'origine apparaît comme un « simple » thème dans la traduction anglaise. Très souvent le contexte (et le co-texte) semblent suffisants pour indiquer au lecteur la valeur focale d'un tel élément en position préverbale. Considérons l'exemple suivant extrait d'un article traitant des mariages « arrangés » au sein de familles d'immigrés en France.

[8] La catégorie « adjunct » dans la tradition systémique comprend essentiellement les adverbiaux de circonstance (Banks 2004).

(19a) La mesure du phénomène reste délicate. Il n'existe aucune statistique officielle. Pour Emmanuelle Piet, ce sont[9] la quasi-totalité des jeunes Turques, des Africaines du fleuve Sénégal et un grand nombre de jeunes Maghrébines qui sont menacées.

(19b) La mesure du phénomène reste délicate. Il n'existe aucune statistique officielle. Pour Emmanuelle Piet, la quasi-totalité des jeunes Turques, des Africaines du fleuve Sénégal et un grand nombre de jeunes Maghrébines sont menacées.

Dans (19a) l'élément clivé, le sujet de la proposition canonique correspondante, concerne l'identification des populations qui risquent d'être touchées par les mariages arrangés. Le fait que le risque existe est déjà signalé : ce qui a valeur d'actualité est l'identification précise de la population touchée : la « *quasi-totalité »* des jeunes Turques, les filles africaines originaires de *Sénégal* et *un grand nombre* de filles nord-africaines. La structure clivée est nécessaire pour diriger l'attention du lecteur vers l'importance de ces éléments. Comme Lambrecht l'a fait remarquer, il existe en français une contrainte forte sur l'appariement entre focus et sujet (1994 : 22). Autrement dit, le français a souvent recours à une structure clivée afin d'éviter un focus en position initiale[10]. Bien que le commentaire de Lambrecht porte sur le français oral, je pense que de telles considérations sont également pertinentes pour le français écrit (et expliquent en partie pourquoi la forme clivée en « c'est » est plus utilisée que la forme anglaise équivalente). Sans la structure clivée, le lecteur serait enclin à attribuer une valeur « d'actualité » ou de focus au prédicat *sont menacées*, dans la partie rhématique de la phrase (voir 19b). En extrayant un élément pour le mettre en exergue, l'énonciateur le privilégie, mais atténue du même coup la valeur des autres éléments.

Or, dans la traduction anglaise (20a) de cette partie de l'article, la forme clivée n'est pas employée.

(20a) The extent of such practices is difficult to evaluate. There are no precise figures. In Dr. Piet's view, almost all Turkish, Senegalese and a large number of North African girls are at risk.

[9] Notons en passant que la forme plurielle « ce sont » est employée au lieu de « c'est », même si cet accord tend à disparaître dans les genres informels.

[10] Un élément extra-prédicatif peut plus facilement recevoir une focalisation (Lambrecht 1994 : 31). Caffarel (2000) fournit également des exemples de thèmes détachés focalisés.

La clivée n'est pas nécessaire. L'anglais est beaucoup plus souple par rapport au positionnement de l'accent focal que le français. On s'attend à ce que le lecteur anglophone accorde naturellement un statut focal aux expressions fournissant des détails sur la population à risque. En employant une forme clivée (voir 20b), le traducteur aurait pu atteindre le même objectif sur le plan pragmatique, mais la présence supplémentaire de la forme clivée *it* serait d'une lourdeur excessive.

> *?(20b) In Dr. Piet's view, it's almost all Turkish, Senegalese and a large number of North African girls who/that are at risk*

Le lecteur peut recréer mentalement la structure informationnelle de la proposition sans indice grammatical supplémentaire.

D'après mes recherches, les clivées portant sur les adverbiaux sont plus susceptibles d'être maintenues dans la traduction anglaise que les GN sujet (douze cas sur dix-neuf). Dans le cas contraire, on peut retrouver l'élément clivé en français dans la partie rhématique de la phrase anglaise :

> (21) Cette notoriété aiguillonne les chercheurs. Dans le monde anglo-saxon, d'abord. C'est aux Etats-Unis que la redécouverte de Le Gray a eu lieu.
>
> *(22) His rise to fame has stimulated a great deal of research, initially in English-speaking countries. Le Gray was first rediscovered in the United States.*

Dans cet extrait d'un article portant sur photographe G. Le Gray, l'adverbial spatial mis en valeur dans la clivée du français original (21) apparaît dans la position rhématique dans la traduction anglaise (22). Toutefois, l'absence de clivée ne semblerait pas avoir d'effet néfaste sur la clarté ou la précision de la traduction. Comme l'identité de Le Gray et sa « redécouverte » en tant que telle sont déjà connues par les lecteurs, l'indication du lieu de cette découverte se révèle progressivement comme le focus de la phrase. Le traducteur aurait pu employer une clivée dans ce cas :

> *?(22b) It was in the United States that Le Gray was first rediscovered.*

S'il ne l'a pas fait, c'est sans doute parce qu'il estimait que le même effet communicatif pourrait être atteint par l'usage d'une structure comportant l'élément mis en valeur dans sa position canonique

Pourquoi en est-il ainsi ? Je pense que l'on peut envisager l'hypothèse que la structure clivée avec *it* en anglais est plus « marquée » ou inhabituelle que la clivée française en *c'est*. Comme nous l'avons vu, le français a recours aux clivées plus fréquemment, en partie à cause de sa structure de focus plus rigide ; et de ce fait la structure peut apparaître comme moins marquée qu'en anglais. Si le traducteur anglais avait systématiquement clivé les éléments mis en valeur dans les textes d'origines – faisant d'eux non seulement le focus mais également le thème – ces éléments auraient peut être acquis plus de proéminence que l'auteur du texte original ne l'aurait souhaité.

Comme c'est souvent le cas dans les corpus de langage naturel, mon corpus contrastif de clivées contient quelques contre-exemples. J'ai trouvé un petit nombre d'exemples où la forme clivée disparaît dans la traduction en anglais, mais où la structuration informationnelle du texte source ne semble pas être respectée. Cela est peut-être dû en partie à une certaine interférence linguistique ou à un certain « bruit » engendré par le processus de traduction lui-même. Le traducteur journalistique est souvent contraint de respecter des délais très serrés, et il n'est pas censé non plus produire un chef d'œuvre littéraire. C'est ainsi tout à fait possible que l'absence de certaines formes clivées dans la traduction anglaise corresponde à des oublis – les nuances du texte source étant supprimées – ou qu'elles n'aient même pas été remarquées. Dans l'extrait suivant, par exemple, l'absence de la clivée dans la traduction anglaise rend la phrase beaucoup moins claire.

> (23) Ironie de l'histoire, c'est le 11-Septembre que les pays membres de l'OEA, dont les Etats-Unis, ont signé une nouvelle charte démocratique prévoyant des mécanismes régionaux de défense des régimes constitutionnelles.
>
> *(24) By a curious twist of history OAS member countries, including the US, signed a new Inter-American Democratic Charter on September 11 2001, which provided for regional mechanisms to defend constitutional governments.*

Dans le texte original en français, l'expression adverbiale « le 11-Septembre » est présentée comme départ thématique et focus de la clause. Dans la traduction anglaise cependant, le groupe nominal *« OAS member countries »* est présenté comme thématique et l'expression adverbiale occupe seulement le

créneau non marqué rhématique à la fin de la proposition. Son positionnement ici, à distance de *« a curious twist of history »*, n'est pas seulement maladroit, mais il occulte l'ironie que le journaliste voulait exprimer. Une traduction comportant une clivée en *it* aurait été plus efficace.

> *(24b) By a curious twist of history, it was on September 11 2001 that OAS member countries, including the US, signed a new Inter-American Democratic Charter, which provided for regional mechanisms to defend constitutional governments.*

L'analyse d'un corpus restreint m'empêche de tirer des conclusions trop hâtives, mais elle a pu faire émerger un certain nombre de tendances générales. La clivée en *it* en anglais n'est en aucune façon aussi cruciale ou essentielle que la clivée *c'est* en français. Bien que parfois les deux langues puissent avoir besoin d'employer des structures clivées (tout particulièrement dans le cas des phrases clivées portant sur des expressions adverbiales), il existe des situations où la clivée anglaise est facultative, et d'autres où il vaut mieux l'éviter. Autrement dit, il n'y a pas de correspondance simple entre la clivée *c'est* et la clivée *it*. Dans l'une et l'autre langue, si la clivée est employée dans un contexte inapproprié, ou n'est pas employée quand c'est nécessaire, l'énoncé résultant, bien que compréhensible, aura l'air étrange pour un locuteur natif. Il est important d'attirer l'attention des apprenants sur ce fait. Il faut enseigner les structures syntactiques spécialisées en tenant compte de leur rôle dans la structuration informationnelle, et non pas seulement comme des variantes ou comme des manipulations grammaticales astucieuses.

Partie C

Pour une approche empirique

Chapitre 5 – Travailler sur corpus

Résumé

Dans le cinquième chapitre, je décris le développement des études sur corpus, en précisant par la suite l'utilisation que je fais des corpus dans mes propres travaux.

Je commence par une réflexion sur l'objet « corpus » (5.1.), passant en revue les différentes acceptations et définitions qui lui sont habituellement associées. J'évoque certains problèmes impliqués dans la constitution d'un corpus : la taille des corpus, leur numérisation, leur représentativité, ainsi que de la différence entre corpus et base de données.

Je passe ensuite à une discussion des différentes approches et méthodologies employées actuellement dans les analyses sur corpus (5.2). Je me penche sur certains enjeux liés à l'emploi d'une méthodologie exhaustive et quantitative, et sur l'apport des méthodes informatisées. J'évoque également certaines incompatibilités potentielles entre les objectifs de la linguistique textuelle et de la linguistique de corpus.

Dans la partie (5.3), je décris l'emploi que je fais des corpus dans mes propres travaux en me focalisant tout particulièrement sur leur utilisation indispensable dans l'analyse de textes spécialisés. Je fais part de certaines difficultés inhérentes à l'assemblage et à l'analyse des corpus utilisés dans mes travaux en anglais scientifique, et de la nécessité de recourir fréquemment aux spécialistes du domaine.

Chapitre 5 – Travailler sur corpus

Le travail empirique sur corpus occupe une place centrale dans mes travaux. Tout en m'intéressant à différentes configurations syntaxiques à la surface textuelle, il s'agit surtout d'élucider les fonctionnements linguistiques dans leur contexte de production. Une telle approche ne peut pas se contenter des exemples forgés. En effet, le recours au corpus devient essentiel lorsque la problématique avancée implique la prise en compte du contexte. Afin d'expliquer les choix linguistiques opérés par les locuteurs, une démarche empirique s'impose. Il est préférable de partir, non pas des productions artificielles produites par introspection, mais de textes « réels » produits pour des raisons de communication entre les gens.

5.1 Le corpus comme objet d'étude

5.1.1 Choix et constitution

Si le recours aux données réelles n'a rien de nouveau en linguistique, le développement exponentiel de l'informatique et de l'Internet a quelque peu changé la donne. Le linguiste a maintenant à sa disposition des textes en très grand nombre *via* Internet ainsi qu'un éventail d'outils pour les interroger. Le courant appelé *corpus linguistics* a pris un véritable essor dans la linguistique anglo-saxonne au début des années 1990, avec l'émergence de divers groupes de linguistes qui se définissent même par rapport à cette appellation : *to do corpus linguistics*. Dans son emploi dans la linguistique moderne le terme *corpus* peut ainsi se revêtir de toute une gamme d'acceptations[1]. Parmi les définitions souvent citées par les lexicologues de corpus on trouve les suivantes :

> *A collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of language. (Sinclair 1996)*
>
> *A body of text assembled according to explicit design criteria (…) for a specific purpose.* (Atkins *et al.* 1992)

[1] Le titre de l'ouvrage de Habert, Nazarenko et Salem (1997), *Les Linguistiques de corpus,* est parlant à cet égard.

A body of text put together in a principled way for computer processing. (Johansson 1998)

Dans toutes ces définitions de l'objet « corpus », le choix et la visée spécifique du corpus sont deux critères récurrents, qui me semblent d'une importance capitale : un corpus n'est pas un ensemble de données langagières rassemblés en vrac mais le fruit d'un choix ; en outre, la construction d'un corpus ou la sélection d'un corpus à exploiter (dans le cas d'un déjà constitué) dépend des visées de l'étude que l'on va mener (cf. Condamines 2003, 2005 ; Habert *et al.* 1997 ; Martin 1997). Comme l'explique J. Martin :

> Il ne peut y avoir de recherche sans que soit formulée, au préalable, une hypothèse de recherche qui va guider non seulement la constitution du corpus mais sa consultation (…) Le corpus, quelle que soit son origine ou sa taille, ne peut dispenser le chercheur d'un questionnement. (Martin 1997 : 80)

Dans le cas d'un corpus de textes spécialisés, la réflexion sur les critères à mettre en œuvre dans sa construction doit être très élaborée (cf. 5.3 *infra*). Mais pour tout corpus, grand ou petit, général ou spécialisé, il est important d'avoir mené une réflexion poussée sur les objectifs de sa constitution.

Aujourd'hui, il existe des collections de textes électroniques de plus en plus vastes. La capacité de stockage de textes semble en effet transformer les données de recueil de corpus. Lorsqu'il s'agit de corpus qui sont censés être représentatifs d'une langue, des corpus de référence, il existe des corpus de plusieurs millions de mots (par exemple, cent millions de mots pour le BNC[2]). Mais la possibilité d'amasser de très grands corpus ne doit pas dispenser le linguiste d'une réflexion sur la construction du corpus. Selon les termes de M.-P. Péry-Woodley (1995) la maxime « gros c'est beau » n'est pas forcément appropriée. Comme le recueil des textes est désormais aisé, on oublie parfois la finalité. Il est vrai également que dans la communauté du Traitement automatique des langues (TAL), on appelle souvent *corpus* les grandes collections de documents qui servent à mettre au point les traitements (Habert *et al.* 1997). Or la création ou la consultation d'un corpus telle que je la comprends ne devrait pas être un but en soi mais un moyen d'atteindre un objectif précis.

[2] British National Corpus.

La taille des corpus n'est pas un critère exclusif. Tout dépend de l'utilisation qu'on veut en faire. Plus le corpus est important, plus nous risquons de perdre en spécificité. Comme le démontre plusieurs auteurs dans l'ouvrage de M. Ghadessy *et al.* (2001), de petits corpus de vingt mille mots peuvent suffire pour certaines tâches. C'est notamment le cas pour les corpus constitués dans une optique didactique (*ibid.*). C'est également souvent le cas dans l'étude de langues spécialisées, par exemple [THOM 08b et 09b], où les « petits » corpus sont répandus (cf. 5.3 *infra*). Un corpus de langue spécialisée n'a pas besoin d'être d'aussi grande taille qu'un corpus de langue générale. Pour Martin :

> Dès l'instant où c'est la spécificité qui est en question, et non la représentativité comme dans les corpus généralistes, l'étoffement du corpus s'arrête dès que cette spécificité cesse de se manifester de manière significative. (1997 : 79)

La taille du corpus est également à mettre en relation avec la nature et le volume des données envisagées. Pour un phénomène grammatical fréquent, tel que l'emploi de la négation, le nombre d'occurrences relevés sera forcément beaucoup plus important que pour une structure linguistique plus rare, telle que la phrase clivée. Il existe aussi de nombreux cas où l'emploi d'un petit corpus est imposé par nécessité. Dans les langues rares, parlées par peu de locuteurs comme l'inuit [THOM 04a et 05c] ou dans les langues anciennes (Prévost 2005) les données disponibles sont rares et ne peuvent pas forcément constituer de grands ensembles.

Comme nous le rappelle la définition de S. Johansson (1998) citée plus haut, sous l'influence du TAL, on a pris l'habitude de concevoir le corpus comme un corpus électronique : *« A body of text put together in a principled way for computer processing. »* Cependant, comme pour la taille, la numérisation du corpus ne dépend pas forcément du choix du linguiste. Il est possible que les corpus disponibles n'existent pas en version électronique. Les corpus d'oral transcrits et numérisés sont par exemple rares[3], en partie à cause de l'investissement nécessaire à la fois sur le plan technique et humain. Dans le cas de corpus de petite taille, écrit ou oral, la numérisation n'est peut-être pas

[3] *Le nouveau corpus du français parlé parisien des années 2000 (CFPP2000)* est bienvenu à cet égard (Branca-Rosoff, Fleury, Lefeuvre et Pires 2009).

non plus toujours utile. La numérisation reste néanmoins très souhaitable pour toute étude quantitative (cf. *infra)* et évidemment une étape obligatoire avant une annotation du corpus en vue des traitements TAL.

5.1.2 Représentativité

Le critère de représentativité est étroitement lié à la question de visée qui doit motiver le choix de corpus. On peut identifier plusieurs cas de figure selon les possibilités d'étendre ou de généraliser les résultats (cf. Condamines 2005 ; Péry-Woodley 2000). Selon l'objectif de l'étude envisagée, la représentativité du corpus se pose différemment. Dans certains cas, la généralisation des résultats n'est pas l'objectif premier de l'étude. Le corpus est conçu comme une simple source de données pour l'étude d'un phénomène linguistique particulier (l'emploi de l'imparfait ; le positionnement phrastique d'une expression adverbiale etc.). L'analyse est focalisée sur cet aspect et sur les éléments du corpus qui contribuent à le décrire. Il s'agit souvent de corpus de petite taille où l'étude est conçue comme une étude pilote, capable de révéler des tendances mais qui n'a pas d'ambition de généralisation immédiate. Les analyses effectuées sur les copies de mes étudiants rentrent dans ce cas de figure [THOM 00a] ainsi que le corpus de traductions utilisé pour effectuer une comparaison entre le fonctionnement de phrases clivées en anglais et en français [THOM 02 ; THOM 09a]. Pour certains commentateurs (par exemple, McEnery et Wilson 1996), de telles études bien qu'en étant fondées sur corpus, ne relèvent pas du *« corpus linguistics »* au sens strict du terme à cause de cette absence de visée généralisatrice ou de méthodologie quantitative (cf. section 5.2 *infra*).

Dans un tout autre cas de figure, le corpus est envisagé comme représentatif de **la** langue générale (cf. par exemple la définition de Sinclair ci-dessus). Toutefois, répertorier tous les usages de la langue et les mettre au même niveau quantitatif reste un travail peu réalisable. Même si le BNC par exemple, est assez exemplaire à cet égard, contenant une grande variété de situations

de communication, la mise en garde de D. Biber sur le caractère artificiel d'une langue générale en tant qu'entité reste d'actualité :

> *Global generalizations are often not accurate at all, because there is no adequate overall linguistic characterization of the entire language; rather there are marked linguistic differences across registers (or sublanguages).* (Biber 1993 : 220)

Le développement de la *Longman Grammar of Spoken and Written English* (Biber *et al.* 1999) répond à des préoccupations de ce type. Dans cette grammaire de la langue anglaise, les caractéristiques grammaticales de la langue sont évaluées et discutées en fonction de la distribution et de leur fréquence dans quatre grands registres : la fiction, l'anglais académique, l'anglais journalistique et la conversation. Le principe sous-jacent n'est pas de prendre le corpus diversifié dans son ensemble comme représentant la langue générale, mais de considérer chaque sous-corpus séparément, afin d'aboutir à une image composite de la langue.

Une alternative à la quasi-impossibilité de tout représenter est fournie par l'étude de corpus de langue de spécialité – c'est-à-dire l'étude de l'usage de la langue à travers les textes produits par différentes communautés de spécialistes socialement préconstruites : des communautés généralement ancrées dans les corps de métiers ou dans les professions (cf. Lerat 1995). De tels corpus représentent par définition un échantillon restreint des usages de la langue. Se focaliser uniquement sur l'usage de la langue au sein de telles communautés aux contours bien définies, c'est se donner les possibilités d'approcher de plus près de ce qui crée les spécificités linguistiques des usages adoptés. L'idée que les langues de spécialités se définissent sur le seul critère du lexique est depuis longtemps dépassée (Resche 2001). Les langues spécialisées font aussi une utilisation sélective des possibilités offertes par la langue générale. Une des tâches esentielles des chercheurs en langue de spécialité – une tâche qui m'intéresse tout particulièrement – est justement d'identifier ces spécificités syntaxiques.

La notion de genre (cf. chapitre 6) est une notion clé pour le linguiste travaillant sur les textes spécialisés. L'approche que j'adopte dans mes travaux sur les

textes spécialisés distingue en effet les paramètres liés aux domaines de connaissance des communautés de spécialistes de ceux liés aux genres employés par ces communautés. En effet, les communautés de spécialistes ont recours à différents genres textuels pour véhiculer leurs connaissances spécialisées : articles de recherche, articles de vulgarisation, communications de colloques, rapports internes etc. La prise en compte à la fois des contraintes génériques et des contraintes disciplinaires dans la constitution du corpus permet ainsi de créer des corpus très ciblés, qui seront moins soumis à la variation que le corpus de textes tout-venant. De tels corpus font un point de départ solide pour des comparaisons ultérieures sur les usages spécifiques aux genres et sur celles spécifiques aux pratiques disciplinaires.

5.1.3 Corpus *vs* base de données

C'est la question de « visée » qui nous permet aussi de faire une distinction entre un corpus et une base de données. En France, à l'heure actuelle, il n'existe pas de corpus de référence de la langue française, comme est censé être le BNC en Grande-Bretagne. Les CD-ROM du journal *Le Monde* sont très loin de représenter toutes les variétés de la langue française[4] et ne forment pas un corpus dans le sens strict du terme. En revanche, le linguiste peut s'en servir pour forger son propre corpus d'articles de presse. D'une façon analogue l'archive textuelle Frantext, composé essentiellement d'ouvrages littéraires des XIXe et XXe siècles, constitue une ressource très riche pour la linguistique de corpus. La frontière entre un corpus et une base de données n'est donc pas fixe. Tout comme une base de données peut se transformer en corpus, un corpus peut plus tard devenir une partie d'une base de données. Comme le précise C. Marcehello-Nizia :

> Il n'est pas rare qu'un corpus finalisé puisse se transformer par la suite en base de données généraliste offerte à d'autres chercheurs et ouvertes à de nouvelles

[4] Le journal *Le Monde,* en outre, ne représente pas à lui seul toutes les variétés du langage journalistique non plus. Au sein de l'opération EIOMSIT que je coordonne avec S. Prévost dans notre laboratoire, nous avons ainsi décidé pour cette raison d'étudier aussi d'autres registres journalistiques moins formels. Le corpus dont nous nous servons pour notre analyse des éléments initiaux dans le français journalistique contient à la fois des articles tirés d'un quotidien national, *Le Monde,* et d'un quotidien régional, *La Dépêche du Midi*.

> recherches ; et inversement il n'est pas rare qu'une base de données soit adoptée pour telle ou telle recherche. (Marchello-Nizia 1999, cité par Prévost 2005 : 155)

En fin de compte, ce qui importe dans le choix et/ou la constitution du corpus est sa dépendance à l'égard d'une application ou d'une gamme d'applications car un corpus est consulté pour une recherche déterminée :

> Un corpus est adéquat ou non à une tâche en fonction de laquelle on peut déterminer les critères de sa représentativité ou de son homogénéité. (Rastier 2005 : 32)

5.2 Linguistique de corpus comme méthode

5.2.1 Principes d'exhaustivité et de quantification

Au-delà d'une linguistique qui se fonde **sur** le corpus, une approche **de** corpus suppose une approche résolument empirique, dans la laquelle l'observation du corpus constitue le point de départ de l'étude. M.-P. Jacques (2005 : 25) isole quatre spécificités qui me semblent bien résumer les points forts de telles approches :

– mettre en lumière des fonctionnements linguistiques qui échappent à l'intuition ;

– corriger les intuitions sur le fonctionnement de la langue ;

– avoir des indications en terme de fréquence ;

– rendre compte de la variation.

Ces spécificités sous-tendent aussi une question méthodologique importante qui est celui de l'exhaustivité de la démarche. Dans une approche du corpus, nous ne sommes plus dans le domaine du possible mais dans le domaine de ce qui est effectivement réalisé. Le linguiste a pour mission de rendre compte de toutes les occurrences du phénomène linguistique étudié :

> *Data are used exhaustively: there is no prior selection of data which we are meant to be accounting for and data which we have decided to ignore as irrelevant to our theory.* (Leech 1992 : 113)

Ce n'est pas une question de « butinage » à la recherche d'exemples appropriés comme dans le cas d'une approche fondée exclusivement sur des exemples attestés. Dans une véritable approche de corpus le linguiste est obligé de tout traiter. De cette manière il est censé éviter certains écueils des analyses trop subjectives et biaisées.

Toutefois, il est difficile d'éviter toute implication de l'analyste. Dans la mesure où le choix ou la constitution du corpus à traiter relève d'un choix, il existera toujours à mon avis un élément de subjectivité. La sélection des données peut être délibérément orientée en fonction d'un objectif particulier. Le choix du phénomène à analyser, quant à lui, peut également demeurer la projection d'une construction théorique ou des partis pris théoriques préexistants (Péry-Woodley 2005). C'est pour cette raison que certains linguistes du courant *corpus linguistics* proposent une démarche beaucoup plus radicale. G. Williams (2005), par exemple reprend à E.Tognini-Bonelli (2001) la distinction entre les études *corpus-based* et les études *corpus-driven.*

> Deux approches fondamentales se côtoient, l'une déductive, *corpus-based* en anglais, qui utilise le corpus pour confirmer ou infirmer une hypothèse, et une linguistique de corpus inductive, *corpus-driven*, qui cherche à explorer les données sans *a priori*. (2005 :13)

Autrement dit, tandis que les études *corpus-based* se fondent sur le corpus comme support, les études *corpus-driven* considèrent les corpus comme apport. Dans cette dernière approche le linguiste est simplement considéré comme un explorateur. C'est le corpus qui le dirige.

Tout en étant conscient des dangers d'une trop forte implication de l'analyste dans son étude, de mon point de vue le corpus lui-même ne peut pas être considéré comme un objet totalement autonome. Il est choisi par l'analyste et est interprété par l'analyste. Il serait plus apte de considérer le corpus comme « une médiation consciente entre le chercheur et le fait linguistique », ou un lieu de confrontation entre la théorie et l'empirie (Mellet 2002 : 9).

Un autre point méthodologique important qui sous-tend la grande majorité des approches de corpus est l'accent mis sur les données quantitatives, particulièrement en ce qui concerne la fréquence d'usage des phénomènes observés. Si l'introspection permet parfois au locuteur d'évaluer la fréquence d'un phénomène linguistique qui lui est familier, ces impressions restent approximatives et peuvent même souvent être erronées. Le recours au corpus et *a fortiori* à un corpus numérisé permet au linguiste de quantifier ses résultats de façon scientifique. Cet apport est d'autant plus appréciable dans le cas d'un

texte spécialisé (texte juridique, médical etc.) pour lequel le linguiste ne peut forcément faire appel à son intuition.

Toutefois, si la numérisation des corpus et les possibilités offertes par les différents outils d'analyse disponibles rendent techniquement faciles décomptes et calculs, l'interprétation et la reproductabilté de ces résultats restent beaucoup plus complexes. Comment savoir si les usages mis en évidence sont vraiment représentatifs ou simplement le fruit du hasard ? L'insistance sur la quantification a, en particulier, des implications importantes pour le volume de données considérées : la validation statistique sur des faibles volumes de données étant particulièrement difficile (Péry-Woodley 2005). Dans mes propres recherches sur la langue scientifique, j'ai essayé surtout d'aborder ce problème en amont. En ciblant de façon très précise l'échantillon de textes étudiés, au sein d'une approche de genre (cf. chapitre 6 *infra*), j'ai pu m'assurer de la cohérence de l'ensemble et ainsi de la représentativité des phénomènes dégagés au sein des confins imposés. Il serait en effet dangereux de mélanger les genres et les finalités des discours, au simple motif que le thème général est le même : textes « scientifiques », texte « techniques » etc. En revanche, un corpus très ciblé et homogène permet une étude en profondeur et une multiplication des angles d'approche.

5.2.2 Analyse textuelle et linguistique de corpus

La contribution de la linguistique du corpus est aujourd'hui omniprésente dans tous les travaux linguistiques s'appuyant sur des données textuelles. Dans mes propres travaux, l'apport des concordanciers[5] (notamment ceux de *Tact* et de *WordSmith*) ont souvent été d'une aide précieuse, pour l'élaboration de listes contextuelles et/ou des tris dans les stades initiaux d'analyse. Grace à de tels outils, les gains pour le chercheur, en temps et en fiabilité, sont appréciables. Dans une étude portant par exemple sur les usages des pronoms personnels dans les communications de colloque de différentes disciplines scientifiques

[5] Cependant, l'emploi de tels concordanciers simples n'est pas adapté pour l'analyse des corpus enrichis (balisés en SGML ou XML) où il faut un véritable outil de navigation (cf. Williams et Burnard 2005).

[THOM 05a], les listes générées par ces logiciels ont permis l'établissement d'une base de départ fiable pour l'analyse des différentes valeurs à attribuer à ces pronoms. Toutefois, malgré cette utilité évidente, les outils et méthodes de la linguistique de corpus ne font pas toujours bon ménage avec les objectifs de la linguistique textuelle.

Tout d'abord, les concordanciers sont habituellement peu adaptés à la visualisation du corpus élargi. En effet, les simples lignes de texte dégagées ne suffisent pas la plupart du temps à permettre une analyse textuelle approfondie. Dans le cas de mes analyses sur les pronoms personnels [THOM 05a], il était souvent nécessaire de remonter aux textes sources et d'examiner de grands espaces de texte afin de pouvoir distinguer entre l'utilisation d'un *we* collectif (*nous* les chercheurs responsables pour ces recherches en question) et d'un *we* inclusif *(nous* les chercheurs responsables pour ces recherches en question et *vous* les lecteurs avertis). Les informations fournies par le concordancier ne permettaient pas non plus d'examiner la distribution des pronoms au fil du texte. Afin de savoir si l'usage de tel ou tel pronom était plus répandu dans les parties introductives de l'article ou dans la présentation des résultats, une nouvelle analyse manuelle s'imposait de nouveau.

Dans une analyse sur les liens entre la linguistique textuelle et la linguistique de corpus, Péry-Woodley (2000 : 148) identifie trois grandes difficultés soulevées par l'étude en corpus de l'unité « texte », à savoir le principe de l'exhaustivité, la constitution de corpus bilingues et le respect du *texte* en tant qu'unité dans la constitution du corpus. Théoriquement, une approche de corpus ne met pas en cause la distinction entre acceptabilité et non-acceptabilité. Mais doit-on forcément, par exemple, inclure les occurrences manifestement malformées en toute circonstance ? La comparabilité de textes produits dans des cultures linguistiques différentes, dans la perspective d'une étude contrastive, pose également un grand problème. Le recueil de textes bilingues (les corpus parallèles de traduction) est souvent peu fiable à cause du risque de calque de la langue cible sur la langue source. Toutefois, la constitution de corpus

vraiment « comparables », où les deux séries de textes sont strictement comparables sur tous les plans est également difficile à garantir.

Dans mes propres travaux je me suis souvent confrontée à ce problème dans la constitution de corpus de textes scientifiques « comparables ». Comment trouver, par exemple, l'équivalent en français d'une revue médicale spécialisée tel que *The Lancet* ? Comme les chercheurs dans ce domaine sont encouragés (ou presque contraints) à publier en anglais, les revues médicales en français sont devenues rares. Il est très difficile de trouver deux revues qui correspondent parfaitement sur tout un ensemble de paramètres : discipline scientifique, degré de diffusion, niveau de spécialisation [THOM 07].

Toutefois, la plus grosse difficulté pour l'étude en corpus de l'unité *texte* concerne à mon avis la question du respect de l'unité texte en lui-même. Beaucoup de corpus, et surtout beaucoup de grands corpus, ne sont pas constitués de textes entiers mais d'extraits ou d'échantillons. La notion de *« pieces of language »* présente, par exemple, dans la définition du corpus par Sinclair (citée en 5.1.1) est problématique car l'unité textuelle en tant que telle n'est nullement pas prise en compte :

> *A collection of pieces of language… to be used as a sample of language.* (Sinclair 1996)

Ce point de vue va à l'encontre de ceux principaux tenants de la linguistique textuelle développés dans les chapitres précédents (cf. 1.3 *supra*). Poussée à l'extrême, la définition de Sinclair suppose la vision d'un texte comme étant un amas de données langagières dans lequel on peut se permettre de prélever seulement ce qui nous intéresse. Tout ce qui concerne la cohérence et l'organisation textuelle n'est tout simplement pas accessible. Recueillir *du* texte au lieu de recueillir *des* textes interdit à l'analyste toute possibilité de reconstituer le discours d'origine car le discours en lui même n'est pas un observable, ce que l'on observe c'est le texte. C'est à travers la matérialité du texte en tant que langage – du matériel sonore ou visuel – que l'on reconstitue le discours.

La citation suivante de F. Rastier appuie succinctement mes sentiments à ce sujet :

> Le texte est pour une linguistique évoluée l'unité minimale, et le corpus l'ensemble dans lequel cette unité prend son sens. (2005 : 31)

D'autre part et en corollaire, il est essentiel dans ma conception d'un corpus d'avoir accès non seulement aux textes intégraux, mais aussi de pouvoir situer très précisément le contexte des produits textuels sous étude. En effet, la description et l'exploitation de traces à la surface textuelle ne prennent sens et n'ont de pertinence que dans la mesure où ces marques sont rapportées au discours par lequel elles sont produites. Dans un corpus d'échantillons, il manque non seulement des informations sur le contexte linguistique élargi, mais aussi des informations sur la situation extralinguistique à l'origine de l'interaction. Or il est essentiel pour moi de pouvoir caractériser de façon précise la réalité extralinguistique du corpus avant de commencer toute analyse linguistique. Le potentiel d'un corpus et la « réussite » du travail de linguiste résident pour une grande part dans la qualité du travail de choix et la constitution du corpus accompli en amont. Comme le souligne Péry-Woodley :

> La validité des conclusions qu'on peut tirer des observations fournies par l'analyse d'un corpus dépend de ce qu'on sait de l'objet étudié, et de la possibilité de le situer par rapport aux autres objets possibles, en fonction de critères externes à l'analyse. (1995 : 218)

5.3 Corpus et langue spécialisée[6]

Depuis mes premières recherches effectuées sur les travaux de mes étudiants ingénieurs en télécommunication [THOM 94], jusqu'à certaines analyses récentes portant sur des corpus d'articles de recherche en médecine [THOM 07 ; 08a et 08b], j'ai toujours privilégié le recours aux données réelles. Toutefois, dans le cadre de mes analyses sur les textes spécialisés, ce recours au corpus revêt une importance toute particulière.

[6] Tout en étant dévouée à des fins spécifiques, la langue spécialisée fait partie de la langue générale. Pour cette raison l'appellation de *langue spécialisée* (Lerat 1995 : 20), de préférence à celle de langue de spécialité, me semble plus apte à saisir cette idée d'un continuum et non d'une rupture avec la langue générale.

Le corpus est en effet la pierre angulaire du travail de recherche des linguistes de la langue spécialisée. Tout comme pour le linguiste travaillant sur les langues disparues, les corpus sont souvent le seul moyen d'accéder à l'objet d'étude. Ce sont les textes produits par une communauté de spécialistes qui permettent d'avoir accès aux connaissances et d'observer les formes d'expression en fonction des objectifs de communication. Autrement dit, pour le linguiste en langue spécialisée, son champ de recherche se confond avec son champ d'exploration. A la différence du linguiste de « langue générale » qui peut à la rigueur se contenter de sa propre intuition linguistique pour décider des critères d'acceptabilité, le linguiste travaillant sur les textes spécialisés n'est pas forcément à même de porter des jugements d'acceptabilité. Certaines contraintes syntaxiques et pragmatiques peuvent rester opaques à celui qui ne fait pas partie de la communauté de discours en question. L'examen des régularités rencontrées au sein du corpus est alors l'un des seuls moyens pour accéder à la « grammaire » sous-jacente.

Mais ceci ne veut pas dire que l'intuition ne joue aucun rôle. Fournir des interprétations les plus justes possibles ne peut se faire sans une très bonne connaissance des conditions d'énonciation. Il est important de compléter ce que révèlent les corpus par une maîtrise de leurs caractéristiques externes et de leurs conditions de production. Le linguiste doit en effet connaître son « terrain ». Citant à ce propos A. Condamines :

> Il y a un temps pour comprendre le terrain, ses besoins, ses traditions, ses attentes et puis, une fois intégrée la situation, il y a un temps pour interpréter le corpus constitué, pour en repérer les régularités « immanentes » puis à nouveau un temps d'évaluation des résultats. (2003 : 34)

Dans le cas des analyses sur des corpus très spécialisés, la question peut se poser du degré réel de compréhension que le linguiste doit avoir du terrain sur lequel il travaille. En tant que linguistes avons-nous la compétence discursive nécessaire pour analyser n'importe quel texte spécialisé (Dudley Evans 1993 ; Petit 1994) ? Il s'agit d'une question souvent débattue parmi les anglicistes du

GERAS[7]. Pour ma part, tout en admettant le fait que je ne pourrai pas normalement évaluer la pertinence cognitive du contenu d'un corpus d'articles de recherche portant, par exemple, sur l'électromagnétisme en physique, j'estime qu'il est essentiel en revanche de me familiariser avec les épistémologies de la discipline concernée et avec les conventions et traditions de ses acteurs. Le linguiste doit comprendre au mieux la situation dans laquelle il intervient : le contexte précis de production et de réception (culturel et environnemental) ; les rapports entre les acteurs (rapports d'autorité, communication entre pairs, etc.), les finalités du discours. Afin d'acquérir ces connaissances, une collaboration avec les acteurs de la discipline en question est évidemment très bénéfique.

Le recours aux spécialistes du domaine peut s'avérer utile, voire essentiel, à divers moments de l'étude du corpus. Au tout début, concernant le choix même du corpus l'avis des experts peut se révéler crucial. Dans le cadre d'une étude effectuée récemment sur l'emploi des phrases conditionnelles dans le domaine médical [THOM 07], des chercheurs en médecine étaient consultés afin de sélectionner les revues médicales les plus représentatives de la discipline médicale concernée. L'avis des spécialistes francophones du domaine était particulièrement important pour la constitution d'un corpus bilingue anglais/français d'articles médicaux portant sur l'oncologie. Afin de trouver des revues, portant non seulement sur les mêmes problématiques, mais également d'une notoriété similaire dans les deux langues, l'opinion informée des praticiens m'était indispensable.

La valeur d'une étude en langue spécialisée est directement liée à la construction du corpus. La cohérence dans la démarche de sélection et la recherche rigoureuse d'une adéquation entre l'objectif visé par le linguiste et les textes choisis en tant que matière première sont des préalables nécessaires à la réussite du projet de recherche. L'un des reproches formulés le plus souvent à l'égard des premières études sur le « discours scientifique » concernait

[7] GERAS : Groupe d'étude et de recherche en anglais de spécialité. Il s'agit d'une société savante regroupant les anglicistes travaillent sur les langues de spécialités en France.

l'hétérogénéité des corpus employés. Il est facile de comprendre, par exemple, qu'un article de vulgarisation de *Time Magazine* sur la volcanologie et un article de recherche primaire d'une revue spécialisé en médecine nucléaire, tout en traitant de la « science », sont très loin l'un de l'autre. Il est en effet difficile à parvenir à des résultats cohérents si le corpus étudié n'offre pas une véritable spécificité. Dans la constitution d'un corpus spécialisé, les principaux écueils à éviter sont selon Martin : le syncrétisme de domaines hétéroclites, la confusion des finalités discursives, les mélanges de types de spécification et les disparités liées aux différents niveaux de spécificité (Martin 1997 : 78-79). Dans mes propres recherches, j'ai essayé de constituer des corpus les plus homogènes possibles en respectant au maximum les recommandations de cohérence interne, énumérés ci-dessus.

Un autre critère qu'il est important de prendre en considération concerne l'identité culturelle des auteurs des textes spécialisés. Si, par exemple, l'objectif est de constituer un corpus homogène d'articles scientifiques en anglais, il est aussi impératif de vérifier l'identité linguistique des auteurs. Comme chacun le sait, l'anglais tient lieu de langue commune aux scientifiques du monde entier. En effet, pour diffuser ses travaux et se faire connaître de ses pairs, le chercheur doit souvent recourir à l'anglais pour la publication et la présentation de ses travaux. Toutefois, en ce qui concerne la communication scientifique, les implications linguistiques de ce passage d'une langue nationale à une langue commune ont été peu étudiées. Autrement dit, l'impact de la langue maternelle du chercheur sur la formulation et l'organisation de ces idées en anglais a été peu évalué [THOM 05e]. Peut-être en raison de la conception empiriste de la science, léguée par l'histoire, comme « une vérité universelle », les éventuelles variations linguistiques et culturelles associées à l'emploi de l'anglais scientifique n'ont pas vraiment attiré l'attention. On parle toujours de l'anglais scientifique au singulier (et non pas des langues anglaises scientifiques au pluriel) comme s'il existait une version standardisée à la disposition de tout chercheur [THOM 05a].

Or l'impact de la langue et de la culture maternelle peut être considérable. J'ai consacré plusieurs travaux à ce sujet et j'y reviendrai dans le chapitre 8. Les résultats de ces travaux confirment l'importance du critère de la culture d'origine dans la configuration du corpus. Si l'on veut créer des corpus vraiment homogènes d'articles en anglais il est impératif de ne retenir que des articles provenant des auteurs anglophones travaillant dans des laboratoires anglophones, où au contraire de créer des sous-corpus en fonction de l'origine géopraphique et linguistique des auteurs : articles rédigés par des auteurs d'origine francophones, articles rédigés par des auteurs d'origine russe et ainsi de suite [THOM 05a].

La constitution d'un corpus en langue spécialisée peut être effectuée essentiellement selon deux méthodes ; l'une qu'on peut appeler verticale consiste à sélectionner à l'intérieur d'un domaine, par exemple la médecine, une discipline médicale particulière, par exemple l'oncologie, la pharmacologie, la chirurgie dentaire. On peut ensuite comparer les différents usages de la langue spécialisée au sein de cette spécialité médicale : l'article de recherche primaire (communication entre pairs), l'éditorial d'une revue spécialisé, l'article de vulgarisation (communication spécialiste /non-spécialiste). L'autre méthode, que l'on peut qualifier d'horizontale, consiste à comparer deux ou plusieurs exemples du même genre de texte, par exemple l'article de recherche, au sein de disciplines scientifiques différentes et/ou de cultures linguistiques différentes.

Dans mes propres recherches sur les textes spécialisés, j'ai employé ces deux méthodes. Dans une analyse sur l'emploi des phrases conditionnelles en anglais, c'est la méthode verticale qui a été adoptée. A partir d'un corpus discipline médicale spécifique (l'oncologie), une comparaison était effectuée sur l'emploi de la phrase conditionnelle dans trois genres bien distincts : les articles primaire de recherche médicale ; les éditoriaux de revues spécialisées et les communications (orales) de colloques [THOM 08a]. Pour d'autres études c'était la méthode horizontale qui était privilégiée. Dans un certain nombre de travaux entrepris avec E. Rowley-Jolivet nous avons cherché à déterminer les

caractéristiques d'un genre particulier, la communication de colloque, dans trois disciplines différentes : en physique, en géologie et en médecine [THOM 04d ; THOM 05b]. Dans une autre étude cette méthode était employée pour caractériser la communication de colloque à travers deux cultures linguistiques différentes. Nous avons comparé les communications de colloque faites par les chercheurs anglophones avec celles des chercheurs francophones, contraints à présenter leurs travaux en anglais [THOM 05d].

Dans tous ces travaux, la recherche et la constitution d'un corpus d'étude homogène ont joué un rôle capital. Afin de pouvoir examiner les spécificités et les invariants d'un discours scientifique donné, il est essentiel de pouvoir cerner très précisément les caractéristiques internes et externes de ce corpus. Les descriptions proposées par le linguiste peuvent ensuite être liées à ces conditions contextuelles précises. Dans cette approche, le corpus reste ainsi beaucoup plus qu'une simple source de données. Toutefois, il reste le problème de la fiabilité de l'interprétation donnée aux faits linguistiques relevés. Comment s'assurer de la « *reproductibilité* » des résultats proposés ? Autrement dit, est-ce qu'un autre analyste arriverait aux mêmes interprétations ? C'est ici que la notion de genre textuel (cf. chapitre 6) me paraît particulièrement pertinente. La notion de genre nous permet de fixer un cadre d'interprétation et de mieux évaluer ces affinités entre les caractéristiques extralinguistiques et les fonctionnements langagiers.

Chapitre 6 – Expliquer les variations : l'approche de genre

Résumé

Ce chapitre est consacré à la notion de genre – une notion qui se situe au cœur de mes travaux d'analyses textuelles. Pour essayer de comprendre ce que peut être l'interprétation d'un texte, il est nécessaire de se donner des cadres fixes d'interprétation. La notion de genre permet à la fois de prendre en compte la réalité des variations langagières et de repérer les régularités de fonctionnement.

Dans la première section du chapitre (6.1), « cohérence et généricité », je défends l'hypothèse selon laquelle les effets de généricité et les effets de textualité sont indissociables. Autrement dit, la capacité d'un récepteur à reconstituer la cohérence d'un texte reste fortement influencée par sa capacité à identifier et à régler ses attentes sur un modèle d'interprétation générique.

Cependant, les difficultés à dresser une typologie de genres sont nombreuses et les classifications sont abordées à partir de perspectives fort différentes. Dans la section 6.2, je présente différentes démarches classificatrices, en précisant les choix adoptés dans mes propres travaux.

Dans la troisième partie du chapitre (6.3), je focalise tout particulièrement sur le rôle de genre dans l'analyse de textes spécialisés. Je présente l'approche de J.M. Swales (1981, 1990) pour définir les genres « académiques » de la recherche universitaire – une approche adoptée dans mes propres analyses sur les présentations scientifiques orales et sur les articles de recherche scientifique.

Chapitre 6. Expliquer les variations : l'approche de genre

6.1 Cohérence et généricité

Genres et textes sont intimement liés. Si nous acceptons qu'un jugement sur la cohérence d'un texte est lié au caractère pertinent et acceptable du texte ([THOM 00] et chapitre 1 *supra*), celui-ci passe aussi par un jugement sur l'adéquation du texte par rapport à certaines attentes génériques. Les effets de textualité me semblent en effet difficilement dissociables des effets de « généricité » (cf. Adam et Heidmann 2004).

M.A.K. Halliday et R. Hasan (1976), tout comme plus récemment L. Lundquist (1999), se sont efforcés de démontrer que les locuteurs d'une langue sont aptes à distinguer un texte d'un non-texte, avec le même degré d'efficacité qu'ils sont capables de distinguer une phrase grammaticale d'une non-phrase[1]. Dans une expérimentation avec ses étudiants, Lundquist démontre que les sujets savent intuitivement reconnaître si une séquence donnée de phrases constitue ou non un texte potentiel. Face à une séquence de phrases volontairement désordonnée, 90% des étudiants ont qualifié la séquence de *non-texte*. La notion de la cohérence textuelle existerait bel et bien dans l'esprit des locuteurs (cf. chapitre 1 *supra* et Charolles 1988). Les procédures mises en œuvre par les étudiants pour « retrouver » la cohérence sont selon Lundquist de deux ordres : le repérage de marques linguistiques et le calcul d'inférences. Or cette capacité de repérer des marques explicites et de traiter ces marques et inférences (et ainsi de distinguer un texte d'un non-texte) est aussi étroitement liée, à mon avis, aux capacités des locuteurs à identifier des genres textuels.

.

Sur le plan de l'organisation des contenus, un texte apparaîtra d'autant plus cohérent qu'il est conforme aux lois du genre. Je le formulais comme suit en 2000 :

> À chaque genre sont associées des stratégies interprétatives spécifiques qui se manifestent par des instructions intrinsèques. Notre interprétation de la cohérence

[1] En linguistique de phrase, il n'existe pas non plus de consensus sur ce qui constitue une phrase acceptable. Les nombreux débats autour de l'acceptabilité sémantique de l'exemple célèbre de N. Chomsky : *« Colorless green ideas sleep furiously »,* en témoignent.

> passera par la facilité que nous avons à suivre ces instructions et à arriver ainsi à une interprétation de la cohérence textuelle au sein du genre en question. [THOM 00 : 37]

Le genre peut affecter potentiellement tous les niveaux de l'organisation textuelle – tant structurels ou formels, que cohésifs ou stylistiques. Une lettre de candidature qui ne ressemble pas physiquement à une lettre de candidature (par sa présentation et sa mise en page) sera difficilement cohérente pour la plupart des lecteurs. Toutefois, ces instructions ne sont pas forcément « connues » par les interlocuteurs et peuvent être plus ou moins conscientes selon le cas. Comme l'explique A. Condamines : « On a conscience de parler dans une langue ; beaucoup moins de s'inscrire dans un genre donné. » (2005 : 25)

L'impact du genre existe en amont même du décodage linguistique. Avant de lire la première page d'un roman ou le titre d'un article scientifique, le lecteur aura certaines attentes vis-à-vis du texte qu'il a sous les yeux. L'adéquation du texte à la situation de communication passe par le respect du genre type de texte attendu. Selon la formule de G. Vigner :

> Reconnaître un genre, c'est pouvoir régler sa lecture sur un système d'attente, l'inscrire dans une trajectoire prévisible. (1979 : 64)

Un texte qui correspond aux attentes du lecteur aura plus de chances d'être considéré comme « réussi » ou « cohérent » qu'un texte où certaines conventions génériques relatives au genre en question sont absentes ou mal signalées. Si le texte ne correspond pas aux attentes du récepteur, ce dernier risque d'être initialement déstabilisé.

Suivant les situations de communication, il est possible de procéder à quelques généralisations sur les types d'attente qu'un récepteur risque d'avoir. Le recours à la notion de genre est ainsi d'une grande utilité pour le linguiste, en tant qu'analyste. S'il est communément admis que la prise en compte de données contextuelles est d'une importance cruciale dans toute analyse textuelle (cf. chapitre 1 *supra*), la nécessité de prendre en compte ces aspects pose un certain nombre de problèmes pour l'analyse. Le linguiste lui-même est

rarement le destinataire du texte qu'il analyse. C. Kerbrat-Orecchioni compare même la position du linguiste à celle d un « intrus » :

> Dans le cas des discours naturels : l'analyste est un récepteur non destinataire, c'est-à-dire qu'il traite un objet qui ne lui est en rien destiné (sauf s'il a lui-même participé à l'interaction qu'il étudie, ce qui pose d'autres problèmes) ; s'immisçant en intrus dans l'échange communicatif, il est généralement incapable de reconstituer la totalité des informations contextuelles pertinentes. (Kerbrat-Orecchioni, 1996 : 47-48).

Pour essayer de comprendre ce que peut être l'interprétation d'un texte, il est nécessaire dans la mesure du possible de restituer le texte dans sa situation d'origine. Une connaissance préalable sur le fonctionnement du genre textuel en question permet à l'analyste d'approcher de plus près cette situation. Au fond, on peut simplement considérer le genre comme ce qui tient lieu de contexte.

Toutefois, si la notion de genre est intuitivement très attractive, sa formalisation crée beaucoup de difficultés. Il est impossible de prévoir toutes les situations de communication : la diversité des genres est infinie. Les genres sont également mouvants et s'entrecroisent. Une délimitation précise reste ainsi délicate, voire impossible. Il peut exister aussi des décalages importants entre les régularités linguistiques réelles et des bases sociales intuitives. Les risques de généralisations abusives sont importants. Comme nous l'avons déjà vu dans le chapitre précédent, c'est aussi à cause de ces types de problèmes qu'il est difficile de garantir la représentativité d'un corpus.

Face à ces difficultés, beaucoup de linguistes expriment une méfiance vis-à-vis de la notion de genre ou d'un quelconque classement des textes. J. Molino résume cette méfiance dans les termes suivants :

> Les classifications maniables [...] ne fournissent qu'un cadre vague sans garantie d'homogénéité ni de régularité tandis que les classifications qui visent à être hétérogènes, rigoureuses, monotypiques et exhaustives sont contraintes de se perdre dans une ramification sans limites qui les rend rapidement inutilisables. [...] Il ne saurait exister de théorie générale du discours ou du texte. (Molino 1990, cité par J.-M. Adam, 1997 : 5)

Toutefois, malgré ces problèmes indéniables, j'estime qu'il existe une voie médiane. Si l'établissement de typologies définitives relève d'une gageure, la

notion de genre permet tout de même de fixer un cadre d'analyse. On ne peut pas étudier chaque texte comme s'il s'agissait d'une production langagière totalement nouvelle et originale. Une approche de genre permet d'avoir des « prises » sur la variation linguistique. On peut certes ne pas tout maîtriser, mais la notion de genre permet de mieux aborder les variétés des situations discursives et les textes.

Je propose dans la section suivante de résumer brièvement certaines des démarches classificatrices les plus connues avant de préciser l'approche privilégiée dans mes propres travaux.

6.2 Démarches classificatrices

La classification des textes peut être abordée à partir de perspectives très différentes. Certaines typologies se concentrent sur les intentions fonctionnelles et d'autres sur les aspects situationnels ou pragmatiques. D'autres encore se fondent exclusivement sur des catégories linguistiques formelles. Dans la tradition littéraire, la notion de genre a été longtemps employée dans une perspective à la fois descriptive et prescriptive. Pour être reconnu comme un bon écrivain il fallait suivre certaines règles de bonne rédaction (Branca-Rosoff 1999), pour ce qui concerne la composition de son œuvre (forme, organisation et contenu). Ce côté très prescriptif a depuis été remis en question. Mais tout en rejetant la classification rigide, le besoin de genres demeure. Afin de pouvoir « désobéir » à un genre (ou de rendre compte d'éventuelles transgressions opérées par l'auteur) il faut convenir de la préexistence d'un genre. Le genre jouerait en quelque sorte un rôle d'interface entre auteur et lecteur, et ceci quelle que soit la nature du texte (Todorov 1978).

La typologie la plus répandue en linguistique (et dans l'enseignement) est sans doute toujours celle issue de la tradition rhétorique. Elle distingue essentiellement quatre grands types d'activité textuelle fondés sur les usages que l'on peut faire du langage : le texte narratif, le texte descriptif, le texte expositif et le texte argumentatif (de Beaugrande et Dressler 1981, Maingueneau 1976, Ruck 1980, Longacre 1976). Toutefois, ces typologies, tout

comme les classifications fondées sur la fonction de communication elle-même (Jakobson 1963), sont difficiles à appliquer telles quelles. Beaucoup de textes sembleraient résulter d'un certain mélange de ces fonctions. Par exemple, dans un récit, généralement dominé par des séquences narratives, il existe souvent des séquences descriptives ou conversationnelles. De même, certains articles scientifiques sont loin d'être seulement « expositifs » ou « informatifs », dès lors qu'ils prennent en compte la présence d'un destinataire – soit pour le convaincre dans le cas d'un article scientifique primaire ou d'un article didactique soit pour le distraire dans le cas d'un article de vulgarisation. De telles typologies textuelles « englobantes » font du discours un objet souvent trop abstrait et de ce fait sont difficiles à appliquer.

Afin d'être vraiment « opérationnelles », les classifications devraient surtout à mon avis prendre en considération une variété de paramètres situationnels. D. Biber (1988 ; 1993) a distingué huit composants (des composants eux-mêmes divisés en sous-composants) susceptibles de jouer un rôle dans la variation langagière des textes :

> *(I) Participant roles and characteristics (individual and group characteristics);*
>
> *(2) Relations amongst the participants (Social roles, degree of shared knowledge etc);*
>
> *(3) Setting;*
>
> *(4) Topic;*
>
> *(5) Purpose;*
>
> *(6) Social Evaluation (of event and content);*
>
> *(7)Relations of participants to the text;*
>
> *(8) Channel.*
>
> (Version abrégée de Biber 1988 : 30-31).

Le poids et l'utilité de tels paramètres peuvent aussi varier selon la situation dans laquelle s'inscrivent les textes analysés. Dans le cas de différents « textes » scientifiques (tels les articles de recherche scientifiques, les articles de vulgarisation, ou les présentations de colloque), les paramètres les plus pertinents à mes yeux sont la finalité du texte, le lectorat ciblé et son mode de transmission (cf. *infra*).

Il est certain que toute démarche classificatrice comporte un risque de généralisation. Il peut être tentant de faire des corrélations un peu « naïves » entre certains faits de langue et l'appartenance à un genre de texte donné. Ce risque est peut-être moins important dans le cas de textes issus du discours scientifique où nous trouvons de toute façon un certain nombre de « rituels », mais ce danger subsiste. Il existe essentiellement deux moyens de définir les critères typologiques, soit déductivement en instituant des classes *a priori* que l'analyse chercherait à valider, soit inductivement par une approche empirique[2] comparative qui permettrait de justifier *a posteriori* certaines classifications. Il y a un piège tautologique qui consiste à confirmer ou prouver l'existence d'un tel type de texte à partir de la simple observation d'une certaine caractéristique linguistique dans ce texte, ou bien au contraire à partir de l'observation de certaines caractéristiques dans le discours pour supposer l'existence d'un certain type textuel. Il est pour cette raison important de souligner ces points de départ fort différents afin d'éviter des corrélations abusives entre formes et fonctions et une certaine circularité.

6.2.1 Une notion biface

Le genre est généralement perçu comme une entité à deux faces, avec une face interne (les fonctionnements linguistiques) et une face externe (les pratiques sociales). Or comme, par exemple, Biber (1988) et P. Achard (1995) l'ont souligné, il est essentiel afin d'éviter le danger de circularité évoqué ci-dessus de distinguer clairement entre ces deux aspects. Pour les classifications à partir de critères externes *« relating to 'speaker/author purpose »*, Biber propose d'employer le terme de *genre* (1988 : 68). En revanche, pour les regroupements de textes s'appuyant sur les critères linguistiques formels, il parle de *« text type »* (*Ibid* : 68). Achard pour sa part désigne comme « registres » les rôles langagiers attendus dans les différentes situations sociales et comme « genre » les fonctionnements linguistiques devant y correspondre (Branca-Rosoff 1999 : 19). Cette opposition entre registre et

[2] Un exemple d'approche empirique serait l'analyse informatisée de Biber (1988) pratiquée sur un corpus de textes de sources diverses, à partir duquel Biber identifie soixante-sept traits linguistiques permettant de classifier les textes.

genre est également courante dans la majorité des courants systémiques du genre, où nous trouvons, par exemple chez Martin la notion de genre définie comme *« a staged, goal-oriented social process realised through register »* (Martin 1992 : 505).

Une fois établie cette distinction entre la face interne et externe de la notion de genre, c'est précisément l'articulation de ces deux faces qui me parait particulièrement intéressante. Si la plupart de mes travaux prennent appui sur l'étude de régularités linguistiques visibles à la surface textuelle (l'analyse des structures syntaxiques et de différents marqueurs), ceci n'est pas fait dans un but strictement descriptif. Je cherche surtout à évaluer l'importance du choix précis des marques relevées en fonction d'un contexte social, à articuler ainsi des formes linguistiques et des fonctionnements sociaux. Dans mes travaux sur l'emploi des phrases conditionnelles[3] en anglais et en français, par exemple [THOM 07 ; THOM 08a], j'ai analysé l'emploi des phrases conditionnelles en relation avec les conditions extralinguistiques qui ont présidé à leur mise en texte. Autrement dit, j'ai essayé d'associer une approche résolument ascendante avec une réflexion sur les conditions de production des textes analysés. Cette analyse m'a permis de constater qu'une modification de la finalité du discours – une modification entre par exemple la finalité d'un article de recherche et celle d'un éditorial spécialisé – entraîne des répercussions importantes sur la forme et fonctions des phrases conditionnelles employées.

6.2.2 Distinction type /genre

Jusqu'ici je n'ai pas distingué le terme de *genre* de celui de *type* de texte ou de discours. Toutefois, comme beaucoup de linguistes (Adam 1987 ; Rastier 1989 ; Maingueneau 1991, 1996 ; Biber 1988) je pense qu'une distinction s'impose. Selon D. Maingueneau, par exemple, le *type* de discours se réfère à une catégorisation à la fois large et instable qui « définit pour les époques et une société données un certain nombre de secteurs de l'activité discursive » (1991 : 213). En revanche le terme *genre* réfère à une catégorie plus restreinte

[3] Cf. chapitre 9 *infra*.

qui spécifie « ces types de discours ». Il s'agit des discours « rapportables à des lieux d'énonciation déterminées » (*ibid.*). On parlerait ainsi d'un éditorial et d'un article scientifique primaire respectivement comme des genres du discours journalistique et du discours scientifique. Tout en étant globalement d'accord avec ces précisons, pour moi le point essentiel concerne la question de finalité ou de but rhétorique. Les genres se distinguent des *types* de discours par l'existence d'un but rhétorique identifié.

De la même façon, il me semble essentiel aussi de distinguer clairement entre type textuel et genre textuel. Les genres de texte sont surtout superordonnés par rapport aux types de texte. Un genre textuel peut relever de différents types de texte. Un article de recherche peut contenir, par exemple, à la fois des séquences de types narratifs et des séquences de type descriptifs (cf. *supra*).

De telles distinctions et définitions nous éloignent des définitions traditionnelles du *genre*, telles qu'elles sont employées dans les études littéraires, où un genre littéraire particulier (telle une comédie ou une tragédie) se définissait en fonction de sa forme et de son contenu. Les genres linguistiques modernes, si nous suivons la définition de F. Rastier, sont essentiellement des codifications sociales et culturelles :

> Un genre est un programme de prescriptions positives ou négatives et de licences qui règlent aussi bien la génération d'un texte que son interprétation ; elles ne relèvent pas du système fonctionnel de la langue, mais d'autres normes sociales. (1989 : 37)

Je crois, comme Rastier, que tout échange linguistique est réglé par les pratiques sociales dans lesquelles il prend place. Un locuteur qui participe à plusieurs pratiques sociales sait intuitivement manier plusieurs « genres » et ainsi exprimer leurs spécificités dans le discours. Ce sont les genres qui permettent de concrétiser les liens entre l'abstraction du texte et la généralité des pratiques discursives :

> Un discours s'articule en divers genres, qui correspondent à autant de pratiques sociales différenciées à l'intérieur d'un même champ. Si bien qu'un genre est ce qui rattache un texte à un discours. Une typologie des genres doit tenir compte de l'incidence des pratiques sociales sur les codifications linguistiques. (*Ibid.* :40)

6.3 Genre et discours spécialisés

Nous retrouvons cette même préoccupation concernant l'influence des « pratiques sociales » dans la conception d'un genre chez les analystes travaillant sur les discours professionnels ou spécialisés. Les linguistes à l'origine du courant *English for Specific Purposes* (ESP) tels Swales (1981, 1990), T. Dudley-Evans (1987, 1994, 1995), K. Hyland (1999 ; 2000), ont cherché à souligner les attentes essentiellement culturelles et sociales en rapport avec la reconnaissance d'un genre. Dudley-Evans (1987) définit par exemple le genre comme :

> *A typified society recognised form that is used in typified society circumstances*. (1987 : 1)

D'une façon analogue, Swales souligne l'idée d'une communauté de locuteurs qui partagent et reconnaissent un certain but ou objectif communicationnel :

> *A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style.* (Swales 1990 : 58)

Les études ESP se démarquent des études sociologiques en soulignant l'existence dans ces communautés de discours, des objectifs communicationnels communs, qui dépassent largement les simples besoins de socialisation ou de solidarité que l'on retrouve dans les communautés de parole d'origine sociale. Le fonctionnement d'un genre serait donc déterminé par des considérations fonctionnelles. Tout en tenant compte de l'ancrage disciplinaire ou cognitif d'un genre, ce qui distingue un genre scientifique d'un autre est surtout cette question d'un but ou d'une fin communicationnelle commune.

Si l'étude du discours scientifique[4] se réalise à travers plusieurs genres écrits (articles de recherche, rapports, lettres, comptes rendus) et oraux (communications de colloque, séminaires), c'est l'article de recherche qui a

[4] Dans ce mémoire, le concept du *discours scientifique* est employé comme l'équivalent du terme anglais *academic discourse*. Le terme *scientifique* n'est pas ainsi nécessairement restreint aux « sciences » dans le sens strict de *sciences de la nature*, mais inclut toute branche académique.

reçu le plus d'attention. Les raisons en sont nombreuses. Il s'agit du genre scientifique sans doute le plus accrédité, car il occupe une place centrale dans la diffusion des savoirs au sein de la communauté scientifique. Il constitue aussi un genre relativement « fermé », dans la mesure où il s'adresse à un lectorat restreint et spécialisé, le plus souvent d'autres spécialistes de la même discipline scientifique que celle de l'auteur. Les auteurs sont soumis à des contraintes de forme et de contenu. Dans les revues, les consignes qu'on leur donne sont très détaillées. On doit souvent aussi appliquer une feuille de style particulière. Même si ces contraintes peuvent varier d'une discipline à une autre, les articles de recherche scientifique ont l'intérêt d'être plus régulés et donc plus observables au niveau formel que certains autres genres de discours scientifique[5].

Toutefois, comme beaucoup de travaux récents sur l'écriture scientifique l'ont démontré, ces régularités peuvent cacher un fonctionnement assez complexe (Miller 1984 ; Bazerman 1988 ; Swales 1990). On aurait tort de penser que le style impersonnel souvent caractéristique des articles de recherche équivaut nécessairement à une perspective objective et neutre. L'ouvrage très influent de Swales (1990), par exemple, a mis en évidence le rôle important de la rhétorique dans l'article de recherche, ainsi que le fonctionnement très spécifique des différentes parties de l'article scientifique.

> *The art of the matter, as far as the creations of facts is concerned, lies in deceiving the reader into thinking that there is no rhetoric.* (Swales 1990 : 112)

Si l'objectif de l'article de recherche a toujours été d'informer, ce n'est pas le seul. Les auteurs cherchent aussi à persuader leur lectorat du bien-fondé de l'argument présenté. Le chercheur doit garder un esprit de compétition afin de créer sa niche et de promouvoir ses propres idées. Cet aspect persuasif de l'article de recherche va de pair avec un aspect interactionnel. Il est important pour le chercheur de montrer qu'il fait partie d'une communauté et de situer sa propre recherche par rapport à celle de cette communauté en reconnaissant,

[5] Les contraintes associées aux textes scientifiques oraux sont, par exemple, beaucoup moins rigides (cf. Chapitre 8 *infra*).

par exemple, l'importance des travaux d'autres chercheurs, afin de délimiter la pertinence de ses propres travaux dans ce travail d'ensemble.

La plupart des articles de recherche dans les sciences exactes et médicales[6] suivent un plan d'organisation précise, le modèle IMRED (Introduction, Matériels et Méthodes, Résultats et Discussion). Mais le plan IMRED est considérablement plus qu'une chronologie imposée de rubriques obligatoires. Chaque section de l'article de recherche remplit des fonctions rhétoriques spécifiques, qui correspondent à leur tour à des formulations linguistiques adaptées. On peut ainsi considérer ces rubriques, tout comme l'abstract contenu dans le péritexte de l'article, comme des sous-genres. La section « Introduction » comprendra habituellement la justification de l'étude, aussi bien que la présentation des objectifs de la recherche proposée. La partie « Matériels et Méthodes » fournira une description précise du processus suivi et des protocoles d'expérimentation, afin de permettre la reproduction de cette méthode. Dans la partie « Résultats » on trouvera les détails des principaux résultats, l'interprétation et l'implication de ces résultats étant réservées à la rubrique « Discussion ». Les traces linguistiques des stratégies argumentatives des chercheurs se trouvent surtout dans les parties « Introduction » et « Discussion » de l'article de recherche – phénomène que j'ai eu l'occasion de constater, par exemple, dans mes travaux sur les phrases conditionnelles [THOM 07 ; THOM 08a].

L'un des aspects des travaux de Swales qui a peut-être eu le plus de répercussion sur les études ultérieures en ESP est son analyse de l'introduction de l'article de recherche. Dans cette analyse, il décortique dans le détail l'organisation rhétorique des introductions. Il distingue trois « tours » (*moves*), qui sont : a) établir un territoire, b) établir une niche, c) occuper la niche. Chaque tour est constitué d'un certain nombre de « pas » (*steps*). Swales démontre la fonction précise de tous ces « pas » et souligne le rôle hautement

[6] Le plan IMRED devient aussi de plus en plus courant dans les sciences humaines et sociales, cf. Breivega *et al.* (2002).

stratégique de l'introduction dans l'article de recherche scientifique. Je reviens sur les détails de ce modèle et sur son éventuelle pertinence pour l'analyse de l'organisation rhétorique de la communication de colloque [THOM 03 ; THOM 05b] dans le chapitre 8.

Dans tous mes travaux sur le discours scientifique, je me suis largement appuyée sur la démarche (d'origine sociétale) proposée par Swales. Le regroupement de textes au sein des genres spécifiés de cette manière permet la constitution de corpus très homogènes (cf. chapitre 5 *supra*). Une analyse linguistique de textes entreprise dans cette optique s'oriente vers le genre textuel comme trait constitutif d'une communauté de lecteurs, en observant la façon dont certaines pratiques langagières sont partagées et mutuellement reconnaissables. Cela permet aussi une analyse linguistique fine des traits discriminants entre les textes regroupés de cette manière (cf. *supra*). Toutefois, une fois devant les corpus, c'est l'analyse linguistique qui prend le pas sur toute préoccupation sociologique. Mon approche reste résolument textuelle. Les textes demeurent l'objet de départ, la source d'observation de mes analyses linguistiques. Il s'agit surtout d'une approche textuelle, ancrée dans la linguistique de corpus : le corpus représente un support incontournable et non pas uniquement un moyen de vérifier des hypothèses.

6.3.1 Genres scientifiques en contraste

Toute caractérisation étant par essence différentielle, la caractérisation d'un genre s'effectue nécessairement par champs de contraste. Ces contrastes peuvent être établis de diverses manières : les variables de départ étant sélectionnées en fonction des hypothèses à vérifier.

De nombreux genres sont utilisés conjointement par plusieurs disciplines. Afin de délimiter la spécificité, par exemple, d'un corpus d'articles de recherche dans le domaine de la biologie, il est très utile de le comparer à un corpus d'articles d'une autre discipline proche. Dans la série de travaux sur la

caractérisation des présentations scientifiques de colloque, nous avons ainsi comparé les communications orales des chercheurs en médecine avec celles des chercheurs en physique [THOM 05a]. On a vu que les conventions épistémologiques exerçaient une influence forte sur la forme et l'organisation de la présentation orale. Les présentations orales des chercheurs en physique contiennent, par exemple, beaucoup plus d'indications métatextuelles sur la structuration de leurs exposés que celles des chercheurs en médecine : une différence que nous avons attribuée au format très conventionnel et standardisé des présentations médicales. Les participants à un colloque médical savent exactement à quoi s'attendre et ont peu besoin d'être guidés.

Une autre méthode, que l'on peut qualifier de verticale (cf. chapitre 5 *supra*) consiste à comparer l'emploi de différents genres scientifiques au sein d'une même communauté disciplinaire. L'approche contrastive permet de bien cibler l'adaptation d'un texte sur les attentes d'une communauté dans une situation énonciative bien spécifique et de procéder ainsi à l'analyse des traits discriminants. C'est cette méthode que j'ai essentiellement employée dans mes travaux sur l'emploi des phrases conditionnelles où j'ai comparé, par exemple, les fonctions des phrases conditionnelles dans les communications de colloque avec leurs fonctions dans les articles de recherche scientifique [THOM 08a] ou encore leurs fonctions dans un genre journalistique (l'éditorial) par rapport à leurs fonctions dans l'article de recherche [THOM 09b].

Dans mes travaux j'ai souvent aussi privilégié des analyses contrastives anglais/français [THOM 07 ; THOM 08c ; THOM 09a]. Toutefois, dans le cas de l'étude contrastive des genres scientifiques, l'établissement de cette comparaison n'est pas toujours aisé. La mondialisation de la science a conféré à la langue anglaise un statut de *lingua franca* scientifique. Dans certaines disciplines scientifiques, telles les sciences de la nature, les articles majeurs sont presque tous publiés dans des revues anglophones. On accorde plus de valeur aux articles publiés en anglais qu'à ceux publiés par exemple en français. Le genre de l'article scientifique de biologie en français aura par

conséquent un statut bien différent de celui de l'article de biologie rédigé en anglais. Pour l'analyse contrastive de l'article de recherche en anglais et en français ceci pose évidemment un problème de « comparabilité ». Même s'ils portent la même appellation, les genres ne renvoient pas forcément à la même réalité sur le plan social. C'est pour cette raison aussi qu'il est primordial de pouvoir consulter les spécialistes de la discipline, avant de constituer le corpus d'étude (cf. chapitre 5 *supra*).

Toutefois, cette tendance à faire de l'anglais une *lingua franca* scientifique signifie qu'une proportion importante des travaux publiés dans les revues ou présentés oralement lors des colloques sont le fait de chercheurs non anglophones. Une autre approche contrastive intéressante consiste à faire des comparaisons au sein d'un même genre entre les travaux en anglais produits par les chercheurs anglophones d'une part et par les chercheurs non anglophones d'autre part [THOM 05d ; THOM 08b].

6.3.2 Genres et pédagogie

Sur le plan pédagogique, une approche générique de l'analyse textuelle passe par une reconnaissance et une évaluation des caractéristiques prototypiques d'un genre.

> *Exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience. If all high probability expectations are realized, the exemplar will be viewed as prototypical by the parent discourse community. (Swales 1990 : 58)*

Dans cette approche, un genre textuel donné n'aura pas seulement une certaine similarité de contenu, mais possédera également certaines configurations organisationnelles et stylistiques communes. C'est ainsi le rôle du professeur d'anglais spécialisé de familiariser les étudiants et/ou les chercheurs novices avec les conventions génériques partagées par la communauté de locuteurs en question :

The imparting of genre knowledge involves increasing awareness of the conventions of writing[7]*, and teaching students to produce texts that by following the conventions appear well-formed and suitably structured to native-speaker readers.* (Dudley-Evans 1995 : 182)

Les intérêts d'une telle approche me semblent manifestes. Un rédacteur non anglophone ou un chercheur novice qui tente de faire publier un article de recherche pourra tirer grand profit de ces modèles de genres, qui fournissent certaines indications fiables sur les formes linguistiques et les fonctions rhétoriques tout en soulignant les aspects socioculturels relatifs à l'article de recherche. J'emploie régulièrement ces méthodes dans mes cours aux doctorants et aux étudiants « Master of Science » à l'institut Télécom (« *Scientific writing: structure and constraints* »). J'ai également conçu un certain nombre d'exercices, fondé sur la notion de genre dans les articles de recherche pour le logiciel Navlilire[8].

Toutefois, il existe quelques dangers inhérents à ces approches, surtout en ce qui concerne un certain degré de prescriptivisme. Même si les modèles schématiques de « *Moves* » ne sont pas censés représenter un carcan mais plutôt une configuration susceptible d'être modifiée selon les besoins du rédacteur, un rédacteur manquant de confiance en soi pourrait être tenté de suivre ces modèles à la lettre sans vraiment chercher à comprendre les considérations théoriques à l'origine des choix opérés. Il existe aussi un risque que le genre conditionne non seulement la façon de présenter la recherche, mais, également, la question de savoir quels aspects de la recherche présenter.

Le fait que les travaux des chercheurs « natifs » sont souvent présentés comme des modèles posent aussi un certain nombre de problèmes plus délicats. Les styles d'argumentation peuvent varier d'une culture linguistique à une autre [THOM 05a ; 05e]. Or selon Dudley-Evans *(*cf. citation ci-dessus*) ;* le respect des conventions génériques passe aussi par le respect des normes de style

[7] Comme je l'ai déjà souligné, les approches génériques en ESP ne se sont évidemment pas réservées aux textes écrits. Il n'empêche que, jusqu'à récemment, la plupart des recherches entreprises dans ce domaine portaient sur les documents écrits.

[8] Le logiciel Navilire (Couto *et al.* 2005 ; Lundquist 2008), est un outil d'aide à la compréhension de textes complexes pour les étudiants de français, d'anglais et d'espagnol langue étrangère.

d'écriture anglo-saxonne. Ce serait le rôle du professeur d'ESP de veiller à ce que les travaux de leurs étudiants soient *« suitably structured to native-speaker readers ».* Même si le chercheur novice anglais ou américain doit aussi « apprendre » comment écrire d'une manière propre au genre scientifique considéré, il semble évident que le chercheur non anglophone risque d'être désavantagé. Mais on peut aussi se demander si le gommage éventuel des spécificités culturelles est une chose positive pour la science. Certains commentateurs ont même qualifié comme un type de « linguicisme » l'intolérance manifestée par quelques éditeurs anglo-saxons vers d'autres styles d'écriture (Canagarajah 1999). Je reviens sur ces questions liées à la culture linguistique des chercheurs dans le chapitre 8.

Malgré ces réserves, les approches de genre me semblent globalement très pertinentes pour les cours d'ESP. Tout d'abord, l'impact du genre peut être décisif dans l'identification (et la compréhension) de la macrostructure rhétorique des textes scientifiques. Une sensibilisation à ces aspects peut aider considérablement les étudiants (doctorants ou jeunes chercheurs) à décortiquer et ainsi comprendre les travaux de recherche dans leur domaine. Elle peut aussi les aider dans l'élaboration de leurs propres travaux de recherche (thèses ou articles). Mais les bénéfices d'une approche de genre ne s'arrêtent pas au niveau de l'organisation rhétorique. Dans mes propres cours, j'essaie aussi de privilégier un enseignement de la syntaxe anglaise fondé sur le genre. Les étudiants en troisième cycle ont souvent de bonnes notions de grammaire anglaise mais ne savent pas en revanche adapter leurs connaissances en fonction des besoins de genres. Une sensibilisation à l'emploi de la voix passive ou à l'emploi du présent simple dans le contexte précis de l'article de recherche peut ainsi être très utile. Différents genres de textes exigent des comportements linguistiques différents. Ce qui est considéré comme acceptable pour un essai philosophique ne sera pas forcément acceptable dans un rapport formel. Les étudiants apprennent à réfléchir sur l'adéquation des structures grammaticales en jeu et non pas uniquement sur leur degré de « correction » formelle.

Partie D

Analyser les textes spécialisés

Chapitre 7 – Applications de la structure informationnelle

Résumé

La première partie du chapitre est consacrée à différentes applications de la notion de progression thématique. Après avoir détaillé certaines applications en didactique (7.1.1) j'examine plus spécifiquement les liens entre le genre textuel et le ou les types de progression thématique mis en place : progression à thème constant, progression linéaire, progression à thème dérivé (7.1.2). Mon hypothèse de départ est la suivante : différents genres de discours font appel à des stratégies informationnelles appropriées. Autrement dit, les types de progression thématique employés et les éléments linguistiques susceptibles d'apparaître en tant que thèmes peuvent varier selon le genre de texte abordé [THOM 99a ; THOM 00 ; THOM 08d].

La deuxième partie du chapitre (7.2) porte sur le choix de marqueurs de la structure informationnelle dans différents textes scientifiques. J'étudie non seulement la fréquence mais également le type d'utilisation de certaines structures dites « spécialisées », telles que les phrases clivées, l'extraposition, la diathèse passive et les énoncés existentiels, dans différents genres de textes scientifiques (oraux et écrits). Diverses comparaisons sont réalisées (7.2.1) afin de mieux caractériser l'influence précise de la situation énonciative sur les stratégies informationnelles adoptées, par exemple [THOM 01 ; THOM 04d ; THOM 05d]. Dans la dernière partie du chapitre, je compare la manière dont ces stratégies sont utilisées par les chercheurs anglophones et non anglophones (7.2.2). Les locuteurs non anglophones sont-ils sensibles au même degré à l'influence de la situation énonciative ? Une de mes hypothèses de départ est que ce sont précisément leurs difficultés à adapter des stratégies informationnelles appropriées au contexte d'énonciation qui les différencient de leurs confrères « natifs ».

Chapitre 7 – Applications de la structure informationnelle

7.1 La progression thématique appliquée

La théorie de la progression thématique ébauchée par F. Daneš (1974) est présentée de façon assez détaillée dans le chapitre 2 de ce mémoire. Ce modèle a fait l'objet d'un certain nombre de remaniements, mais il a surtout été à l'origine de nombreux travaux en didactique. Je propose de faire quelques remarques sur ces travaux ainsi que sur l'utilisation faite de cette théorie et de ses prolongations dans ma propre recherche et dans mon enseignement. Je passerai ensuite aux applications potentielles de la progression thématique dans l'analyse de textes spécialisés. À travers une analyse de deux genres de textes très différents, article de recherche primaire et article de vulgarisation scientifique [THOM 99a], je démontre comment les types de progressions thématiques adoptés varient en fonction des objectifs spécifiques de ces deux genres.

7.1.1 Applications en didactique

Le modèle de Daneš a été repris par de nombreux chercheurs. Certains l'adoptent dans son intégralité et d'autres proposent des affinements. Parmi ces derniers, B.-L. Dubois (1987), par exemple, propose quelques modifications à la typologie de Daneš. D'abord, Dubois simplifie cette typologie en supprimant le troisième type de progression – la progression à thème dérivé. Pour elle, la progression linéaire et la progression à thème constant sont les deux types principaux. Cependant, à l'intérieur de ces deux catégories, elle isole deux directions possibles. Soit le nouveau thème est « simple » dans la mesure où il émane d'une seule source, soit il s'agit d'un thème multiple – le thème est dérivé de plus d'une source. Tout comme Daneš, elle accepte l'idée qu'un nouveau thème peut être « contigu » (« *contiguous* ») ou « décalé » (« *gapped* »), dans le cas où la reprise d'un élément rhématique ou thématique se fait à distance.

Les raffinements proposés par Dubois me semblent utiles dans l'ensemble. La typologie initiale de Daneš reste certes un peu abstraite, dans la mesure où les exemples sont tous des exemples construits. L'analyse de la progression thématique dans le discours authentique soulève naturellement des problèmes d'analyse dont plusieurs sont directement abordés dans l'article de Dubois. L'insistance de Dubois sur la fréquence relative de thèmes « décalés » par rapport aux thèmes contigus me semble significative. En revanche, il me parait nécessaire de maintenir la catégorie de progression à thème dérivé. Au contraire de la progression constante et de la progression linéaire, les progressions dites dérivées permettent de lier les référents de plusieurs énoncés à une entité extérieure, spécifiée ou non. Ce type de configuration figure relativement souvent dans les textes scientifiques et permet d'organiser des énumérations autour d'un point de départ donné. Dans l'exemple suivant c'est le titre d'une sous-section d'un rapport (*« The Different partners »*) qui joue le rôle d'un hyperthème, fournissant les thèmes des différentes propositions successives.

> *(1) The different partners*
>
> *Space System Loral, a branch of Loral Corporation is responsible for the space segment and launching aspects.*
>
> *Alcatel Espace will provide the payload modules."*
>
> *Qualcomm is in charge of specifications and development of gateway and user terminals.*
> [THOM 00 : 91]

Dans ce cas précis, le lecteur est aidé dans ses tentatives pour suivre cette progression thématique par la présence d'un schéma sur la composition du consortium. La progression à thèmes dérivés permet ainsi au scripteur de convertir cette construction arborescente en une suite de phrases sémantiquement homogènes.

Plusieurs linguistes français ont souligné le potentiel de la théorie de la progression thématique dans le domaine de la linguistique appliquée : J.-M. Adam (1985), B. Combettes (1977, 1978, 1983), Combettes et R. Tomassone (1988), G. Kassai (1976) et D. Slakta (1975). Cependant, tous rappellent aussi que les principaux types de progression thématique identifiés par Daneš se

trouvent très rarement à l'état « pur ». Combettes, par exemple, insiste sur le fait que les ruptures, auxquelles fait brièvement référence Daneš sont en réalité très fréquentes et qu'il faut s'attendre à en trouver de nombreuses dans les textes longs (1978 : 81). La notion de progression thématique semblerait difficile aussi à appliquer à des textes non « planifiés ». À l'exception de Dubois, qui travaillait sur les conférences transcrites (monologiques) semi-préparées, la plupart des travaux utilisant les concepts de progression thématique se fondent sur des corpus de textes écrits très planifiés. Dans un oral spontané, habituellement caractérisé par des digressions, des hésitations et des phrases inachevées, une analyse de la progression thématique semblerait en effet difficile à mettre en œuvre. (cf. Grobet 2002)

Si le modèle de la progression thématique n'est pas utile pour l'analyse de tout texte, je crois néanmoins que ce modèle présente un intérêt certain dans l'apprentissage de l'expression écrite en milieu scolaire. L'ouvrage de Combettes et Tomassone de 1988 permet par exemple de détailler un véritable éventail d'applications didactiques découlant de ces théories. Dans cet ouvrage, destiné essentiellement aux enseignants du secondaire, ils proposent d'étudier la densité et la répartition de l'information dans les manuels scolaires. Ils se limitent à ce qu'ils appellent le texte *informatif*, ce qui se traduit dans le cadre précis de leur étude par une analyse des manuels de géographie et d'histoire de niveau quatrième et cinquième. Selon les auteurs, les difficultés de compréhension de la part des élèves ne sont pas toujours dues à une méconnaissance du lexique ou à une incapacité à déchiffrer une structure syntaxique complexe ; les difficultés résident surtout dans l'incapacité de l'élève à suivre l'organisation et la répartition de l'information dans un texte donné. Ils soulignent l'importance de l'organisation thématique dans les processus de compréhension et de production de textes. La reconnaissance de tel ou tel type de progression dans un passage peut être considérée « comme l'un des paramètres de l'acte de compréhension » (1988 : 10).

Toutefois, s'il est important de pouvoir reconnaître un certain type de progression thématique, il est aussi important de pouvoir le reproduire soi-même. Comme le soulignent Combettes et Tomassone (*ibid. :* 100), les différents types de progression thématique correspondent à des objectifs informationnels différents. Il est donc important de familiariser les étudiants avec ces différents fonctionnements, afin qu'ils soient non seulement capables de les reconnaître mais aussi de les reproduire dans leur propre écriture.

Les travaux de Combettes et Tomassone portent sur l'enseignement du français et sur un enseignement dans la langue maternelle. Or les mêmes problèmes se posent et à un niveau plus critique dans l'apprentissage d'une langue étrangère. En effet, les étudiants d'une langue étrangère sont souvent tellement angoissés par les questions de précision phrastique locale, qu'ils ne pensent pas aux enchaînements interphrastiques ou aux découpages en paragraphes. Il faut ajouter à cela que dans les grammaires pédagogiques et ouvrages didactiques destinés aux étudiants en anglais langue étrangère la dimension organisationnelle est rarement abordée. Or il est très important de montrer aux rédacteurs en anglais langue étrangère qu'un texte (même un texte rédigé dans une langue étrangère !) est le produit d'un tissage, plutôt qu'une agglomération de bribes de signification. Une sensibilisation à la notion de progression thématique peut ainsi s'avérer fort utile.

Dans le cadre de ma thèse et des travaux entrepris à sa suite, j'ai appliqué l'analyse pragoise de la progression thématique (cf. chapitre 2, section 2.1) à un corpus de rapports techniques rédigés en anglais par mes étudiants francophones. À partir de ces analyses j'ai proposé une typologie des différentes maladresses relevées au plan de l'organisation thématique, à savoir : 1) les difficultés dans l'identification de thèmes, 2) le problème de rhèmes insuffisamment accentués, 3) le mauvais ordre de présentation des éléments donnés et nouveaux, et 4) les progressions thématiques illogiques.

Je ne propose pas de revenir ici sur les détails de cette typologie (cf. par exemple, [THOM 98]), mais il s'agit des concepts que je mets régulièrement en pratique dans mes enseignements de l'expression écrite.

Depuis quelques années, j'essaie systématiquement de familiariser mes étudiants d'anglais langue étrangère (de jeunes adultes) avec la notion de progression thématique. Je demande aux étudiants d'identifier les thèmes phrastiques dans un extrait de leur propre rédaction et d'établir un schéma montrant les différents types de progression qui y figurent. Quand il y une « rupture », je leur propose d'analyser les conséquences de la rupture.

Une rupture peut être plus ou moins programmée dans le discours. L'émetteur peut choisir par exemple d'orienter une proposition autour d'un référent « nouveau », c'est-à-dire un élément qui n'a pas été l'objet d'une mention préalable et qui n'est pas non plus facilement récupérable par le récepteur[1]. Toutefois, si le rédacteur signale au récepteur qu'il s'agit effectivement d'un nouveau point de départ dans le discours, et si ce nouveau thème reste en quelque sorte prévisible dans le contexte précis du texte en question, il n'y aura pas de véritable rupture. Si, par exemple, le nouveau thème est précédé d'un connecteur de type contrastif ou s'il coïncide avec le début d'un nouveau paragraphe typographique, on peut estimer que le lecteur est prévenu du changement de direction que le rédacteur souhaite impulser. Cependant, très souvent la rupture est involontaire. Le lecteur n'arrive pas à trouver un référent à sa portée (dans le texte ou en dehors) qui lui permette d'arriver à une interprétation logique, et les conséquences sont donc potentiellement néfastes pour sa compréhension du texte. En travaillant sur la progression thématique d'un texte et en réfléchissant sur le discours en construction au-delà de la phrase, j'essaie de montrer aux étudiants comment « jongler » avec la langue et mieux gérer le flux d'information.

[1] Cf. Chapitre 2 (section 2.3.1) pour une discussion sur la distinction importante entre le statut de l'information et le statut des référents dans l'univers du discours.

La notion de progression thématique me sert aussi dans mes activités de relecture de documents scientifiques à l'Institut Télécom. Je suis fréquemment sollicité par mes collègues enseignants-chercheurs francophones pour la relecture de leurs articles scientifiques en anglais avant publication dans des périodiques spécialisés. Il s'agit en général de chercheurs avec un bon niveau d'anglais (qui maîtrisent parfaitement le vocabulaire spécialisé de leur spécialité scientifique : informatique, téléphonie mobile, systèmes d'information). En revanche, leurs écrits ne possèdent pas toujours la fluidité qu'on trouverait chez un locuteur anglophone. On ressent parfois un certain piétinement du texte (qui peut provenir d'une série des rhèmes insuffisamment accentués) ou au contraire une certaine confusion dans les enchaînements phrastiques proposés (progression thématique illogique). Un recours au modèle de la progression thématique (ainsi qu'au concept de donné-nouveau[2]) m'aide souvent à expliquer aux scripteurs la raison de la maladresse ressentie. Dans de très nombreux cas, une légère réorganisation des blocs informationnels suffit pour rendre la cohérence à ces passages défaillants.

Le modèle de progression thématique me parait particulièrement bien approprié pour l'analyse d'articles scientifiques. Ce sont des textes planifiés et hautement structurés. Ils contiennent des sections et sous-sections clairement délimitées. Il devient alors possible d'examiner le ou les types de progression thématique mis en place dans une partie spécifique de l'article et de déterminer les stratégies informationnelles qui ont été favorisées. Comme l'a souligné, par exemple, Slakta, les schémas élaborés par Daneš ne sont pas de structures abstraites :

> *Il ne s'agit pas seulement d'une méthode descriptive ; les schémas proposent une explication.* (1975 : 39)

Les différents types de progression thématique correspondent à des objectifs informationnels précis. Dans le cas de l'article de recherche, on peut ainsi comparer l'utilisation des différents types de progression thématique employés en fonction des visées rhétoriques des différentes sections de l'article

[2] L'ordre donné-nouveau est généralement considéré comme un ordre cognitif non marqué (Vande Kopple 1986). Cf. Chapitre 2 *supra.*

(« Introduction », « Méthodes », « Résultats » etc.). Afin de préciser le profil informationnel global de l'article scientifique, on peut également envisager des comparaisons entre articles scientifiques et d'autres genres de textes spécialisés. Dans les parties suivantes, je vais décrire certains de mes travaux qui explorent ces relations entre structure informationnelle et genre textuel. Je propose en particulier d'examiner les liens entre les considérations de genre (cf. chapitre 6 *supra*) et les types de progression thématique mis en place dans les textes spécialisés.

7.1.2 Applications aux textes spécialisés

L'analyse de la structure informationnelle peut jouer un rôle de premier plan dans l'étude des genres textuels. Il me semble qu'il existe essentiellement deux champs d'action communs pour une approche qui privilégie à la fois une réflexion sur le genre et une analyse informationnelle. Nous pouvons en premier lieu analyser la nature des éléments thématiques ou rhématiques au sein d'un genre particulier. Quels sont les éléments les plus susceptibles de figurer dans le groupe thématique au sein d'un genre textuel particulier ? Dans l'article de recherche, par exemple, les thèmes sont essentiellement réalisés par des noms communs inanimés, souvent des groupes nominaux complexes (Vande Kopple 1994). On constate également la présence fréquente de constructions existentielles avec un *there* non référentiel [THOM 01, THOM 04c]. Dans les articles de vulgarisation, en revanche, les thèmes sont plus souvent composés de noms propres (le nom d'un chercheur par exemple), de pronoms personnels ou de pronoms démonstratifs. Le ton est plus « personnel » car le journaliste a besoin d'impliquer ses lecteurs dans son article [THOM 99a].

Une autre piste consiste à analyser les types d'enchaînements thématiques qui figurent dans un genre de texte particulier afin de cibler les configurations préférentielles d'organisation. Les différents genres de texte, selon leur visée discursive, vont faire appel à des stratégies informationnelles appropriées.

Selon P.H. Fries, les schémas de progression thématiques sont même étroitement liés aux genres :

> *Different patterns of thematic progression correlate with different genres, i.e. patterns of thematic progression do not occur randomly but are sensitive to genre.* (1995 : 319)

Autrement dit, les types de progression thématique employés vont varier selon le genre[3] de texte abordé.

J'ai souhaité vérifier cette hypothèse en examinant deux genres de discours scientifique spécifiques dans le domaines des télécommunications : l'article de recherche primaire et l'article de vulgarisation scientifique [THOM 99a]. J'ai constaté en effet que ces deux genres scientifiques présentent des profils thématiques assez différents. Globalement la progression linéaire est plus courante dans l'article de vulgarisation, tandis que la progression à thème constant est plus courante dans l'article de recherche primaire. Mes résultats vont dans le même sens que ceux de K.N. Nwogu (1990) et Nwogu et T. Bloor (1991) portant sur différents genres de texte médical. Cependant ceci n'est guère surprenant car, comme J.M. Swales (1990), par exemple, l'a démontré, c'est le genre, plus que le domaine ou sujet du texte, qui joue un rôle déterminant dans le fonctionnement textuel.

Deux facteurs principaux – deux facteurs qui sont d'ailleurs déterminants dans la définition d'un genre (cf. chapitre 6 *supra*) – peuvent expliquer ces préférences : l'objectif du texte et le type de lectorat visé. L'une des fonctions majeures d'un article de vulgarisation est d'expliquer le fonctionnement d'un processus ou d'un appareil à un lectorat non spécialiste. La progression linéaire semblerait très bien adaptée à cette tâche. L'extrait suivant sur la transmission par fibre optique tiré d'une célèbre revue de vulgarisation scientifique, *The New Scientis*t, en fournit une illustration :

[3] Mon utilisation du terme genre est détaillée dans le chapitre 6.

(2) The signal must be boosted every 40 miles or so.

↙

<u>In the past this</u> was achieved by electronic repeaters which convert the light pulses into electronic signals, boost them and convert them back before sending them on.

↙

<u>But electronic repeaters</u> must be pre-set to process a certain amount of information.

↙

<u>So increasing the information capacity of the cable</u> would mean replacing every repeater along its length.

↙

<u>This problem</u> can be avoided by using optical repeaters.

↙

<u>These</u> are stretches of cable that contain the light-emitting element erbium and a 'pump' laser.

↙

<u>The laser</u> keeps...[THOM 99a : 142/143]

Dans cet extrait, le journaliste[4] s'efforce d'expliquer les avantages de l'amplification optique du signal transmis par la fibre optique par rapport à une amplification électronique. Le thème de chaque phrase est issu directement du rhème de la phrase précédente : Th1→Rh1

↓

→Th2 (Rh1)→Rh2

En conséquence, les explications apparaissent claires et faciles à suivre.

Le recours fréquent à la progression linéaire dans les articles de vulgarisation semblerait aussi étroitement lié à la question des connaissances partagées. Les lecteurs d'un texte de vulgarisation auront, de manière générale, peu de connaissances sur le sujet en question. Pour cette raison, l'auteur sera souvent contraint de développer son argumentation à partir des informations qui ont déjà figuré dans les parties rhématiques des propositions antérieures afin d'augmenter le stock des connaissances mutuelles. Le thème issu d'un rhème

[4] Il se peut aussi que l'auteur de l'article de vulgarisation soit non pas un journaliste mais un chercheur qui a choisi de vulgariser ses travaux afin d'atteindre un lectorat plus vaste.

précédent introduit une précision ou un complément d'information permettant d'approfondir un topique donné par paliers successifs.

En revanche, dans l'article de recherche, où le degré de connaissances partagées sera normalement très élevé, le rédacteur aura moins besoin de puiser dans les informations présentées dans la partie rhématique précédente. Ce sont des spécialistes qui s'adressent à des spécialistes. Les auteurs peuvent ainsi plus facilement prendre appui sur un stock de connaissances déjà partagées avec leur lectorat [THOM 99a :143]. Les explications pédagogiques du type de l'exemple (2) ci-dessus, organisées autour de progressions linéaires, sont pour cette raison plus rares et réservées en général aux parties les plus argumentatives de l'article de recherche : la partie « Introduction » et la partie « Discussion »[5]. Dans l'article de recherche on trouve au contraire un nombre élevé de séquences descriptives, particulièrement dans les sections de l'article consacrées à la présentation de la méthodologie ou à l'énumération des résultats. Or une organisation textuelle fondée sur la progression à thème constant est très adaptée pour ces fonctions descriptives et c'est pourquoi nous trouvons un recours fréquent à la progression à thème constant, comme dans l'extrait « méthodologique » (3) ci-dessous :

(3) Two different plastic clumps hold down the silicon vessel array (…)

↓

<u>The smaller clamp</u> has a total contact area of approximately 5.6mm.

↓

<u>while the larger clamp</u> has a contact area or approximately 18.6mm.

↓

<u>To provide an adequate clamping force</u>, four screws with springs push down on the corners of the clamp.

L'emploi de la progression à thème constant ici permet à l'/aux auteur(s) de focaliser sur le déroulement précis de l'expérimentation.

[5] Les articles de recherche scientifiques sont souvent divisés en quatre grandes sections : « Introduction », « Méthodologie », « Résultats », « Discussion/Conclusions ». Les fonctions principales de ces différentes sections sont détaillées en chapitre 6 section 6.3).

Le troisième type principal de progression thématique identifié par Daneš, la progression à thème dérivé, bien que moins fréquent dans les deux genres scientifiques examinés, permet aussi de remplir des fonctions très spécifiques. Ce type de progression semble particulièrement adapté à la présentation de certaines explications nécessitant une énumération de plusieurs points connexes. Il permet par exemple au scripteur de convertir une construction arborescente en une suite de phrases sémantiquement homogènes, comme dans l'exemple (1) cité ci-dessus. On retrouve fréquemment aussi ce type de configuration thématique dans les « Introductions » aux articles de recherche, dans le paragraphe où l'auteur fait l'inventaire de toutes les parties de son article :

(4) Section 1 describes (…).

Section 2 reviews (…).whilst section 3 concludes and discusses (..)

Dans de tels cas c'est l'article lui-même qui joue le rôle de l'hyperthème :

T (hyper-thème) = Article

↙ Th1 (Section 1) → Rh1 ↓ Th2 (Section 2→Rh2 ↘ Th3(Section 3)→Rh3

Le thème dans chaque phrase correspond à la section de l'article concerné, et la partie rhématique apporte des renseignements sur le contenu de chaque partie.

Après ces réflexions sur les applications de la progression thématique pour la didactique et pour l'analyse de genres scientifiques, je propose maintenant de porter mon attention sur les choix de marqueurs de la structure informationnelle dans différents genres de textes scientifiques. Je vais faire état d'un certain nombre de mes travaux qui ont pour objectif l'analyse des formes syntaxiques utilisées pour exprimer la structure informationnelle au sein de différents genres scientifiques.

7.2 Structures spécialisées et genre

Je reste convaincue qu'une analyse informationnelle implique nécessairement une prise en compte du contexte linguistique et extralinguistique complet. Il est

primordial de caractériser les choix de structuration liés aux conditions de la mise en texte : autrement dit des choix qui ne sont pas contraints par le système de langue, mais qui résultent du contexte de production du texte concerné. Or ce contexte peut être plus ou moins contraignant selon les cas. Dans le cas du discours scientifique, ces contraintes sont habituellement très fortes. Le genre discursif a un impact important sur les stratégies informationnelles adoptées et sur les structures syntaxiques employées pour mener à bien celles-ci. C'est à cet impact du genre sur la structuration informationnelle que je vais m'intéresser dans le reste de ce chapitre.

Afin de cerner l'influence des contraintes de genres il est nécessaire de faire une recherche très ciblée. Avec E. Rowley-Jolivet, spécialiste de la communication scientifique orale, j'ai effectué toute une série de comparaisons entre l'article de recherche et la communication orale correspondante dans le cadre de congrès [THOM 01, THOM 04d, THOM 05d]. Notre objectif premier était d'analyser les formes utilisées pour exprimer la structure informationnelle au sein de ces deux situations énonciatives particulières. Il s'agissait, en particulier, d'étudier la répartition et l'utilisation des structures spécialisées, telles que les phrases clivées, l'extraposition, la diathèse passive et les énoncés existentiels, (cf. chapitre 4 *supra*). En adoptant une approche qui s'inspire à la fois de la linguistique textuelle et des approches de genre (cf. chapitre 6), nous avons souhaité comparer non seulement la fréquence mais également le type d'utilisation de telles structures dans ces deux genres de texte (oral et écrit). L'objectif était de pouvoir mieux caractériser l'influence précise de la situation énonciative sur les stratégies informationnelles adoptées.

7.2.1 Contraintes de genre : communication de congrès *vs* article de recherche

L'article de recherche scientifique et son pendant oral la communication de colloque partagent un certain nombre de caractéristiques génériques dans le mesure où ils sont tous les deux « exploités » par la même communauté discursive (la communauté de chercheurs). Toutefois, il est évident que lorsque le chercheur scientifique parle de ses travaux, il s'exprime d'une façon

complètement différente de celle qu'il utilise lorsqu'il écrit. Ces différences sont liées à certaines spécificités des deux situations énonciatives – des spécificités détaillées dans plusieurs de mes travaux (par exemple, [THOM 03 ; THOM 04c et d]. On peut identifier au moins trois contraintes importantes, spécifiques à la situation de colloque : les contraintes cognitives liées au traitement de l'information en temps réel (par opposition à un traitement en « différé »), les rapports avec des auditeurs présents dans la salle (par rapport aux lecteurs « absents ») et la gestion de deux canaux de communication : la communication visuelle (les transparents ou les diapositives Powerpoint) et la communication verbale (le commentaire de l'orateur).

La contrainte sur le traitement en temps réel (Chafe 1987) aura une incidence importante sur le type d'information que l'orateur pourra transmettre, et sur les structures dont il aura – ou n'aura pas – besoin pour ce faire et pour faciliter le traitement de l'information par son public. Le discours scientifique possède en général une haute densité informationnelle. Dans l'article de recherche scientifique, le chercheur doit inclure dans son texte une somme importante de connaissances établies (les références à la littérature du domaine concerné sont très importantes). Sur le plan syntaxique, cette haute densité d'information est obtenue surtout grâce à la nominalisation et aux groupes nominaux complexes. Or les groupes nominaux complexes de l'article rédigé sont difficiles à produire ainsi qu'à assimiler en temps réel. Dans les communications orales les locuteurs seront contraints de manière générale d'employer des phrases simples et des groupes nominaux thématiques très courts.

Les rapports entre les interlocuteurs sont aussi assez différents. Dans l'article de recherche écrit, il existe un grand souci d'objectivité. Le chercheur est habituellement très prudent dans tout ce qu'il affirme et se présente à son lecteur comme un observateur objectif, dépourvu de toute subjectivité qui pourrait diminuer la validité et la portée scientifiques de son propos. Cependant, devant un auditoire physiquement présent, un monologue trop impersonnel serait mal adapté au contexte (d'où on peut déjà prévoir un emploi du passif

plus réduit). Il faut aussi noter que dans de nombreuses disciplines, les orateurs scientifiques ne lisent pas un texte rédigé mais commentent librement leurs projections visuelles sans recourir à des notes. Cela semble être particulièrement le cas dans les sciences expérimentales et dans la médecine, où il s'agit donc d'un oral certes préparé, mais communiqué d'une manière assez spontanée.

La troisième contrainte importante concerne la nature même de l'événement scientifique. Dans une présentation de colloque, l'orateur doit habituellement intégrer deux canaux de communication : le canal verbal et le canal visuel. Lors des présentations (surtout dans les disciplines scientifiques dites « dures »), un très grand nombre de transparents visuels sont projetés à l'écran et accompagnent en permanence les paroles de l'orateur. Ces transparents utilisent parfois la sémiotique linguistique, mais le plus souvent font également appel à des sémiotiques non linguistiques – des images ou des symboles mathématiques – pour transmettre l'information (Rowley-Jolivet 2002). Une des tâches importantes de l'orateur de congrès est de veiller constamment à bien intégrer les informations visuelles dans la trame linguistique de son commentaire verbal.

Comment alors le locuteur gère-t-il toutes ces contraintes ? Nous avons surtout cherché à cibler la manière dont ces contraintes se répercutaient sur les stratégies informationnelles adoptées par les locuteurs à travers une analyse ascendante ou *« bottom-up »* des formes morphosyntaxiques employées dans les deux genres.

7.2.2 Corpus et méthodologie

Afin d'évaluer l'impact précis de ces contraintes énonciatives, nous avons cherché à neutraliser les autres variables. Il était particulièrement important

pour nous d'écarter le variable disciplinaire[6]. Le corpus employé pour l'étude concerne de ce fait seulement une discipline scientifique : la physique. Il se compose de la transcription de l'enregistrement sur bande vidéo de neuf communications orales faites par des chercheurs scientifiques anglophones lors d'un congrès en physique et des neuf articles de recherche correspondant à ces communications dans la version publiée dans les Actes de ce même congrès (40 000 mots approximativement). Dans chaque sous-corpus, les thèmes de recherche sont ainsi identiques. Une deuxième constante concerne les chercheurs eux-mêmes. L'orateur est dans tous les cas l'un des auteurs de l'article écrit. Une troisième constante est la manifestation scientifique : le congrès en physique.

Ce corpus a fourni le point de départ pour une série des comparaisons portant sur les marqueurs de la structuration informationnelle dans les deux genres. Dans un premier volet de l'étude [THOM 01] nous avons analysé l'emploi du passif, de l'extraposition, des énoncés existentiels, de l'inversion et des clivées (phrases clivées en *« it »* et phrases pseudo-clivées). Dans un deuxième volet [THOM 04d], nous avons étendu notre analyse à différentes formes de structures détachés : dislocations avec reprise anaphorique, topicalisations sans reprise et topicalisations avec terme introducteur. Je propose de faire quelques commentaires sur les résultats de ces analyses dans la section suivante.

[6] Dans le chapitre 8 *infra*, je présente un autre volet de ma recherche portant justement sur les influences éventuelles de différentes cultures disciplinaires et des traditions sur la structuration des communications scientifiques [THOM 05a].

7.2.3 Distribution de structures spécialisées dans les deux genres

Structure	Article	Comm	Acad	Conv
Passifs	480 32%	172 9%	25%	2%
Extraposition Enoncés existentiels	54 13	9 51	+ ≅	– ≅
Inversion	0	61	–	+
Clivées	2	2	+	–
Pseudo-clivées	0	31	–	+
Détachements : Terme introducteur Avec reprise pronominale Topicalisation (sans reprise)	 7 0 0	 9 6 15	–	+

Tableau 1. Comparaison quantitative. Distribution des structures syntaxiques spécialisées (tableau adapté de [THOM 01]).

Les deux premières colonnes du tableau indiquent le nombre d'occurrences pour chacune des structures spécialisées dans les deux sous-corpus : articles de recherche (Article) et communications orales (Comm.). À titre de comparaison, nous avons aussi reporté dans les deux dernières colonnes des indications sur la fréquence relative de ces structures dans les textes académique écrites (Acad) et la conversation courante (Conv). Ces indications sont tirées de l'ouvrage de Douglas Biber *et al.* (1999) sur la grammaire de l'anglais écrit et parlé, ouvrage basé sur l'analyse d'un corpus de quarante millions de mots. Les signes « + », « – » ou « = » signifient que la structure en question est plus fréquente, moins fréquente ou d'une fréquence équivalente dans les corpus examinés.

Ce tableau montre clairement des différences importantes dans l'utilisation des structures syntaxiques spécialisées dans l'anglais scientifique écrit et dans les présentations orales. Des structures qui sont très fréquentes dans un mode de communication sont absentes ou peu sollicitées dans l'autre. D'une façon générale, si nous comparons les chiffres relevés dans notre propre corpus avec les indications fournies dans l'ouvrage de Biber *et al.*, il semble évident que les caractéristiques de la communication scientifique de congrès s'apparentent beaucoup plus à celles de la conversation qu'à celles de la recherche académique écrite, et ce quelle que soit la structure spécialisée examinée.

Je propose ici de reprendre le cas de quatre de ces structures. Tout d'abord je ferai quelques brèves remarques sur les passifs et sur l'extraposition, des structures qui ne sont vraiment productives que dans la communication scientifique écrite. Ensuite, je ferai quelques remarques sur le cas des pseudo-clivées et les détachements à gauche : des structures qui semblent en revanche presque exclusivement réservées à la communication scientifique orale.

Passifs

Le pourcentage élevé de passifs dans le corpus d'articles de recherche (32%) correspond aux résultats d'autres études sur l'anglais scientifique (cf. chapitre 4 *supra* et Banks 1994 ; Tarone *et al.* 1998). Cependant, il est important de souligner que l'emploi fréquent de cette structure concerne seulement l'écrit. Dans notre corpus de communications orales, le passif est beaucoup moins employé (seulement 9% des propositions). Comme je l'ai souligné dans le chapitre 4, l'une des fonctions premières du passif est de faire correspondre la structure thématique et la structure syntaxique. En employant un passif l'émetteur thématise ce qui serait apparu en position d'objet syntaxique et par conséquent aurait été normalement rhématique dans une phrase active. Sans doute y a-t-il une corrélation à établir entre la proportion importante d'éléments thématiques inanimés à l'écrit par rapport au nombre élevé d'éléments thématiques animés (pronoms personnels) dans la communication scientifique orale. Dans une communication de colloque, la personne du chercheur, en tant qu'orateur, est bien présente. En revanche, dans les articles de recherche, les auteurs en tant que personnes sont en retrait. Le choix d'articuler une phrase autour d'un repère concret inanimé nécessitera dans bien des cas, suivant l'orientation du verbe, une construction passive.

Extraposition

Tandis que l'extraposition est peu exploitée dans les communications scientifiques de colloque, la structure semble jouer un rôle important dans les articles de recherche (cf. tableau 1). Nous avons proposé plusieurs explications pour cette distribution très différente dans les deux genres.

L'extraposition est une opération qui permet de déplacer en fin d'énoncé une proposition entière et de créer de cette manière une répartition plus équilibrée de la charge informative (cf. chapitre 4). Voici un premier exemple du corpus d'articles de recherche :

> *(5)* ***it*** *is advantageous to stress the blocks to the maximum electric field that can be safely tolerated without breakdown.* (Article)
>
> (Il est avantageux de soumettre les blocs au maximum de champ électrique qui peut être toléré sans danger avant claquage)

C'est la proforme *it* qui remplit le rôle de thème dans la phrase matrice, son contenu se trouvant inclus dans la proposition extraposée. Dans une version non extraposée le thème serait surchargé :

> *(5*) to stress the blocks to the maximum electric field that can be safely tolerated without breakdown is advantageous.*

L'utilisation de l'extraposition évite cette surcharge informationnelle en début de proposition et respecte ainsi ce que les grammairiens anglais ont appelé le principe de *« end-weight »* (Quirk *et al.* 1985). À cela s'ajoute le fait qu'un thème long et complexe comme celui de la version non extraposée contient certainement plusieurs éléments nouveaux. En ayant recours à l'extraposition, le scripteur garantit ainsi la rhématisation de ces éléments.

Une autre fonction majeure de l'extraposition a trait à la modalisation des énoncés. L'extraposition permet au chercheur de thématiser le jugement qu'il porte sur la validité de l'information contenue dans la proposition extraposée :

> *It is advantageous to…*
>
> *It can be implied that…*
>
> *It seems that…*

En outre, en étant positionné dans la phrase introduite par *it*, ce jugement a l'air plus « objectif ». Le chercheur peut ainsi concilier son souci d'objectivité avec la nécessité rhétorique de convaincre le lecteur. Remarquons en passant que dans la linguistique systémique, l'extraposition est justement appelée *« thematised comment »* (Thompson 1996 : 129-130).

Afin d'expliquer la quasi absence de l'extraposition dans les communications orales, nous avons comparé tous les passages d'articles où figurent l'extraposition avec les passages d'oral où les mêmes informations scientifiques sont présentées, afin de voir quelles structures syntaxiques sont utilisées dans ce second mode de communication. Dans certains cas, une telle comparaison s'avère impossible car l'information contenue dans la proposition extraposée de l'écrit est simplement omise à l'oral. Ceci est particulièrement le cas lorsque l'extraposition est employée dans deux sections de l'article scientifique – « Introduction » et « Méthode expérimentale ». Ces sections sont souvent très brèves, voire absentes, dans la communication orale pour des raisons à la fois de contraintes de temps et d'attentes génériques, entraînant des différences de contenu entre l'article et la communication orale correspondante. Dans les cas restants, où les mêmes informations scientifiques sont présentes dans les deux modes, il n'y a pas de parallélisme syntaxique : aucune des occurrences d'extraposition à l'écrit ne donne lieu à une proposition extraposée à l'oral.

L'une des raisons en est que, tandis que l'emploi de l'extraposition permet au rédacteur, à l'écrit, d'éviter le positionnement d'une proposition complexe en tant que thème, cette stratégie n'est simplement pas pertinente à l'oral Dans les communications orales, le thème est souvent un simple pronom et la modification du groupe nominal est de manière générale très « allégée » et fractionnée en plusieurs segments brefs. L'orateur n'a donc pas besoin de l'extraposition. Si nous comparons l'exemple d'extraposition déjà cité ci-dessus en (5), avec le passage oral qui correspond « scientifiquement » (6), nous voyons que, dans ce dernier, il n'y a pas de surcharge au niveau des thèmes. Les thèmes des deux énoncés sont déjà courts et la progression linéaire opérée dans la deuxième proposition permet au locuteur de fournir progressivement à l'assistance les explications nécessaires.

> *(5) it is advantageous to stress the blocks to the maximum electric field that can be safely tolerated without breakdown.* (Article)

(Il est avantageux de soumettre les blocs au maximum de champ électrique qui peut être toléré sans danger avant claquage.)

(6) you ***want to stress it as far as possible***

Th1→Rh1

↓

and ***that****'s limited by the breakdown field*)

→Th2 (Rh1)→Rh2

(on veut le contraindre autant que possible, et c'est limité par le champ de claquage)

L'emploi de l'extraposition comme stratégie de modalisation ne s'applique pas non plus aux communications orales. Les orateurs ont tendance à exprimer leurs jugements et commentaires d'une manière plus directe, en utilisant la première personne et un verbe à la forme active. Ils n'ont pas besoin de recourir à cette stratégie de modalisation.

ARTICLE	COMMUNICATION ORALE
It is shown that	We've demonstrated
It seems that.	We think that probably
It is possible to define	We can define
It seemed that	We've developed an opinion
It has been found that	We find in practice
It can be implied	We know
It is believed that	We believe
It is shown	We've shown
It is hoped that X	X was a thought

Tableau 2. L'extraposition. Articles *vs* communications de colloque (tableau adapté de [THOM 04d]).

Structures détachées

Comme on pouvait s'y attendre, le phénomène de détachement est beaucoup plus important dans les communications de colloque que dans les articles. Dans le chapitre 4 j'ai brièvement décrit trois grandes catégories de structures détachées : les détachements avec reprise pronominale, les détachements sans reprise pronominale et les détachements avec terme introducteur que je reprendrai ici [THOM 04d] :

(7) The beam, (…) it's still slower

(8) The repetitions rates, we've been up to a kilohertz…

(9) Moving on to the carbon fibre cathodes, these are…

Le détachement avec terme introducteur est le seul à apparaître également dans le corpus d'articles écrits. Dans ces exemples, le terme introducteur, la plupart du temps *« for »* (ou *« with »* également dans le cas de la communication orale) semble permettre à l'énonciateur d'activer un nouveau référent, qui, tout en étant lié au sujet général de la séquence, signale une délimitation particulière de ce sujet. Considérons un exemple de notre corpus écrit :

(10) For ferroelectric ceramic dielectrics in the frequency range of consideration here, the dominant contribution to the overall dielectric comes from…

L'auteur vient d'évoquer les difficultés pour trouver un matériel adapté dans le cadre de ses expériences scientifiques. Il passe en revue maintenant les différentes possibilités. Le terme introducteur *« for »* peut être compris comme *« en ce qui concerne »* et permet d'ouvrir un nouveau cadre thématique ou s'inscrit le reste de l'énoncé.

Les détachements avec reprise pronominale et les détachements sans terme introducteur n'apparaissent pas du tout dans notre corpus d'articles**.** Nous avons avancé plusieurs explications pour cette distribution très différente dans les deux genres scientifiques. Les détachements avec reprise pronominale semblent particulièrement bien adaptés aux exigences de discours en temps réel. En segmentant ainsi son discours, au moyen d'un thème préjeté qui sera repris par un pronom dans l'énoncé prédicatif, le locuteur facilite sa propre tâche ainsi que celle de l'auditeur.

(11) the beam, although it's relativistic, it's still slower than the speed of light…

Les détachements de ce type permettent plus de souplesse quant à l'ordre selon lequel les choses sont dites – rendant possible, par exemple, comme dans ce cas, l'insertion de parenthèses/commentaires avant la reprise anaphorique.

Les détachements sans reprise (et sans terme introducteur) sont aussi très productifs dans le corpus de présentations orales. Il n'y a donc pas d'intégration syntaxique, le référent thématisé n'est pas directement impliqué dans des relations prédicatives avec d'autres constituants. Examinons les deux exemples suivants :

(12) Magnetic field, as you see we've got insulating field here, at the end we have some wave field.

(Champ magnétique, comme vous voyez nous avons le champ isolant ici, au bout on a le champ associé à l'onde.)

(13) The repetition rates, we've been up to a kilohertz for avalanche pulsers and on field effect transistors the repetition rate is limited only by the power supply. (P9)

(Les taux de répétition, on est arrivé jusqu'à 1 kHz pour les générateurs d'impulsion en avalanche et pour les transistors à effet de champ, le taux de répétition n'est limité que par l'alimentation.)

Ici la mise en avant des groupes nominaux (*« magnetic field »*, *« repetition rates »*) fonctionne comme un type d'annonce, signalant un nouveau départ thématique. Chaque groupe porte un contour ascendant suivi d'une légère pause. Il s'agit en fait des titres de différents schémas qui figurent sur les transparents que l'orateur est en train de projeter. Les référents sont donc déjà accessibles, visuellement, dans le contexte énonciatif, mais ils ont besoin d'être activés afin d'être intégrés dans le commentaire verbal.

Il semblerait donc que ce dernier type de détachement soit surtout motivé ici par la nécessité pour l'orateur d'intégrer l'information visuelle dans son discours. La sémiotique visuelle permet la représentation simultanée de plusieurs référents – tous disposés dans un seul schéma ou une seule configuration spatiale, ou tous présents sur la même diapositive. Dans son commentaire verbal, en revanche, l'orateur est contraint d'introduire ces référents de manière séquentielle, et de les intégrer l'un après l'autre, dans son texte. En les plaçant ainsi en tête de l'énoncé (et en même temps en les montrant physiquement sur le transparent) le locuteur les identifie explicitement comme thème.

Il est important de souligner que l'utilisation de ces différentes constructions détachées ne correspond en aucune manière à de simples lapsus ou ruptures dans la trame discursive, et c'est en partie pour cette raison que j'emploie délibérément le terme de *détachement* et non celui de *dislocation*. Leur usage répond à de véritables besoins sur le plan communicatif. Dans le discours

spontané et même dans les monologues semi-préparés de cette étude, l'utilisation des structures détachées peut être considérée comme résultant d'un effort d'adaptation de l'énonciateur aux données de la situation.

Phrases pseudo-clivées

Notre corpus scientifique comporte peu d'occurrences de la clivée en *it*, mais on y trouve, en revanche, un certain nombre de phrases pseudo-clivées, toutes dans les communications orales. Les pseudo-clivées peuvent être de deux types (cf. chapitre 4 *supra* ; Weinert et Miller 1996) : les pseudo-clivées régulières et les pseudo-clivées inversées, les deux étant courantes dans les communications orales. On peut avancer plusieurs hypothèses pour expliquer la productivité de ces structures dans cette situation spécifique d'oralité.

La pseudo-clivée régulière en *« wh »*, tout en respectant la structure informationnelle non marquée thème/rhème permet d'accentuer la division par une coupure, ou clivage, de l'énoncé de départ :

> *(14) well, what we're talking about here // is pulses in the voltage range of 10 to 40 KV with a peak current in the range of (…)*

Par cette « mise en scène », la structure permet de rendre plus saillantes les informations véhiculées par le rhème. En outre, le flot discursif est ralenti. La pseudo-clivée permet à l'orateur de focaliser l'attention sur la précision à venir : elle prépare le public ou l'avertit qu'une information importante est imminente. Sans la pseudo-clivée, cette information aurait reçu moins d'impact. Autrement dit, l'emploi de pseudo-clivées permet de fournir aux récepteurs des consignes précises pour l'interprétation de l'information. Lorsque l'on compare les passages comportant une pseudo-clivée à l'oral avec les passages correspondants de l'article où la même information scientifique est présentée, on ne trouve pas une telle focalisation explicite.

En effet cette structure est souvent employée par les orateurs dans les colloques scientifiques pour attirer l'attention sur les résultats de leurs travaux :

> *(15) What we found was that (…).*
>
> *So what we're saying here is (…).*
>
> *What this work has shown is (…).*

La communication de colloque a souvent une fonction épistémologique complémentaire distincte de celle de l'article de recherche. L'accent est un peu moins mis sur les connaissances déjà établies (les références à la littérature du domaine sont beaucoup plus rares). En revanche, les chercheurs présentent souvent des travaux en cours, dans lesquels ils tiennent à mettre en évidence l'originalité ou la nouveauté de leur contribution scientifique

On pourrait également considérer l'emploi de la phrase pseudo-clivée comme une stratégie discursive interactive. En employant une pseudo-clivée, le locuteur répond en quelque sorte, par le focus, à une question implicite que son interlocuteur serait en train de se poser (« Qu'est-ce que vous avez fait ensuite ? Qu'est-ce que vous voulez dire ? Ce que nous voulons dire c'est que... ») De cette manière, l'orateur fait intervenir un certain rapport dialogique dans son monologue.

Tout comme la pseudo-clivée régulière, les clivées inversées servent à segmenter le flot discursif continu en de petits segments et à créer des unités textuelles clairement bornées à gauche et à droite. Toutefois, à la différence le la pseudo-clivée régulière, c'est l'élément initial qui est focalisé dans la clivée inversée et la proposition subordonnée apparaît en deuxième plan :

(16) it's really important how fat the dipole is, ***that's*** *what this slide shows.*

Comme je l'explique dans mon article sur les phrases clivées [THOM 09a), de telles séquences sont difficile à traduire directement en français (****Cela*** *est ce que montre cette diapositive*). Les pseudo-clivées inversées sont de manière générale très rares en français. Il est en effet difficile d'attribuer un focus à un élément intra-prédicatif en tête de phrase. Il serait plus habituel de traduire la clivée inversée du (16) de la manière suivante :

(17) C'est cela que démontre cette diapositive.

K. Lambrecht invoque même l'existence en français d'une *« powerful grammatical constraint against the* co-mapping of subject and the pragmatic relation focus » (1994 : 22).

Les clivées inversées remplissent des fonctions très précises dans la communication de colloque en anglais. Dans notre corpus, le constituant clivé (X) de la clivée inversée est dans la majorité des cas le démonstratif *Thi*s ou *That*. Le fait que les démonstratifs soient fréquemment employés dans la présentation orale scientifique me semble pareillement pertinent : l'orateur est constamment en train de montrer le chemin afin que les auditeurs puissent s'orienter dans le discours. La pseudo-clivée inversée avec *That*, comme en (16) ci-dessus, se rapporte davantage au canal verbal et résume ce qui vient d'être dit, tandis que la pseudo-clivée avec *This* a une fonction déictique et pointe sur une information nouvelle communiquée par le canal visuel.

(18) ***This*** *is kind of like what it all looks like in the LAB (…),*

Here's an OIL tank here.

L'utilisation complémentaire des deux démonstratifs dans les clivées inversées sert ainsi à intégrer les deux canaux de communication, le verbal et le visuel, et contribue considérablement à la clarté du discours. Comme les pseudo-clivées régulières, les clivées inversées ralentissent le flot du discours et créent des sections plus « digestes », répondant ainsi aux contraintes particulières du traitement en temps réel d'un discours oral à haute densité informative. L'article scientifique écrit n'est pas éphémère, et n'a sans doute pas autant besoin de ces rappels et signaux pendant le déroulement du discours.

7.2.4 Comparaison entre chercheurs anglophones et non anglophones

Comme les paragraphes ci-dessus l'ont mis en évidence, les stratégies informationnelles adoptées par les chercheurs peuvent être assez différentes dans ces deux situations d'énonciation. Même si les articles de recherche et les communications de colloque relèvent de genres assez proches, un certain nombre de contraintes liées à la spécificité de la situation orale d'un colloque (cf. *supra*) obligent les conférenciers à recourir des stratégies assez différentes de celles adoptées dans un contexte d'écriture. Ceci a pour conséquence une répartition très différente des structures syntaxiques spécialisées dans les deux cas.

Dans un autre volet de mes travaux, j'ai souhaité voir si ces différences étaient aussi frappantes dans les articles de recherche et les communications de colloque des chercheurs non anglophones. Autrement dit, les locuteurs non anglophones sont-ils sensibles au même degré à l'influence de la situation énonciative ? Une de mes hypothèses de départ est que ce sont précisément les difficultés des chercheurs non anglophones à adapter des stratégies informationnelles appropriées au contexte d'énonciation qui les différencient de leurs confrères « natifs ».

Les analyses présentées dans les sections ci-dessus ont été ainsi étendues à un corpus plus large, comportant cette fois-ci les articles et les transcriptions des communications de colloque de deux groupes de chercheurs : des chercheurs anglophones « natifs » et les chercheurs non anglophones rédigeant et présentant leurs travaux de recherche en anglais. Les chercheurs non natifs avaient pour la plupart un niveau assez fort en anglais. Ils faisaient très peu de vraies « fautes » à un niveau syntaxique superficiel. Toutefois, comme je l'ai démontré dans plusieurs travaux antérieurs, la maîtrise de la langue à un niveau syntaxique local ne garantit pas forcément la réussite de la communication à un niveau plus global (cf. chapitre 1 *supra*). Il est aussi très important de voir dans quelle mesure les configurations syntaxiques adoptées répondent aux exigences de la situation de communication.

Une autre série d'analyses entreprise avec E. Rowley-Jolivet a ainsi cherché à cerner les éventuelles différences de comportement syntaxique entre les deux groupes de locuteurs [THOM 05d]. Nous avons comparé neuf communications orales et les articles correspondants de chercheurs scientifiques anglophones (NS) avec neuf articles de recherche et neuf communications orales des chercheurs non natifs (NNS). Afin de réduire au maximum le biais lié aux influences disciplinaires, les thèmes de recherche étaient quasiment identiques. Il s'agissait dans les deux cas de congrès de physique.

Nos recherches ont démontré que la distribution des structures spécialisées parmi les deux groupes de locuteurs est très différente. Comme le tableau ci-dessous (tableau 3) le montre, les chercheurs NS et NNS n'emploient pas ces structures dans les mêmes proportions. Les chercheurs NNS semblent beaucoup moins différencier les deux situations de communication que leurs confrères anglophones.

	NS Comm.	NNS Comm.	NS Articles	NNS Articles
Passives	171 (8,3%)	212 (17,1%)	480 (32,2%)	405 (31,7%)
Extraposition	8 (0,4%)	18 (1,5%)	51 (3,6%)	31 (2,4%)
Inversion	64 (3%)	21 (1,6%)	1	2
pseudo-clefts	28 (1,4%)	1 (0%)	0	0
It-clefts	3	0	2	0
Existential *There*	52 (2,5%)	26 (2%)	13 (0,9%)	12 (1%)

Tableau 3. Occurrences des structures syntaxiques spécialisées : NS *vs* NNS (tableau adapté de [THOM 05d : 46]).

Les passifs, par exemple, sont fréquents dans les articles des deux groupes de chercheurs, 32% et 31% respectivement. En revanche dans les communications orales, ils sont employés très différemment. Les chercheurs NS emploient seulement 8% de passifs dans leurs communications orales tandis que les chercheurs NNS emploient le double (17%). Dans le cas de l'extraposition, nous pouvons observer une tendance similaire. Bien que l'extraposition soit moins fréquemment employée à l'oral que dans les articles dans les deux groupes, cette différenciation est beaucoup moins nette chez les NNS. Il semble que les chercheurs NNS sont plus sous l'emprise de l'écrit et ont donc tendance à réemployer la syntaxe de l'écrit dans leurs communications orales. A titre d'exemple, voici un extrait d'un article publié dans les Actes et le passage correspondant à l'oral de l'un ces chercheurs NNS de l'échantillon étudié :

	Article NNS	*Communication NNS*
(19)	The mass reduction benefits associated to the adoption of ion propulsion can be, in particular, retained a key factor also when referred to small, resource limited satellites (called also lightsats) … The massive utilisation of lightsats (generally in the range 300-800 kg), alone or in constellations, has been diffusely proposed during the last years for a number of applications. (Article 11)	Electric propulsion can be a key issue when applied to small satellites... In particular, the small satellites in the range between three hundred and eight hundred kilograms have been proposed, alone or in constellations to perform a variety of missions... the drag compensation of LEO spacecraft that are generally devoted to earth observation... (Présentation 11)

Tableau 4 [THOM 05d : 50]

Nous pouvons constater dans cet exemple un fort parallélisme lexico-grammatical entre la version écrite et la version parlée. L'extrait de la communication de colloque contient trois passifs, dont un passif réduit. Les deux premières constructions passives en particulier (*« when applied to »*, et *« have been proposed »*) sont calquées de façon très marquée sur la version écrite. On constate aussi que les phrases sont longues et les thèmes phrastique sont des thèmes inanimés complexes (la personne du chercheur scientifique est absente dans cet extrait). Tout cela fait que cet extrait est très difficile à assimiler en temps réel

Une analyse de la distribution des autres structures syntaxiques du tableau 2 permet d'étayer cette hypothèse de l'emprise de l'expression écrite dans les communications orales des chercheurs non anglophones. On constate par exemple que les énoncés existentiels en *there* sont beaucoup moins usités dans les communications des chercheurs NNS que dans celles du groupe anglophone. Les phrases pseudo-clivées et les inversions sont quant à elles quasi absentes dans les communications scientifiques orales du groupe NNS. Comme je l'ai souligné ci-dessus (et à travers plusieurs publications, [THOM

01, THOM 04c et d, THOM 05d]), de telles structures remplissent des fonctions très utiles dans le contexte de la communication orale du colloque. Ces structures facilitent surtout le traitement en temps réel d'une charge informative lourde. Les chercheurs NNS semblent avoir des difficultés à adapter leur syntaxe à ce type de discours oral. En revanche, ils ont tendance à se servir (à contre-emploi) de structures spécialisées de l'écrit et à peu exploiter les structures spécialisées adaptées au contexte de la communication de colloque.

Il est vrai qu'il ne s'agit pas de véritables erreurs de la part des locuteurs NNS. On pourrait arguer que ce ne sont que des points de « détail ». Un locuteur peut très bien se passer, par exemple, de l'utilisation des phrases pseudo-clivées dans une présentation orale. En revanche l'accumulation de tous ces points de détails risque à terme d'avoir une influence néfaste sur l'efficacité rhétorique du message. Nous l'avons formulé de la façon suivante dans un article paru dans *English for Specific Purposes* en 2005 :

> *ll these features, when taken in isolation, may not seem crucial – misapplying or underusing the specialised information packaging structures does not constitute a serious 'error' or make the discourse incomprehensible. When such features co-occur, however, they have a negative impact on the rhetorical appropriacy of the discourse, its cognitive processing, and its persuasiveness.* [THOM 05d : 60]

C'est pour cette raison que nous préconisions d'accorder plus d'importance à l'acquisition d'une certaine compétence pragmatique nécessaire à la manipulation de la structure informationnelle dans un contexte générique spécifique. Dans les cours d'anglais de spécialité il est nécessaire de sensibiliser les apprenants à l'importance d'une analyse fine des conditions et des contraintes de genre, afin d'adopter les stratégies informationnelles les plus appropriées. Le « modèle » syntaxique de l'écrit (l'article de recherche) adopté par les locuteurs NNS de notre échantillon s'est avéré être une entrave à leur efficacité communicative.

Chapitre 8 – Organisation rhétorique et textes spécialisés

Résumé

Ce chapitre présente un programme de recherche sur les textes spécialisés, mené essentiellement en collaboration avec E. Rowley-Jolivet (université d'Orléans), qui examine les modèles d'organisation textuelle axés sur la rhétorique, et plus particulièrement la rhétorique de la communication orale en science (8.1). Dans le cadre d'une approche tout d'abord macroscopique, nous nous attachons à dégager un schéma rhétorique générique (« *move structure* ») des sections introductives de la communication de congrès en anglais (8.1.1).

Par le moyen d'une série d'études sur corpus, cette recherche démontre la présence de stratégies récurrentes spécifiques à la communication de congrès, stratégies qui sont différentes de celles de l'article de recherche [THOM 03 ; THOM 05b]. À la suite de ces analyses nous proposons un modèle rhétorique de *moves* adapté à l'analyse des sections introductives de la communication de congrès (8.1.2). Un volet suivant de cette recherche vise à évaluer l'influence respective de la culture disciplinaire et de la culture linguistique sur la mise en œuvre de ce modèle de *moves*. [THOM 05a]. En limitant notre analyse à deux disciplines scientifiques (médecine et physique) et à deux « groupes culturels » (des chercheurs d'origine anglo-américaine et des chercheurs d'origine francophone), nous avons cherché à cerner le poids respectif de ces deux variables selon la manière dont ce modèle est adopté ou modifié par ces différents groupes de locuteurs (8.1.3).

La deuxième partie de ce chapitre porte sur la relation entre syntaxe et rhétorique (8.2). En adoptant une approche ascendante (et microscopique), je mets plus particulièrement l'accent sur la manière dont les orateurs de communications scientifiques gèrent les relations interpersonnelles avec leur public à travers les stratégies syntaxiques choisies [THOM 04c ; 05b]. Nous évaluons une fois encore les influences respectives de la culture disciplinaire et de la culture linguistique sur ces choix. Seront examinés l'emploi de pronoms personnels (8.2.1), ainsi que de certaines structures syntaxiques potentiellement « interactives » (8.2.2) [THOM 05a].

Chapitre 8 – Organisation rhétorique et textes spécialisés

Les travaux présentés dans ce chapitre se situent dans la tradition de la « nouvelle rhétorique », où l'accent est surtout mis sur les aspects contextuels et sur la finalité argumentative du discours étudié. Pour A. Freedman et P. Medway, par exemple, la rhétorique est définie comme *« speech or writing used to achieve some purpose within a social situation »* (1984 : 2). Dans le cas des discours non littéraires, tel le discours scientifique[1], l'objectif est ainsi de décrire la fonction des textes par rapport aux pratiques socialisées qui les ont engendrées.

Les recherches entreprises sur les textes scientifiques ces dernières décennies ont démontré que la conception traditionnelle du discours scientifique comme neutre, objectif et monologique doit être remise en question (cf. chapitre 6 *supra*). Par son texte, le chercheur doit convaincre son public (ses pairs) de l'intérêt et du bien-fondé de sa recherche. Cet aspect argumentatif est étroitement lié à un aspect interactionnel. Le chercheur doit se positionner dans un milieu compétitif. Tout en promouvant sa propre recherche, il est important pour le chercheur de montrer qu'il fait partie d'une communauté, et de situer ses propres travaux dans un travail d'ensemble. Les articles scientifiques par exemple, ne sont pas ainsi des textes purement informatifs, mais ils sont aussi porteurs de divers procédés rhétoriques qui visent à persuader. Comme l'explique K. Hyland (1999), la réussite d'un texte dans ce contexte passera par le respect de certaines conventions linguistiques et sociales reconnues et partagées par la communauté des pairs :

> *The persuasiveness of academic discourse does not depend upon the demonstration of absolute truth, empirical evidence or flawless logic. Texts are the result of socially situated writers and are persuasive only when they employ social and linguistic conventions that colleagues find convincing.* (1999 : 99)

[1] N.B. Le concept de *discours scientifique* est employé dans ce mémoire comme l'équivalent du terme anglais *academic discourse* (cf. chapitre 6 *supra*). Il n'est pas forcement restreint à l'étude des sciences dites dures.

Les modèles d'organisation textuelle axés sur les schémas rhétoriques (*move structure*) de J. Swales (1981 ; 1990 ; 2004) se sont révélés de puissants outils pour explorer l'organisation rhétorique interne des textes scientifiques (cf. section 6.4). Toutefois, ces analyses, comme la grande majorité des études du discours scientifique, se portent presque exclusivement sur la communication scientifique écrite. Les modèles de *moves* de Swales ont été appliqués, par exemple, aux articles de recherche et les nombreuses autres études entreprises dans le sillage de Swales ont porté elles aussi sur l'écrit. Les caractéristiques linguistiques et rhétoriques de la communication scientifique orale ont fait l'objet de beaucoup moins d'études. En collaboration avec E. Rowley-Jolivet[2], je me suis donc penchée sur une analyse de l'organisation rhétorique de la communication de congrès.

8.1 Une analyse de l'introduction de la communication de congrès

Les communications de congrès jouent un rôle clé dans la recherche universitaire. Il s'agit d'un genre que tout chercheur doit maîtriser s'il veut faire carrière dans la recherche. Les doctorants et les chercheurs novices sont en particulier très demandeurs de conseils sur la conduite à adopter, surtout quand il s'agit de faire une présentation en anglais. Les applications pédagogiques de ces études sont en effet nombreuses, dans la formation à la communication de recherche des jeunes chercheurs et des locuteurs non anglophones. Or il existait jusqu'à très récemment[3] très peu d'études de corpus pour alimenter ces applications didactiques. Une des raisons pour expliquer ce désintérêt est sans doute liée aux difficultés de constituer un corpus oral (cf. Rowley-Jolivet 1998). Mais on peut aussi se demander si en tant qu'« objet » d'étude, il est possible d'appréhender et de décortiquer une présentation orale de la même façon qu'un texte scientifique écrit.

[2] Cette recherche s'appuie sur le corpus constitué par Rowley-Jolivet (1998), composé d'un nombre important de communications scientifiques, enregistrées en vidéo dans des congrès internationaux dans trois domaines – la géologie, la médecine, et la physique – et ensuite transcrites. Le corpus a été enrichi par de nouveaux enregistrements en 2001 et 2004.

[3] Le corpus de MiCase (http://quod.lib.umich.edu/m/micase/) portant sur le langage parlé utilisé dans le milieu universitaire américain permet en partie de pallier ce manque.

Les communications de congrès ne sont pas soumises aux mêmes types de contraintes formelles et organisationnelles que les articles de recherche. S'il existe presque toujours des contraintes fortes sur le temps de parole accordé, les autres « règles » du genre sont souvent beaucoup plus tacites. Il n'y a pas, par exemple, de feuille de style. Les processus de filtres ou de « garde-fous » pour l'évaluation de la recherche ne fonctionnent pas non plus de la même manière que dans le cas de la communication scientifique écrite. Même si le colloque ou congrès en question se dote d'un comité scientifique, l'évaluation de la proposition de communication est faite au niveau du contenu et non pas en jugeant le degré de conformité de la présentation à une norme générique ou linguistique prédéfinie. Nous l'avons exprimé de la façon suivante :

> *The relative lack of interest in spoken academic genres means that there is a corresponding shortage of material available for deconstruction and genre-based teaching methodologies. Unlike in the case of the research article, there are also no gate-keeping processes or formal criteria to guide scientists unfamiliar with either the genre or the appropriate English usage and to enable them to hone their rhetorical skills. Conference presentation proposals are accepted by the conference committee on the basis of the informational content alone, and not on the appropriateness of the oral presentation itself to the contextual constraints of the genre or on the communicative proficiency of the speaker.* [THOM 05b : 46]

Nous avons donc voulu savoir si malgré cette absence de « règles » explicites, la section introductive de la communication de congrès possède une structure rhétorique bien définie. Le modèle CARS[4] de *move analysis* élaboré par Swales (1990) pour les articles de recherche scientifique (cf. chapitre 6 *supra*) peut-il s'appliquer aussi à l'analyse de la communication de congrès ? Ce sont ces interrogations qui sous-tendent le programme de recherche présenté dans ce chapitre.

8.1.1 Corpus et méthodologie

Nous avons restreint notre étude à l'analyse des introductions des communications de congrès, premièrement afin de pouvoir les confronter au

[4] CARS signifie *Create a Research Space* (Swales 1990 : 140).

modèle de CARS de Swales (*ibid.*), mais aussi à cause du rôle primordial de l'introduction sur le plan rhétorique. C'est surtout dans son introduction que le chercheur doit positionner sa recherche et établir sa propre crédibilité.

Notre analyse a porté sur les sections introductives de quarante-quatre communications de congrès dans trois disciplines scientifiques : géologie, physique et médecine[5]. Dans le premier volet de la recherche, seules ont été analysées les communications de chercheurs anglophones (anglo-américains), afin d'écarter les variables liées aux origines linguistiques (cf. *infra*). Nous avions aussi à notre disposition un deuxième corpus contrastif plus restreint, contenant treize communications de colloque en physique et les articles correspondants publiés dans les Actes du colloque. Ce deuxième corpus permettait de faire des comparaisons directes entre les deux genres.

Notre analyse s'est vite heurtée à une première grande difficulté d'ordre méthodologique. Le lecteur sait où l'introduction commence mais il est plus difficile de fixer sa fin avec précision. À la différence de l'article de recherche, la communication de congrès ne possède pas de sections formellement délimitées[6] (« Introduction », « Méthode », « Résultats »). Afin de nous aider à identifier les bornes des sections introductives, nous nous sommes donc servi de trois types d'indices : le contenu scientifique, les indications métatextuelles données par l'orateur et les indications visuelles présentes sur les diapositives. Un changement de section dans une présentation scientifique, par exemple, s'accompagnera normalement d'un changement de diapositive. Pour les cas les plus délicats, nous avons consulté un spécialiste de la discipline concernée et/ou l'article correspondant dans les Actes s'il était disponible. En tant qu'analystes, nous avons aussi employé un système d'évaluation double jusqu'à ce qu'un consensus soit atteint.

[5] Les détails sur ces corpus peuvent être consultés [THOM 03 ; THOM 05b].
[6] Pour des explications sur la division des articles de recherche suivant le modèle IMRED, voir chapitre 6 *supra*.

Une fois les sections introductives délimitées, nous avons ensuite essayé d'appliquer le modèle argumentatif de *move analysis* postulé par Swales (1990) pour l'article de recherche scientifique. Il pourrait paraître logique de trouver une organisation rhétorique similaire. Les communications de congrès partagent de nombreuses caractéristiques avec les articles de recherche analysés par Swales. Il s'agit dans une grande mesure des mêmes communautés (la communauté des chercheurs universitaires) et il est vrai que l'on peut constater le même mouvement rhétorique global que dans les articles de recherche, c'est à-dire celui du général vers le particulier. Les chercheurs essayent de situer leur objet de recherche particulier par rapport au contexte général afin de montrer la pertinence de leurs travaux et de justifier ainsi leur recherche. Toutefois, les conditions spécifiques de la communication de congrès (cf. section 7.2 *supra*), et en particulier les contraintes liées à son mode de transmission, font qu'il est difficile d'appliquer ce modèle complètement. Nous avons, par exemple, dégagé plusieurs étapes rhétoriques dans la communication de colloque qui ne figurent pas dans le modèle CARS de Swales. L'ordre des *« moves »* proposé par Swales ne semble pas non plus refléter l'ordre de celui des communications de congrès.

Afin de nous guider dans notre analyse, nous nous sommes aussi appuyées sur deux autres modèles d'introduction : celui proposé par S.E. Thompson (1994) pour l'analyse de cours magistraux (« lectures ») et celui de B-L. Dubois (1980) pour l'analyse des conférences biomédicales. Comme ces deux derniers modèles portaient sur des monologues oraux, il semblerait donc logique d'y trouver un certain nombre d'analogies avec notre propre corpus. Les trois modèles sont présentés dans la figure 1 ci-dessous.

SWALES (1990) *CARS Moves model for research articles introductions*		**DUBOIS (1980)** *Biomedical CPs*		**THOMPSON (1994)** *Lectures*	
		A : Listener orientation		**Function 1 : Set Up Lecture Framework**	
		1	To Chairperson	SubF1	Announce topic
		2	To audience	SubF2	Indicate scope
		3	To projectionist	SubF3	Outline structure
				SubF4	Present aims
Move 1 : Establishing a territory		**B. Content orientation**		**Function 2 : Put Topic in Context**	
Step 1	Claim centrality	1	Non-technical	SubF1	Show importance/ relevance of topic
Step 2	Make topic generalizations	2	Technical	SubF2	Relate 'new' to 'given'
Step 3	Review previous research	2a	Subject qualification	SubF3	Refer to earlier lectures
Move 2 : Establish a Niche		2b	Amplification		
St. 1a	Counter-claim	2c	Pre-hypothesis		
1b	Indicate gap	2d	Hypothesis		
1c	Question-raising	2d*	Implications		
1d	Continue tradition				
Move 3 : Occupy the Niche					
St. 1a	Outline purposes				
1b	Announce present research				
2	Announce main findings				
3	Indicate RA structure				

Figure 1. Trois modèles d'introductions dans les genres scientifiques [THOM 03 : 72].

Les modèles de Dubois (1980) et de Thompson (1994) nous ont aussi fourni des indications précieuses pour l'analyse de la structure des communications de congrès. Dans le modèle de Thompson, les catégories métatextuelles concernant la mise en place du cadre général du cours magistral (*Lecture Framework*), par exemple, sont très pertinentes pour l'analyse des communications de congrès. De la même façon, la distinction faite par Dubois entre *Listener orientation* et *Content orientation*, est aussi fort utile. En revanche les sous-catégories de Dubois, conçues dans une perspective tagmémique, semblent beaucoup moins appropriées que les catégories de *moves* et *steps* de

Swales. Le modèle de Thompson n'est pas non plus parfaitement adapté à nos besoins. Comme ce modèle portait sur le discours didactique et non pas sur un discours de recherche universitaire, il ne nous permet pas d'appréhender la visée rhétorique d'une communication de colloque. Dans un cours magistral le professeur cherche plus, par exemple, à synthétiser un domaine de recherche et non pas à promouvoir de nouvelles idées (qui sont peut-être aussi des idées controversées), alors que c'est le cas dans une communication de congrès.

8.1.2 - Une proposition pour un modèle de *« moves »*

Comme aucun des trois modèles consultés n'étaient entièrement satisfaisant, nous avons donc conçu notre propre modèle. Ce modèle est présenté dans la figure 2 ci-dessous.

A. Setting Up the Framework		
1. Interpersonal framework		
	1a.	Listener orientation
And/or	1b.	Acknowledgements
2. Discourse framework		
	2a.	Announce topic
And/or	2b.	Outline structure/Indicate scope
B. Contextualising the topic		
	1a.	Conference context
and / or	1b.	General research context
C. Research rationale		
1. Motivation		
	1a.	Problems/Gaps/Counter-claims
And/or	1b.	Relevance/Centrality/Need
And/or	1c.	Continuation of previous work
2. Response		
	2a.	Question-raising/Hypotheses
and / or	2b.	Preview results or solutions
3. Outline research goals		

Figure 2. Un modèle rhétorique pour les sections introductives de la communication de congrès [THOM 03 : 65].

Les détails de ce modèle sont exposés en [THOM 03] et [THOM 05b]. Je propose ici de commenter brièvement la spécificité de ce modèle en le comparant au modèle CARS de Swales pour les articles de recherche – le modèle qui reste en somme le plus proche du nôtre. Comme je l'ai déjà détaillé dans les chapitres précédents (voir sections 6.2 et 7.2), le mode de transmission fournit un paramètre utile dans la caractérisation d'un genre et les différences entre les deux modèles tiennent dans une grande mesure à l'influence du mode de transmission. Ces différences peuvent être résumées sous trois grandes rubriques : les relations interpersonnelles avec le public dans la salle, les contraintes cognitives liées au traitement de l'information en temps réel et le contexte épistémologique spécifique de la communication de congrès.

L'une des différences les plus frappantes dans l'organisation rhétorique de la communication de congrès concerne l'importance accordée à la mise en place d'un cadre interpersonnel (*Interpersonal framework)*. Devant un auditoire physiquement présent, l'impersonnalité typique de l'article de recherche est moindre. Le fait que les interlocuteurs sont présents dans la salle de colloque ou l'amphithéâtre encourage en revanche l'orateur à instaurer un rapport direct avec eux, une démarche que nous avons appelée *Listener orientation*. Il s'agit de salutations aux auditeurs, des remerciements aux organisateurs ou de toute autre remarque amicale destinée à installer un rapport avec la salle. De tels propos peuvent s'insérer à d'autres moments dans la présentation mais ils semblent particulièrement courants au tout début de l'introduction. Une autre étape rhétorique importante dans la mise en place de ce cadre interpersonnel est la catégorie *Acknowledgements.* La recherche dans les sciences est essentiellement collaborative et il s'agit donc d'une stratégie de politesse où l'orateur reconnaît publiquement l'aide de ses collaborateurs, de ses employeurs ou des agences de subvention :

> *(1) I want personally to thank Dr. {NTD} for his spearheading this project and supplying the patients. I've always been amazed at the interest Dr. {NTD} has in fostering interdisciplinary types of research. I'd also like to thank Dr. {NTD} for his continued support during these studies.* (Présentation en médecine) [THOM 05b : 52]

Bien que souvent présents dans l'article de recherche, ces remerciements figurent d'une façon moins proéminente, insérés par exemple dans une note de bas de page où tout en fin d'article.

Les contraintes liées au traitement cognitif de l'information (pour l'orateur et pour les auditeurs) sont aussi très importantes. Même s'il s'agit d'un monologue planifié et semi-préparé, la communication de congrès se déroule en temps réel. L'orateur doit accommoder la haute densité informationnelle de son propos à ces contraintes et adopter des stratégies organisationnelles appropriées. À la différence du lecteur d'article de recherche, l'auditeur dans l'amphithéâtre d'un congrès est contraint de suivre la communication linéairement. Il est ainsi important d'instaurer le *Discourse framework.* Un composant très important dans la trame rhétorique de la communication du congrès concerne l'annonce de son objectif principal (*Announce topic*) et l'indication de son plan et/ou l'étendue de la communication (*Outline structure / Indicate scope*). Lors d'un congrès, chaque nouvelle présentation entraîne un mouvement de personnes qui sortent et qui entrent. Le début de la présentation doit donc capter l'attention de l'auditoire et fournir un maximum d'informations pour permettre son suivi linéaire. De telles informations sont moins fréquentes dans les introductions de l'article de recherche et, quand elles sont présentes, elles figurent plutôt vers la fin de l'introduction qu'à son début.

À cause des contraintes liées au déroulement de la communication en temps réel ainsi qu'à cause des contraintes de temps de parole imparti, les orateurs sont limités à la fois par la quantité et le type d'information qu'il est possible de transmettre. Les références à la littérature du domaine, caractéristiques des introductions dans les articles de recherche, figurent parfois sur les diapos de l'orateur mais sont rarement incluses dans ses commentaires verbaux. En revanche, les sections introductives des communications de congrès de notre corpus contiennent de fréquentes références aux présentations des autres orateurs. Les orateurs s'efforcent de montrer comment leur propre recherche

fait partie d'une épistémologie plus « locale », partagée avec les auditeurs. Nous avons appelé cette démarche *Conference context*.

> *(2) This paper follows very logically on from the previous one and compliments to the organizers for that and it will save quite a lot in terms of background information.* (Présentation en géologie) [THOM 05b : 53]

L'absence d'un véritable état de l'art est aussi liée à la fonction épistémologique un peu différente de la communication de congrès. L'accent est mis un peu moins sur les connaissances déjà établies. En revanche, les chercheurs présentent souvent des travaux en cours, où ils mettent en évidence l'originalité ou la nouveauté de leur contribution scientifique. Dans le temps de parole qui lui est accordé, l'orateur doit se démarquer et promouvoir l'importance de sa propre recherche. La nouveauté ou l'intérêt de cette recherche sera donc souligné aux dépens d'une présentation détaillée de l'état de l'art ou de références bibliographiques explicites.

Nos recherches ont démontré que le « modèle » présenté dans le figure 2 ci-dessus était suivi d'assez près dans la grande majorité des communications de colloque de notre corpus. Il existe évidemment quelques exceptions et idiosyncrasies. Toutes les catégories ne figurent pas nécessairement non plus dans toutes les communications. Toutefois, malgré l'absence de « règles » explicites, la section introductive de la communication de congrès semble posséder une structure rhétorique bien définie. Nos études ont montré qu'il existe bien des stratégies rhétoriques récurrentes spécifiques à la communication orale, différentes de celles de l'article de recherche.

À ma connaissance, il semble que les seuls travaux portant sur l'analyse d'un discours scientifique oral à l'aide des principes de *move analysis* soient les nôtres. Nos recherches, tout en confirmant la validité générale et la puissance du modèle de Swales, ont permis ainsi de proposer un modèle adapté à l'analyse rhétorique de ce genre oral important qu'est la communication de congrès. Ces résultats fournissent à mon avis un point de départ fort intéressant pour des applications didactiques diverses (cf. *infra*).

Une autre piste de recherche liée à l'identification de ce modèle concerne l'existence d'éventuelles différences entre les disciplines scientifiques. Dans l'élaboration de notre modèle, nous n'avons pas distingué entre les origines disciplinaires des orateurs. Or si les introductions aux communications de congrès possèdent leur structure rhétorique propre, on peut cependant s'attendre à trouver aussi un certain nombre de variations selon les conventions disciplinaires. Les différentes disciplines, selon leurs objets d'étude et leurs épistémologies, peuvent exercer un contrôle « normalisant » sur la façon dont la recherche est présentée et diffusée. Une autre influence potentielle sur les stratégies rhétoriques des orateurs provient de leur origine linguistique. Notre modèle initial était fondé sur l'analyse de présentations par des chercheurs d'origine anglo-américaine. Or, de nos jours, de plus en plus de chercheurs d'origines linguistiques diverses choisissent (ou sont contraints) de présenter leurs travaux en anglais. Une autre question à étudier est celle liée à l'empreinte de la culture linguistique d'origine sur l'organisation de la présentation en anglais. Dans la prolongation de cette première étude, nous avons donc cherché à évaluer le poids respectif de ces deux variables sur les stratégies organisationnelles adoptées par ces différents groupes de locuteurs.

8.1.3 Impact de la culture disciplinaire *vs* impact de la culture linguistique

Afin de structurer cette comparaison, nous avons restreint notre analyse à deux disciplines scientifiques, la médecine et la physique et à deux groupes culturels, des chercheurs d'origine anglo-américaine et des chercheurs d'origine francophone. Travaillant en France, il était évidemment utile pour nous de cerner de plus près les éventuelles différences spécifiques aux chercheurs francophones. Ces comparaisons avaient en plus une application potentielle directe pour moi-même car j'enseigne des modules de présentation scientifique aux étudiants ingénieurs en formation initiale et un autre aux doctorants de diverses disciplines scientifiques à l'Institut Télécom. Il était surtout utile de savoir si l'on constatait d'éventuelles spécificités d'organisation rhétorique chez les intervenants francophones, ou si en revanche les normes disciplinaires et génériques prévalaient sur les variations linguistiques potentielles.

Pour cette étude, nous avons ainsi examiné quarante communications en tout : vingt communications en médecine (Méd), vingt d'intervenants anglophones (EM) et dix d'intervenants d'origine francophone (FM), et vingt communications en physique (Phys), dix d'intervenants anglophones (EP), dix d'intervenants d'origine francophone (FP) :

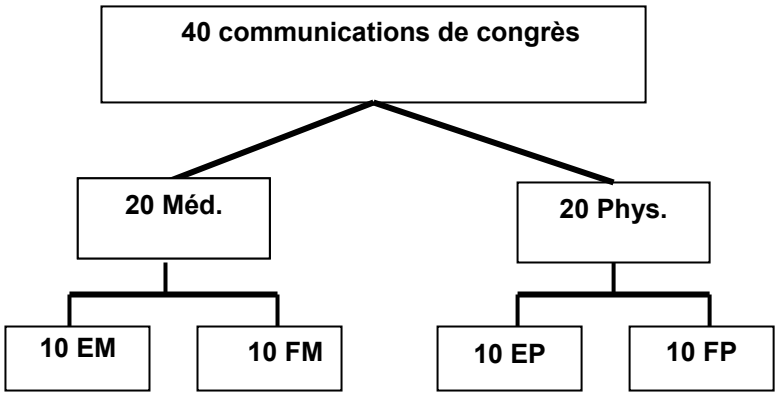


Figure 3 : Corpus contrastif de communication de congrès [THOM 05a]

Les résultats de nos analyses suggèrent que notre modèle rhétorique est solide. Tous les orateurs, quelle que soit leur origine linguistique, prennent en compte les contraintes contextuelles et les objectifs spécifiques des introductions de communications de congrès. Les intervenants francophones ont les mêmes connaissances disciplinaires que leurs confrères anglophones sur les normes et les attentes à l'égard de la structure générale de la présentation. L'approche adoptée par les chercheurs pour présenter leur recherche, qui consiste à la « contextualiser » par rapport à une problématique plus générale et ensuite à montrer la pertinence de leurs propres travaux, est sensiblement la même pour les intervenants anglophones et francophones. Chez les orateurs francophones (en médecine et en physique) nous remarquons même une tendance à suivre notre modèle de façon quasi linéaire, par rapport à certains intervenants anglophones qui se permettent parfois quelques digressions. Les introductions des orateurs francophones sont souvent pour cette raison plus courtes que celles des orateurs anglophones [THOM 05a : 304].

En revanche, il existe des différences assez frappantes concernant la répartition des *moves* entre les deux groupes disciplinaires. La catégorie *Outline structure / Indicate scope* est présente dans les vingt communications en physique, alors qu'elle figure dans seulement deux des présentations en médecine. Cette différence s'explique sans doute par l'existence d'une structure de présentation très conventionnelle dans les communications en médecine. Les orateurs peuvent donc se dispenser de signaler cette structure par avance. On constate aussi que notre catégorie *Acknowledgements* qui figure dans la grande majorité de présentations en physique apparaît en seulement trois des présentations médicales.

Nous avons constaté aussi quelques légères différences entre les deux groupes culturels, notamment celles liées à la gestion des relations interpersonnelles. Comme il a été mentionné ci-dessus, la présence physique des auditeurs dans la salle incite l'orateur à instaurer un rapport direct avec eux. Les intervenants anglophones semblent en particulier bien exploiter l'étape que nous avons intitulé *Listener orientatio*n. Ils utilisent leur introduction pour s'adresser directement à l'auditoire, sous forme de plaisanteries, d'anecdotes ou de brèves observations. En voici un exemple :

> *(3)* Bonjour, mesdames et messieurs. *Very pleased to be here. …My work comes from Buffalo, whose name originally is not from the bison, which once populated our western plains, or so I'm told, but rather an evolution from the French "*buffle*".*
>
> (Conférencier anglophone en médecine)

En revanche, chez les orateurs francophones, les contacts directs avec les auditeurs sont plus réduits, se résumant souvent à une simple salutation. Dans leur introduction, les intervenants anglophones semblent aussi faire davantage allusion aux autres orateurs et au contexte même du colloque (*Conference context*) que les intervenants francophones. Ces différences ne sont certes pas très importantes sur le plan scientifique, mais elles contribuent à créer une ambiance ou un style un peu différent. Les interventions des orateurs anglophones peuvent paraître plus interactives et plus proches de la conversation et celles des orateurs francophones plus formelles et plus distantes.

Dans la section suivante, je propose de poursuivre la réflexion sur la question des rapports interpersonnels dans la communication de congrès. Comme on l'a vu ci-dessus, l'aspect persuasif du discours scientifique va de pair avec l'aspect interactionnel. Tout en promouvant sa propre recherche, le chercheur doit situer ses travaux au sein d'une somme de recherches déjà accomplies. Il doit rester attentif aux avis de ses lecteurs et prévenir leurs éventuelles critiques. Si ces efforts à prendre en compte les lecteurs restent souvent implicites dans les articles scientifiques (Fløttum 2005 ; Hyland 1999), on peut s'attendre à ce qu'ils soient plus explicites dans la communication de congrès. Les passages qui suivent s'attachent donc à décrire ceux de mes travaux qui explorent la manière dont les orateurs de communications scientifiques gèrent les relations interpersonnelles avec leur public.

8.2 Syntaxe et relations interpersonnelles dans la communication de congrès

Dans les travaux rapportés ici, je m'intéresse particulièrement à la manière dont les orateurs de communications scientifiques gèrent les relations interpersonnelles avec leur public à travers l'emploi qu'ils font des pronoms ainsi que de la syntaxe qu'ils ont choisie. Il s'agit ainsi d'une approche ascendante. En partant des traces laissées à la surface textuelle, l'objectif est d'évaluer comment ces marques peuvent nous renseigner sur les rapports interpersonnels entre les conférenciers et leur public et contribuer à la force persuasive du discours.

L'un des indices les plus évidents relatifs aux types de relations interpersonnelles est fourni par l'étude des pronoms personnels employés. Or, si dans le discours scientifique écrit les pronoms personnels sont traditionnellement assez peu présents, dans les communications de congrès la situation est tout autre. Dans une comparaison entre treize introductions de communications de colloque en physique et les treize articles correspondants publiés dans les Actes du colloque, nous avons trouvé 40% de pronoms personnels en tant que sujet grammatical dans les présentations orales, contre

seulement 5% dans les articles [Thom 04c ; Thom 05b]. Les conférenciers font une utilisation très stratégique des pronoms de première et de deuxième personne *I*, *you* et *we* dans la gestion de leurs relations avec les auditeurs. Ils n'hésitent pas à s'impliquer dans leurs textes, mais aussi à impliquer directement les auditeurs, comme l'emploi fréquent des pronoms de la deuxième personne et du *we* inclusif en témoigne. En contraste, non seulement l'emploi de pronoms personnels est rare dans les articles, mais nous ne trouvons, par exemple, aucune occurrence de *you*. L'auteur d'un article de recherche ne s'adresse que très rarement aux destinataires de façon directe.

Plus subtil, mais tout aussi stratégique, est l'emploi de certaines structures syntaxiques potentiellement interactives, dont par exemple l'impératif avec *let* et les phrases pseudo-clivées, qui permettent d'instaurer un dialogue avec le public. Il s'agit d'un type de syntaxe peu employé dans la communication scientifique écrite, mais qui, dans le contexte d'une conférence se déroulant en temps réel devant les auditeurs présents dans la salle, contribue à la force persuasive du discours. Si de tels moyens sont en théorie à la disposition de tout orateur, nous avons voulu savoir si leur emploi est aussi influencé par les variables culturelles et/ou disciplinaires [THOM 05a].

Le corpus analysé pour ces comparaisons, le même que celui décrit dans la section 8.1.3 *supra*, comporte quarante présentations orales : vingt communications en médecine, dix d'intervenants anglophones (EM) et dix d'intervenants d'origine francophone (FM), et vingt communications en physique, dix d'intervenants anglophones (EP), dix d'intervenants d'origine francophone (FP).

8.2.1 Relations interpersonnelles et choix de pronom

Comme je l'ai indiqué dans les sections précédentes (par exemple sections 7.2.1 et 8.1.2), les contraintes liées au traitement en temps réel de l'information, aussi bien par l'orateur que par les auditeurs, font que les groupes nominaux complexes ainsi que les propositions longues, caractéristiques de la

communication scientifique écrite, sont difficiles à produire ainsi qu'à assimiler dans ce contexte. De telles structures créent aussi plus de distance entre les interlocuteurs. La présence physique des auditeurs encourage en revanche à adopter une position plus personnelle. Tous les orateurs de notre corpus prennent globalement en compte ces contraintes contextuelles. Les propositions sont de manière générale simples, les nominalisations très réduites et les groupes nominaux thématiques consistent souvent en un seul pronom. On trouve en moyenne 46% de propositions avec un sujet thématique pronominal, dont 34% avec un pronom de première ou de deuxième personne [THOM 05a : 308]. Il est important de noter aussi qu'aucun des présentateurs de notre échantillon ne lit un texte rédigé. Comme c'est est souvent le cas dans les sciences de la vie et les sciences expérimentales, les présentateurs commentent librement leurs projections visuelles.

Toutefois, malgré ces similarités dans la gestion des contraintes cognitives et contextuelles, la distribution précise des pronoms *I*, *you* et *we* parmi les quatre groupes de chercheurs est assez différente. Une étude des fonctions et valeurs assurées par ces mêmes pronoms révèle aussi des différences frappantes d'utilisation, tant entre les groupes disciplinaires qu'entre les groupes linguistiques (cf. tableau 1.ci-dessous).

		EP	FP	EM	FM	Moyenne
Occurrences (taux pour 1000 mots)	I	8.6	4.7	11.4	6.8	7.2
	You	18.3	8	10	6.2	11.3
	We	19.6	25	25.3	17.5	18.4
Fonctions et valeurs (%)	*I*- organisationnel	72		40		
	I - évaluatif	14		34		
	You –personnel	27	55	45	85	
	You –générique	73	45	55	15	
	We – organisationnel	18.4		9.6		
	We – générique	10.5	15	11	20.7	

Tableau 1. Distribution de pronoms de 1re et 2e personne dans la communication de congrès [THOM 05a : 309].

EP= intervenants anglophones en physique, FP= intervenants francophones en physique, EM= intervenants anglophones en médecine, FM= intervenants francophones en médecine.

Si nous observons d'abord la fréquence globale de ces trois pronoms (le taux d'occurrence pour mille mots), on constate, par exemple que les orateurs francophones des deux disciplines, FP et FM, emploient deux fois moins le pronom *I* que leurs confrères anglophones. L'emploi du pronom *I* permet au chercheur de projeter une identité rhétorique forte en tant que chercheur et en tant qu'orateur. Son emploi fait souvent partie d'une stratégie métatextuelle. L'orateur signale aux auditeurs l'organisation de sa présentation dans le cadre du *« move »* que nous avons appelé *Outline structure/Indicate scope* :

> *(4) My talk today on… has three objectives. The first,* ***I'll*** *look at a simple axisymmetric Milo configuration,…* ***I'll*** *then go on to… And* ***I'll*** *also throughout the talk (EP).*

Dans d'autres cas, l'emploi du *I* par l'orateur peut permettre au chercheur d'émettre un avis personnel ou une évaluation sur la pertinence du projet de recherche présenté :

> *(5) I think this is probably due to… (EM).*
>
> *(6) And that is one point I hope you go away with (EP).*

L'utilisation moins importante du pronom *I* par les chercheurs francophones dans ces deux fonctions principales fait que leur engagement ou implication vis-à-vis de leurs propres travaux et vis-à-vis du déroulement de leur présentation apparaît moins explicite que celui des chercheurs anglophones. Ces résultats s'assimilent à ceux de I. Vassileva (2002) qui a également trouvé que les présentateurs d'origine bulgare étaient moins susceptibles d'employer le pronom *I* que leurs confrères anglophones.

Il existe aussi des différences assez marquées dans l'emploi du pronom *I* selon l'appartenance disciplinaire des chercheurs. Les chercheurs en physique (EP et FP) emploient le pronom essentiellement pour signaler l'organisation de leur discours (72% des occurrences), tandis que chez les chercheurs en médecine (EM et FM) cette proportion est beaucoup moins forte. Cette différence est évidemment liée à la présence quasi systématique de la catégorie de *Outline structure* dans l'organisation interne de la présentation en physique. En revanche, les orateurs en médecine emploient plus souvent le pronom *I* dans

sa fonction évaluative (souvent avec le verbe *think*), pour exprimer un avis personnel.

Le pronom *you* a deux valeurs principales dans la communication de congrès : une valeur personnelle renvoyant à l'auditoire présent dans la salle et une valeur générique plus impersonnelle renvoyant à la figure du chercheur en général. On remarque les mêmes différences entre les deux disciplines qu'avec l'emploi de *I*. Les conférenciers en médecine sont plus ouvertement interactifs que leurs confrères en physique, dans la mesure où ils emploient plus fréquemment le *you* personnel. Chez les chercheurs en physique on retrouve plus souvent le *you* générique :

> *(7) this could be quite problematic if you have etch rates variations and you combine that with non-uniformity (EP).*

Si nous comparons l'emploi de *you* par les deux groupes linguistiques, nous constatons de nouveau que *ou*, comme *I*, est beaucoup moins employé par les chercheurs francophones que par les anglophones, ce qui crée par conséquent moins d'interaction directe avec l'auditoire. Ce décalage vient surtout de la moindre utilisation du *you* générique par les deux groupes francophones : un décalage sans doute lié aux valeurs différentes de *vous* et de *you* dans les deux langues. Il semblerait que les chercheurs francophones soient peu informés de l'emploi du *you* impersonnel pour rendre le *On* du français.

Le pronom personnel *we* est plus employé que le pronom *I* par les quatre groupes et peut assumer une variété de valeurs. Tous les orateurs emploient le pronom, par exemple, dans son sens exclusif pour faire référence aux travaux entrepris avec leurs proches collaborateurs ou leur équipe, ainsi que dans son sens plus inclusif pour inclure le public présent dans la salle. Toutefois, l'emploi du pronom diffère chez les quatre groupes de locuteurs. À l'instar du pronom *I,* les présentateurs en physique emploient plus souvent le *we* dans sa fonction organisationnelle métatextuelle. Il existe aussi de légères différences d'emploi entre les deux groupes linguistiques. Une plus grande utilisation de *we* par les francophones permet en partie de compenser leur faible emploi du pronom *I*.

On constate surtout un plus grand recours au *we* dans son sens générique (la communauté de chercheurs), à la place du pronom *you* pour les raisons expliquées ci-dessus.

L'examen de la distribution des pronoms personnels permet ainsi de constater de subtiles différences dans la manière dont les relations interpersonnelles sont gérées par les quatre groupes de conférenciers. On constate plus d'engagement direct chez les conférenciers en médecine (à travers par exemple l'emploi du *I* évaluatif), tandis que chez les conférenciers en physique l'emploi de pronoms est surtout lié à des préoccupations organisationnelles. Dans les deux groupes disciplinaires, le plus grand recours global aux pronoms personnels par les conférenciers anglophones, et surtout le plus grand emploi de *I* et *you*, fait qu'ils paraissent plus proches de leur auditoire que les conférenciers d'origine francophone. La culture linguistique d'origine et les conventions disciplinaires semblent toutes les deux influencer la manière dont les orateurs créent leur propre identité discursive et interagissent avec leur auditoire. Ces influences se font aussi sentir dans le comportement syntaxique des quatre groupes.

8.2.2. Relations interpersonnelles et syntaxe

Le type de syntaxe employé par les orateurs dans les communications de congrès est en général plus « conversationnel » que celui employé dans la communication scientifique écrite [THOM 01]. On s'attend à trouver, par exemple, beaucoup moins de verbes à forme passive, de tournures impersonnelles ou de groupes nominaux complexes et, en revanche, une syntaxe plus orale avec des phrases courtes, des verbes à forme active et des structures et formulations plus « interactives ». À l'instar de l'emploi de pronoms personnels, l'orateur a la possibilité, à travers ses choix de mode et de syntaxe, d'instaurer une structuration de discours plus interactive (Camiciottoli 2004). Dans cette étude visant à évaluer l'influence de la discipline et de la culture linguistique sur la gestion des rapports interpersonnels [THOM 05a], nous nous sommes penchées sur quatre structures qui se prêtent bien à un emploi

interactif et sur deux structures plus impersonnelles. Le tableau 2 permet de visualiser les résultats de cette comparaison suivant chacun des quatre groupes de conférenciers.

	EP	FP	EM	FM	Total
Structures syntaxiques interactives					
Questions rhétoriques	19	12	16	12	59
Phrases conditionnelles[7]	22	4	23	2	51
Impératif avec *Let*	20	3	3	0	26
Phrases pseudo-clivées	68	4	19	4	95
Total	129	23	61	18	231
Syntaxe impersonnelle					
Extraposition	17	23	11	41	92
% phrases passives	6%	13%	8%	13,5%	

Tableau 2. Choix syntaxiques dans la communication de congrès : comparaison entre quatre groupes de conférenciers [THOM 05a : 309].

Les questions rhétoriques peuvent être considérées comme un indice d'une interactivité très explicite dans la mesure où leur emploi permet d'établir un pseudo-dialogue avec les auditeurs. Les conférenciers les emploient souvent, par exemple, pour attirer l'attention des auditeurs sur une transition vers un nouveau thème ou sur un nouveau *move* rhétorique :

> *(8) Now, what are the results, the overall results? For the clinical response, we have complete response (FM).*

Toutefois, si les questions rhétoriques sont couramment employées par tous les conférenciers dans leurs communications orales sans distinction[8], ce n'est pas le cas pour les trois autres structures « interactives » examinées. On peut tout d'abord constater un recours moins systématique à l'impératif avec *let* et aux phrases pseudo-clivées par les conférenciers en médecine que par les conférenciers en physique. Le format très standardisé des présentations en médecine fait que les conférenciers peuvent souvent se dispenser de donner des consignes métadiscursives aux auditeurs et qu'ils ont donc moins besoin de ces deux structures qui sont en grande partie dédiées à ces fonctions métadiscursives. Toutefois, les différences les plus frappantes sont entre les

[7] Sont seules incluses les phrases conditionnelles jouant un rôle dans la structuration du discours. Pour une analyse complète des fonctions jouées par les phrases conditionnelles dans le discours scientifique (écrit et oral), voir chapitre 9 *infra.*

[8] On trouve un taux d'occurrence de 7,5 questions rhétoriques pour 1000 mots dans chacun des quatre groupes de locuteurs [THOM 05a : 311].

deux groupes linguistiques. Les conférenciers francophones ont globalement très peu recours aux phrases pseudo-clivées, à l'impératif avec *let* ou aux fonctions pragmatiques des phrases conditionnelles.

Les phrases conditionnelles peuvent être exploitées à des fins interactives et organisationnelles. Leur emploi peut permettre, par exemple, de prévenir les auditeurs d'un changement de direction :

(9) Okay, if I can now turn to particle beam disposition (EP).

Dans une deuxième utilisation pragmatique, elles sont employées comme des injonctions atténuées. L'orateur invite les auditeurs à diriger leur attention vers une image ou un schéma qu'il est sur le point de projeter.

(10) Now if we look at the average ion flux for the main etch step, what's particularly striking here is that… (EP).

Toutefois, les conférenciers francophones de notre échantillon exploitent rarement les fonctions pragmatiques des conditionnelles. Ils emploient bien les phrases conditionnelles dans leur sens logique premier (cf. chapitre 9 *infra*), mais ils semblent peu familiers avec cet emploi potentiellement interactif de la structure.

Concernant l'emploi du mode impératif, on trouve très peu d'injonctions directes dans notre corpus de présentations scientifiques. Les conférenciers estiment sans doute qu'il serait peu approprié de prononcer des commandes directes à l'intention des confrères. En revanche, on relève un certain nombre d'exemples de l'impératif avec *let* (*let me* ou *let's),* presque tous dans les présentations des conférenciers anglophones :

(11) Let's have a look… Let me show you a schematic of… Let me summarize…

De telles expressions permettent aux orateurs d'attirer l'attention des auditeurs sur une information à venir d'une façon polie et courtoise. Comme on ne trouve presque aucun exemple de ce type chez les conférenciers d'origine francophone, il serait légitime de se demander ce qu'ils utilisent à la place comme « accroches » pour mobiliser l'attention des auditeurs. Une analyse détaillée de leurs communications permet de mettre en évidence l'absence de

structures « annonces » de ce type. Leur entrée en matière semble plus directe : *This is a diagram of (…), here is (…).*

D'une façon analogue, les phrases pseudo-clivées sont aussi beaucoup moins usitées par les orateurs d'origine francophone que par leurs confrères anglophones (seulement huit occurrences pour les francophones contre quatre-vingt-sept pour les anglophones). Comme il a été indiqué dans le chapitre 7, les phrases pseudo-clivées remplissent plusieurs fonctions utiles dans un contexte de communications de congrès. Une de ces fonctions concerne la gestion des relations interpersonnelles. En employant une pseudo-clivée, le locuteur répond en quelque sorte, par le focus, à une question implicite que son interlocuteur serait en train de se poser.

> *(12) Now what we found is that these little dots here indicate a boundary (EP).*
>
> *(13) I think what this really means is that they have a high chance of recurring in the loco-regional area (EM).*

Dans toutes les phrases pseudo-clivées il est possible d'identifier une question sous-jacente que les auditeurs seraient susceptibles de poser : « Qu'avez-vous trouvé ? » « Qu'est-ce que ceci veut vraiment dire ? » Le locuteur fait ainsi intervenir un rapport dialogique dans son monologue. La quasi-absence ainsi de cette structure « interactive » chez les conférenciers francophones, tout comme celles des phrases conditionnelles pragmatiques et de l'impératif avec *Let,* renforce l'impression d'un style beaucoup moins conversationnel et plus formel que celui employé dans les communications des chercheurs anglophones.

On peut également constater le niveau moins élevé d'interactivité dans les présentations des chercheurs francophones, en comparant l'emploi fait de la syntaxe impersonnelle par les deux groupes linguistiques. Si le recours à ces deux structures est sensiblement moins important que dans le discours scientifique écrit, on constate néanmoins que les chercheurs francophones emploient plus de passifs et de formes extraposées que les chercheurs anglophones (voir également chapitre 7 *supra*). Globalement, on voit que la transition d'une syntaxe écrite vers une syntaxe interactive est beaucoup moins nette dans les deux groupes francophones. Dans l'article de recherche,

l'extraposition est employée, par exemple, comme stratégie de modalisation (cf. Chapitre 7 et [THOM 04d]). Comme l'exemple suivant le démontre, les chercheurs francophones ont tendance à employer la structure de la même façon dans leurs présentations orales :

> *(14) It is important to note that this method gives one value of the electronic temperature for each argon line used (FP).*

Or dans une communication de congrès se déroulant en temps réel devant un public, les orateurs anglophones préférèrent adopter en général un engagement beaucoup plus direct : *« We know; I believe ».* L'influence de l'écrit dans les communications des orateurs francophones peut aussi être constatée dans la quasi-absence de contractions (*« we are »* au lieu de *« we're »* et *« I am »* au lieu de *« I'm »* etc.). La seule forme contractée employée de façon courante est celle du pronom impersonnel *it's,* ce qui conduit aussi à l'impression d'un style moins interactif.

Les analyses présentées dans les sections précédentes démontrent ainsi le dosage assez complexe des influences disciplinaires et linguistiques sur les stratégies interpersonnelles des conférenciers. Si, par exemple, les conférenciers en médecine n'hésitent pas à intervenir directement au niveau du contenu pour émettre un jugement ou une évaluation, les conférenciers en physique interviennent davantage au niveau métadiscursif. Ils adoptent globalement une gestion du discours plus explicite et plus « directive » que celle de leurs confrères en médecine. Toutefois, ce sont les différences entre les deux groupes linguistiques qui sont les plus patentes. Les communications des conférenciers francophones dans les deux disciplines apparaissent moins conversationnelles et plus formelles que celles des conférenciers anglophones À travers leur emploi accru des pronoms personnels et d'une syntaxe plus interactive, ces derniers interagissent plus directement avec leurs auditoire. En revanche, les conférenciers francophones donnent l'impression de vouloir rester plus en retrait. Ces résultats permettent d'étayer ce qui a été souligné dans la section précédente concernant la répartition des *moves* entre les deux

groupes de locuteurs où nous avons relevé un recours plus réduit à la catégorie *Listener orientation* chez les chercheurs d'origine francophone.

Il est possible d'avancer plusieurs hypothèses pour expliquer ces différences dans la gestion des relations interpersonnelles par les deux groupes linguistiques. Une hypothèse serait d'attribuer ces différences au style d'argumentation inhérent à chacune des deux langues d'origine. Certains commentateurs, par exemple J. Hinds (1987) ou M. Clyne (1998), considèrent que l'anglais est une langue beaucoup plus centrée sur le rédacteur que sur le lecteur (*« writer-responsible »*). C'est la responsabilité du rédacteur ou de l'émetteur du message de rendre explicite son argumentation et de guider au maximum les destinataires du texte, à la différence de beaucoup d'autres cultures linguistiques (*« reader-responsible »*), dont la culture française, où la responsabilité incombe davantage au lecteur/récepteur[9]. Afin de vérifier cette hypothèse il faudrait aussi comparer le comportement linguistique des conférenciers francophones dans les présentations scientifiques données en français.

Une autre hypothèse serait de lier la moindre interactivité observée dans les présentations des conférenciers francophones à l'emprise de la syntaxe écrite sur leurs stratégies interpersonnelles. La transition d'une syntaxe écrite vers une syntaxe plus orale et interactive est globalement beaucoup moins nette dans les présentations orales des chercheurs francophones. Sans apporter un jugement de valeur sur ces deux styles de présentation, il semble néanmoins important de relever ces différences et de montrer aux étudiants et/ou chercheurs novices les éventuelles répercussions des choix syntaxiques sur le type de relation instauré avec le public. Il est certain aussi que le style de présentations très informel adopté parfois par les conférenciers anglo-américains serait peu approprié dans certaines cultures européennes. En revanche, on peut aussi se demander si le style parfois très impersonnel des

[9] Dans une comparaison sur deux jugements juridiques en danois et en français, L. Lundquist (1998) conclut que le texte juridique danois est plus facilement accessible que le texte français qui exige du lecteur un plus grand effort .

orateurs francophones dans leurs présentations en anglais ne s'accompagne pas d'une perte d'efficacité rhétorique.

Dans mes propres cours à l'Institut Télécom, j'essaie de montrer, exemples à l'appui, tous ces aspects organisationnels et interactionnels des communications de congrès. J'essaie aussi de discuter avec les étudiants (doctorants pour la plupart) des éventuelles différences liées aux traditions disciplinaires. Les manuels proposant un « mode d'emploi » pour les présentations scientifiques disponibles dans le commerce sont souvent peu détaillés et ne prennent pas vraiment en compte les réalités du terrain. Comme les recherches rapportées dans les sections précédentes l'ont démontré, les communications de congrès possèdent leurs propres stratégies rhétoriques et organisationnelles. Les structures syntaxiques dont risquent d'avoir besoin les orateurs pour mener à bien ces stratégies dépendent de ce contexte très spécifique et il me semble important de familiariser les étudiants avec ces finalités rhétoriques, ainsi qu'avec la syntaxe nécessaires à la réalisation de ces objectifs.

Chapitre 9 – Une syntaxe au service des genres : le cas des phrases conditionnelles

Résumé

Dans ce chapitre j'analyse le fonctionnement des phrases conditionnelles dans trois genres scientifiques différents : les articles de recherche, les communications de congrès et les éditoriaux spécialisés. À travers une analyse fine de leur fonctionnement en contexte, je montre l'impact du genre sur l'activation de cette réponse syntaxique, mais également le rôle structurant joué par cette structure dans l'élaboration d'un modèle interprétatif par le récepteur.

Dans le premier volet (9.1.), j'identifie certaines spécificités au niveau de la forme et du fonctionnement des phrases conditionnelles dans les articles de recherche médicale [THOM 07]. Je passe ensuite à une analyse contrastive de l'emploi du *si* français et du *if* anglais dans des articles de recherche médicale, en comparant les fonctions rhétoriques remplies par ces constructions, selon les formes verbales et le positionnement de la proposition en *si/if* (initial, médian ou final) [THOM 07].

Le deuxième volet (9.2) du chapitre porte sur l'emploi des conditionnelles dans les communications de congrès. Une comparaison avec les articles de recherche permettant d'identifier les spécificités de l'emploi de cette structure dans ce genre oral est suivie par une comparaison entre l'utilisation faite de cette structure par des locuteurs anglophones et des locuteurs non anglophones [THOM 08b].

Dans un troisième volet (9.3), je décris l'emploi de cette construction dans un troisième genre spécialisé : l'éditorial médical. Après une comparaison avec l'article de recherche [THOM 08a], j'examine l'emploi fait par les éditorialistes de quatre variantes formelles de la structure *[If P, Q],* à savoir : *[If P, then Q], [Q only if P], [if P, Q?], [even if P, Q]* (ou *[Q even if P]* [THOM 09b] et [THOM 08a].

La dernière partie du chapitre (9.4) comporte une réflexion sur le décalage entre la présentation des phrases conditionnelles faite dans la littérature linguistique et dans les grammaires pédagogiques avec les usages auxquels les apprenants risquent d'être confrontés dans leur vie professionnelle [THOM 08a].

Chapitre 9. Une syntaxe au service des genres : le cas des phrases conditionnelles

L'idée qu'une langue spécialisée est un sous-système de la langue générale est depuis longtemps dépassée (Lerat 1997 ; Resche 2001). Il n'y a pas une grammaire de spécialité distincte de celle de la langue générale. En revanche, il peut exister une utilisation spécifique de la syntaxe. Les différentes communautés spécialisées empruntent au fonds commun de la langue les structures syntaxiques les mieux adaptées à leurs besoins épistémologiques et/ou communicatifs. Cette utilisation spécifique de la syntaxe peut se révéler par trois aspects : fréquentiel, fonctionnel et formel [THOM 08b].

Les informations sur la fréquence d'une structure sont depuis longtemps abordées dans le domaine de l'ESP (cf. Barber 1962). Dans différents discours spécialisés, on trouve parfois une utilisation massive de certaines structures. On cite souvent l'exemple du passif, dont l'emploi est beaucoup plus important dans les articles de recherche scientifiques que dans l'usage courant. Les fonctions remplies par la structure syntaxique en question peuvent aussi varier selon le discours spécialisé concerné. Beaucoup de constructions syntaxiques ont des fonctions multiples. Les phrases conditionnelles peuvent, pas exemple, remplir toute une gamme de fonctions. Il est donc intéressant d'identifier les fonctions qui sont les plus pertinentes dans le contexte du discours étudié. Finalement, il semble important d'examiner les variations formelles que l'on peut observer dans l'emploi d'une construction. Les phrases conditionnelles possèdent plusieurs variantes qui permettent d'exprimer une gamme de nuances sémantiques. Dans les travaux présentés ici, j'essaie aussi d'identifier les caractéristiques formelles des phrases conditionnelles les plus répandues dans les discours spécialisés examinés.

L'hypothèse que les textes varient linguistiquement en fonction, non seulement des domaines de connaissance auxquels ils ont trait, mais aussi du genre auquel ils appartiennent, est depuis longtemps explorée dans la recherche en

langue de spécialité (cf. chapitre 6). Certaines particularités dans l'utilisation spécifique de la syntaxe peuvent être attribuées à l'impact du genre et d'autres sont plus liées au domaine disciplinaire. Afin de limiter l'impact de cette deuxième variable, les travaux présentés ici portent tous sur le discours médical et plus précisément sur l'oncologie.

9.1 Emploi des phrases conditionnelles dans les articles scientifiques en médecine

Les phrases conditionnelles jouent un rôle important et difficilement remplaçable, dans l'activité scientifique. Les constructions hypothétiques permettent, par exemple, aux chercheurs de placer leur argumentation dans un monde fictif (Adam 1998). Cet espace fictif est très utile au chercheur médical, en lui fournissant la distance nécessaire pour l'expression de ses hypothèses de recherche. En recherche médicale et en particulier dans les enquêtes cliniques du type étudié dans ce corpus où le raisonnement est inductif et non pas déductif, il est très important d'évaluer l'évidence afin de créer un espace pour toute assertion éventuelle. Les hypothèses proposées doivent être nécessairement émises avec précaution (*« hedging »*), à cause d'un grand nombre de facteurs aléatoires ou de variables (Salager-Meyer 1994).

> (1) **Si** une telle analyse était réalisée de manière systématique, la valeur prédictive des emboles pour déterminer l'envahissement ganglionnaire augmenterait probablement de manière importante.

La certitude absolue n'existe pas et l'emploi de phrases conditionnelles permet aux chercheurs de garder la distance nécessaire avec ces assertions. Cependant, en même temps, les résultats obtenus dans la recherche médicale sont soumis à des conditions strictes (des tests cliniques, des protocoles expérimentaux). Les chercheurs doivent préciser les conditions dans lesquelles la recherche est entreprise. Les constructions conditionnelles peuvent ici jouer une fonction importante en précisant, par exemple, les critères d'éligibilité pour les malades impliqués dans les différents tests ou en expliquant les raisons pour des différentes décisions thérapeutiques adoptées :

> (2) L'association comporte trois à quatre cures de BEP, suivies d'une irradiation de 40 Gy du site tumoral **si** la maladie est localisée ou une irradiation de 40 Gy du site tumoral et de 25 Gy de l'axe craniospinal **si** la maladie est diffuse.

L'importance des phrases conditionnelles dans le monde de la recherche, est également liée à la fonction potentiellement polyphonique de ces constructions. L'article de recherche scientifique a souvent été qualifié de texte à la fois concurrentiel et coopératif (cf. Fløttum 2005 et chapitre 8 *supra*). Le chercheur (et non pas seulement le chercheur en médecine) doit garder un esprit de compétition afin de créer sa niche et de promouvoir ses propres idées. En même temps, il doit démontrer un certain esprit de coopération : en reconnaissant, par exemple, l'importance des travaux d'autres chercheurs, afin de délimiter la pertinence de ses propres travaux dans ce travail d'ensemble. Il peut faire rentrer différentes voix dans son texte – tout en négociant ses propres hypothèses de recherche – en s'appuyant, par exemple, sur des connaissances partagées ou des suppositions communément admises (*Si l'on admet X*), ou bien sur des allusions aux travaux de ses pairs (*Si comme les études de X démontrent*). On prétend souvent (cf. Dahl 2004a) que les auteurs d'articles de recherche en médecine sont un peu « absents » en tant qu'auteurs (ils sont censés relater les faits et ne pas s'impliquer dans leurs textes). Bien que la potentielle polyphonique des phrases conditionnelles soit plus marquée dans les deux autres genres examinés (cf. ci-dessous), mes analyses suggèrent que même dans l'article de recherche, le chercheur médical est assez présent dans son texte et parfois même polémique – même si l'on peut également constater une certaine différence de comportement entres les chercheurs anglophones et francophones [THOM 07].

9.1.1 Corpus et méthodologie

L'étude contrastive entreprise des phrases conditionnelles dans l'article de recherche [THOM 07] était fondée sur l'analyse de quatre-vingt-dix articles de recherche (337 619 mots extraits de quatre revues spécialisées (deux en anglais ; deux en français) : *Bulletin du Cancer* (quinze articles), *Cancer-Radiothérapie* (quinze articles), *International Journal of Radiation Oncology* (trente articles) et *Journal of Clinical Oncology* (trente articles). Ces quatre revues étaient conseillées par un spécialiste de la discipline comme des revues sérieuses (avec comité de lecture), jouissant d'une notoriété similaire dans le

domaine de l'oncologie[1]. Comme il a été souligné dans le chapitre 5, afin de créer un corpus spécialisé homogène, il est souvent nécessaire de recourir aux spécialistes du domaine. Dans ce cas précis, l'opinion informée d'un spécialiste francophone était particulièrement importante pour la constitution de ce corpus comparable anglais/français. Dans les deux revues anglophones, j'ai également distingué deux sous-groupes supplémentaires, les auteurs anglo-saxons (En) et les auteurs francophones (Fse), afin d'évaluer l'influence éventuelle de la langue d'origine qui pourrait résulter de la comparaison anglais/français.

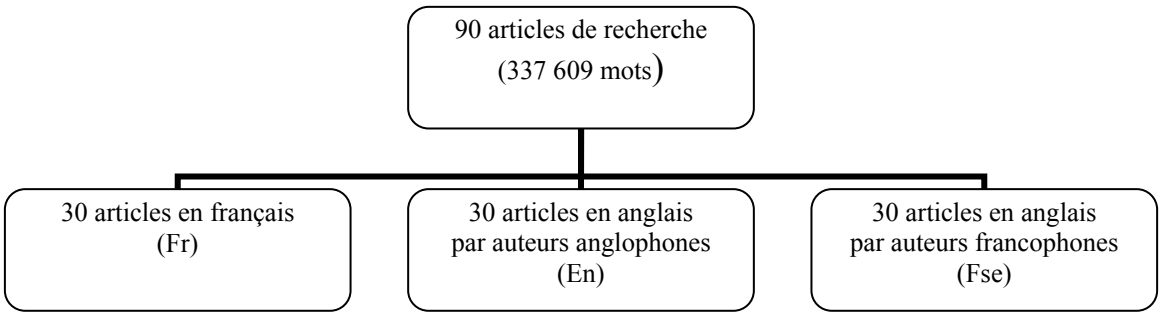


Figure 1 – Corpus contrastif d'articles de recherche en médecine [THOM 07].

L'identification des phrases conditionnelles en elle-même peut s'avérer problématique à cause de la très grande variété des formes employées pour exprimer la notion de conditionnalité et aussi du problème plus général de la définition d'un sens « conditionnel ». Dans cette étude j'ai adopté une approche purement ascendante. Seules ont été relevées les occurrences contenant les mots *si* ou *if,* que j'ai considérés comme les représentants lexicaux typiques de conditionnalité en français et en anglais avec un chevauchement assez large dans leurs fonctions prototypiques. Les autres subordonnants dans les deux langues tels que *à condition que, à supposer que, on condition that, unless,* ou les différents cas parataxiques sans marqueur lexical (*« Should you change your mind, let us know »*) n'ont pas été pris en considération. J'ai ensuite éliminé manuellement toutes les occurrences dans les deux langues d'un *si* ou d'un *if* percontatif ou d'interrogation indirecte, ainsi que les exemples avec *as if* ou *comme si*. En revanche, j'ai pris en compte tous les emplois elliptiques de *si*

[1] L'oncologie a été choisie comme spécialité médicale afin de permettre des comparaisons avec un autre corpus déjà constitué de communications de colloque dans le même domaine (cf. 9.2 *supra*).

ou *if (si possible, si nécessaire),* des empois souvent écartés dans d'autres approches (par exemple chez C.E. Ford et S.A. Thompson 1986) mais qui se sont révélés relativement courants dans ce corpus spécialisé.

9.1.2 Fréquences, formes et fonctions

D'un point de vue purement quantitatif, les subordonnées en *si* et en *if* sont employées avec une fréquence très semblable chez les trois groupes d'auteurs, avec en moyenne une occurrence pour mille mots.

Origine des auteurs	Mots	Occurrences de *si/if*	Taux pour 1000 mots
Tous (90 articles)	337 689	320	0,94
Auteurs – Fr (30 articles)	111 391	103	0,92
Auteurs – En (30 articles	111 907	119	1,06
Auteurs – Fse (30 articles)	114 391	98	0,85

Tableau 1. Fréquence des phrases conditionnelles dans les articles scientifiques
Fr = Articles en français ; En = Articles en anglais par des auteurs anglophones ;
Fse = Articles en anglais par des auteurs francophones [THOM 07 : 153].

Le positionnement de la subordonnée au sein de la phrase est également très similaire dans les trois groupes. Toutes les occurrences des phrases conditionnelles ont été répertoriées selon trois positions : initiale, médiane et finale. Chez les trois groupes d'auteurs, c'est le postpositionnement de la proposition subordonnée (la proposition P) qui est le plus répandu dans plus de 50% des cas : Fr 52% ; En 55% ; Fse 52%.

> *(3) Patients were eligible for participation **if** they were 18 years of age or older.*

Cette tendance vers une proposition P finale semble être une vraie spécificité du genre. Dans la littérature, que ce soit en anglais ou en français, la position initiale de la proposition P est souvent présentée comme le cas par défaut. B. Comrie suit par exemple Greenberg (1963) en considérant l'ordre protase/apodose comme un trait universel :

> *In conditional statements, the conditional clause (protase) precedes the conclusion (apodose) as the normal order in all languages (1986 : 83)*

Dans l'étude de Ford et Thompson (1986), les propositions P finales représentaient seulement 23% des occurrences à l'écrit (et 18% à l'oral). Selon ces auteurs, les cas assez rares de postpositionnement de la proposition subordonnée sont liés aux considérations de lourdeur informationnelle ou de

complexité syntaxique. Or dans ce corpus spécialisé d'articles médicaux, il semble que le post-positionnement est aussi très lié aux fonctions rhétoriques assurées par la phrase conditionnelle et à la position elle-même de la phrase conditionnelle au sein de l'article de recherche. Un très grand nombre de propositions conditionnelles postposées se trouve dans la partie « Méthodes » de l'article de recherche. Cette section contient quatre-ving-cinq conditionnelles postposées pour seulement quarante-deux propositions antéposées.

Or dans la section « Méthodes » de l'article de recherche, nous relevons surtout des phrases conditionnelles de type « factuel » (cf. *infra*). Les auteurs décrivent les conditions dans lesquelles une expérience se déroule ou ils précisent les critères qui motivent le choix d'une thérapie :

> (4) A 7 h 30, elle bénéficiait d'une injection radio-isotopique (Nanocis 30 MBq) en péritumoral **si** la tumeur était palpable ou en subaréolaire **si** la tumeur était très externe ou infraclinique.

Dans la plupart de ces exemples de type factuel (décisions méthodologiques et thérapeutiques), la proposition subordonnée est postposée et favorise ainsi une lecture intraprédicative. La décision de traitement est présentée avant la définition opérationnelle détaillant les critères applicables pour la décision. L'étendue de ces subordonnées finales semble se limiter à la proposition principale antéposée, et ne fournit pas un point de départ en aval, comme ça peut être souvent le cas pour les propositions subordonnées à l'initial (Charolles 2003).

Une autre spécificité d'ordre formel concerne les formes verbales de la principale et de la subordonnée. Chez les trois groupes d'auteurs, les trois séquences conditionnelles traditionnelles telles que celles présentées dans les grammaires pédagogiques dans les deux langues (si+présent+futur ; si+passé+conditionnel, si+plus que parfait + conditionnel passé) étaient peu utilisées. Ces combinaisons représentaient moins de 6% d'occurrences dans les trois groupes. En revanche, les chercheurs ont eu recours à une très grande variété d'autres séquences verbales[2].

[2] Pour les détails et les analyses des séquences verbales dans les articles en anglais, voir [THOM 08b : 48-49].

Les fonctions jouées par les phrases conditionnelles dans l'article de recherche présentaient aussi un profil très particulier. Les classifications de phrases conditionnelles déjà existantes (comme celles de E. Sweester – 1990) – et A. Athanasiadou et Dirven –1997– se sont révélées très difficiles à appliquer à ce corpus spécialisé (cf. *infra*). Pour cette raison nous avons décidé de créer notre propre classification, en nous fondant sur les régularités relevées dans le corpus. Nous avons adopté une approche de genre, identifiant ainsi les fonctions jouées par les constructions par rapport aux objectifs rhétoriques de ce discours très particulier. Trois grandes catégories de fonctions ont été établies : fonctions factuelles, fonctions de réglage (« refocusing »), fonctions discursives.

La catégorie de fonctions factuelles correspond en grande partie aux fonctions appelées habituelles ou génériques (ou *« course of event »*) dans d'autres approches (cf. Athanasiadou et Dirven 1997). Cependant, le terme *factuel* semble particulièrement pertinent dans l'analyse de ce discours scientifique où il est très important de stipuler les conditions dans lesquelles les différents faits décrits peuvent être considérés comme légitimes. La relation P-Q est établie dans le monde « réel » et le *si* ou *if* est très poche d'un *when* ou d'un *quand,* comme dans l'exemple (4) ci-dessus. Néanmoins, la valeur suspensive ou « non assertée » de la proposition subordonnée crée un certain espace libre où le lecteur a la possibilité de remettre en cause les décisions prises par les auteurs (Achard-Bayle 2005). Trois sous-catégories de conditionnelles factuelles ont été distinguées : décisions méthodologiques, décisions de traitement et corrélations. Comme nous venons de le voir, les phrases conditionnelles factuelles se trouvent souvent dans la section « Méthodologie » de l'article de recherche.

La deuxième grande catégorie de fonctions regroupe les fonctions de réglage ou de *« refocusing »*. Afin de promouvoir leurs arguments ou de défendre leurs thèses les auteurs d'article de recherche peuvent avoir besoin de redessiner ou de renégocier l'espace argumentatif. Les phrases conditionnelles de ce type

fournissent un certain potentiel de manœuvre. Sont inclus dans cette catégorie les recommandations, les prédictions et hypothèses, les emplois concessifs, les commentaires et les emplois adversatifs ou contrastifs. Je propose d'illustrer cette catégorie avec l'exemple d'une conditionnelle de recommandation et avec une conditionnelle de type hypothétique[3].

Les phrases conditionnelles permettent au chercheur de qualifier ou d'atténuer des recommandations :

> *(5) Il est donc primordial, si l'on veut se dispenser de curage axillaire systématique comme cela est proposé par différentes équipes [23, 24, 37, 48], d'utiliser la technique la plus fiable possible.*

Dans le contexte d'un article de recherche où les conseils trop directs envers les pairs pourraient être considérés présomptueux, la proposition conditionnelle permet d'atténuer la force de ces conseils ou recommandations. L'exemple (5) fournit également un bon exemple de l'interaction polyphonique, effectuée à l'aide de l'opérateur *si*. Les avantages procurés par cette procédure médicale sans curage, sont attribués à un pronom « on » avec une portée très large et indéfinie, comprenant à la fois l'auteur lui-même, les autres équipes médicales mentionnées dans les références entre crochets, et aussi peut-être les lecteurs. De cette manière l'auteur trouve de bonnes justifications pour la recommandation plutôt directe de la proposition Q : « Si ce que nous (et d'autres) disons s'avère, voici ce qu'on peut conseiller en conséquence. »

Les hypothèses dans l'article de recherche médical sont exprimées d'une façon très prudente. Dans ce corpus il n'y avait qu'un seul exemple d'une prédiction classique (si + présent + futur) exprimant ainsi une hypothèse probable. Dans tous les autres cas il s'agissait d'hypothèses potentiellement réalisables, où l'auteur cherche à se distancier par rapport à ses prédictions au moyen de verbes modaux, de formes verbales du passé ou d'expressions lexicales d'incertitude.

> (6) Si une telle analyse <u>était réalisée</u> de manière systématique, la valeur prédictive des emboles pour déterminer l'envahissement ganglionnaire <u>augmenterait probablement</u> de manière importante.

[3] Les autres types de conditionnelles de réglage (« *refocusing* ») sont illustrés en [THOM 07].

Il existe également des cas où l'hypothèse est attribuée à d'autres auteurs :

> (7) Si l'on en croit la Société de pneumologie de langue française, le Cyfra 21-1 est actuellement le seul marqueur tumoral dont la valeur pronostique est clairement démontrée dans le cancer bronchopulmonaire primitif non à petites cellules [4]. Nous avons néanmoins pu constater (…).

Le recours à l'hypothèse d'un autre est ici un stratagème rhétorique. Comme l'explique G. Achard-Bayle (2005), l'emploi de la conditionnelle peut servir à suspendre la valeur de vérité. Dans cet exemple, l'auteur invoque les théories des autres, à l'aide de l'opérateur *si,* afin de les contester et de promouvoir ses propres idées.

Dans la troisième grande catégorie de fonction dégagée, celle des fonctions discursives, il s'agit d'orienter l'attention du lecteur d'un sujet à un autre ou d'un espace textuel à un autre, passant par exemple du texte au schéma :

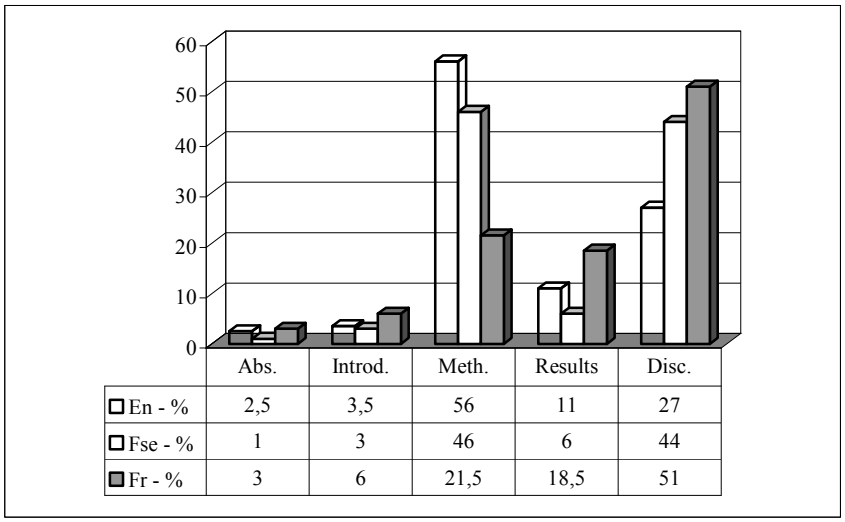
> (8) Si nous étudions plus précisément la nature des 39 discordances, nous notons qu'elles se répartissent en deux types : les sous-évaluations et les problèmes d'échantillonnage.

Dans l'exemple (8) l'auteur attire l'attention du lecteur sur un schéma dans l'article. Il ne s'agit pas vraiment de signaler une relation particulière entre P et Q, mais de signaler un cadre provisoire (espace sémantique) fourni par la proposition subordonnée dans le contexte où il est pertinent d'énoncer Q. Toutefois, cette catégorie est peu représentée dans les articles de recherche, Comme les articles suivent un format très rigide avec de nombreux sous-titres et sous-sections, les indications metadiscursives sont souvent redondantes. En revanche, on relève nombreux exemples de ce type dans les communications de congrès (cf. *infra*).

Globalement, dans les articles de recherche en médecine, ce sont les fonctions factuelles des phrases conditionnelles qui sont les plus employées et les fonctions discursives qui sont de très loin les moins sollicitées. Cependant, il existe des différences assez frappantes dans l'emploi de conditionnelles factuelles et de conditionnelles du type *« refocusing »* chez les trois groupes d'auteurs. Ces différences sont résumées dans la partie suivante.

9.1.3 Comparaison français/anglais

Comme on l'a déjà constaté, sur le plan quantitatif global la fréquence des phrases conditionnelles est assez semblable dans les articles anglais et français. Cependant une comparaison de la répartition de ces mêmes occurrences dans les différentes sections de l'article de recherche fait apparaître des différences marquées.



	Abs.	Introd.	Meth.	Results	Disc.
En - %	2,5	3,5	56	11	27
Fse - %	1	3	46	6	44
Fr - %	3	6	21,5	18,5	51

Figure 2. Distribution d'occurrences selon la section de l'article chez les trois groupes d'auteurs [THOM 07 : 156].

Tous les articles de recherche de ce corpus possèdent des sections bien distinctes et suivent le schéma IMRED. Comme il a déjà été souligné (cf. Swales 1990 ; Salager-Meyer 1994 et chapitre 5 *supra*), les différentes sections de l'article de recherche remplissent des fonctions rhétoriques spécifiques. Il est habituel d'établir une frontière assez nette entre les sections plus descriptives de l'article et les sections plus analytiques. Les sections « Méthodes » et « Résultats » sont essentiellement descriptives et informatives. En revanche, les sections « Introduction » et « Discussion » sont plus analytiques. Les auteurs fournissent une justification de leurs projets de recherche ou proposent des interprétations des résultats obtenus. Dans les articles en anglais des auteurs anglophones (En), la plus grande proportion des phrases conditionnelles se trouvent dans les parties plus descriptives de l'article, surtout dans la partie « Méthodes » (56% des occurrences). En revanche, dans les articles en français (Fr), la plus grande proportion des occurrences se trouvent

dans la partie « Discussion » (51% des occurrences). Pour le troisième groupe d'auteurs (les articles anglais rédigés par des auteurs francophones), la distribution des occurrences se situe entres deux extrêmes.

Ces résultats suggèrent que les auteurs d'articles de recherche en français (Fr) ont un grand recours aux phrases conditionnelles dans des contextes argumentatifs, une tendance qui est aussi observée (dans un moindre degré) dans le cas des auteurs francophones d'articles en anglais (Fse). En revanche, les auteurs anglophones (En) ont un plus grand recours aux phrases conditionnelles dans les contextes descriptifs. Ces tendances se reflètent dans la distribution des trois catégories fonctionnelles de conditionnelles chez les trois groupes d'auteurs.

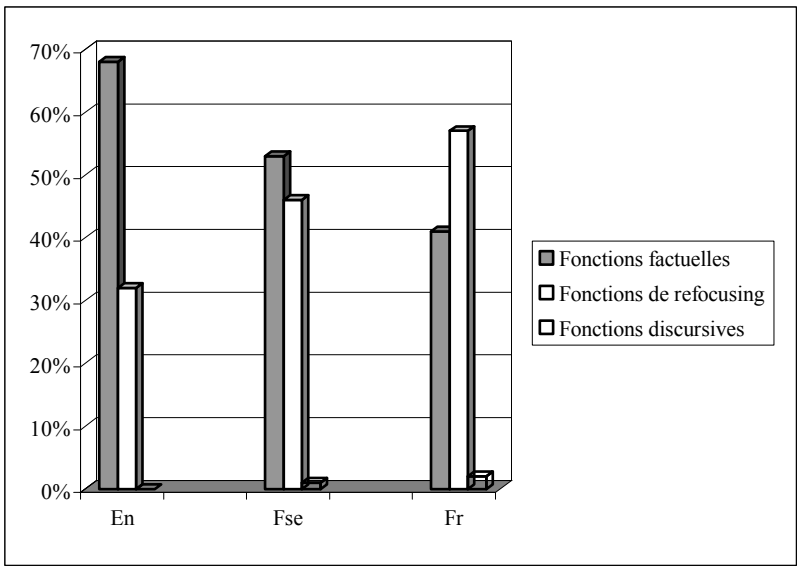


Figure 3. Distribution des catégories fonctionnelles chez les trois groupes d'auteurs [THOM 07 : 161].

Comme la figure 3 le montre, si les fonctions discursives sont peu représentées chez les trois groupes d'auteurs, les deux autres grandes catégories de fonctions identifiées ne sont pas distribuées de la même manière. Chez les auteurs anglais (En), ce sont les fonctions factuelles qui priment (68%). Les auteurs anglais ont plus recours à ce type de construction afin de présenter leurs décisions méthodologiques et thérapeutiques que les auteurs français (41%). (L'emploi des fonctions factuelles par les auteurs Fse se trouve à mi-chemin entre les deux). Les auteurs anglais semblent, par exemple, employer

presque systématiquement une conditionnelle pour établir les critères d'éligibilité pour les malades impliqués dans les différents tests :

> *(9) Patients were excluded from the study **if** they had evidence of metastases, a documented history of cardiac disease or previous cancer (…) FSE*

La légitimité de la relation entre P et Q est ainsi circonscrite à un sous-ensemble. En revanche, dans les articles français les chercheurs adoptent souvent d'autres moyens syntaxiques pour annoncer les critères d'éligibilité pour une expérience :

> (10) Les critères d'exclusion étaient ceux habituels pour la technique du GS : un stade TNM supérieur à T1 ou N1, un cancer multifocal, une chimiothérapie préalable, une allergie, obésité (…)

Le fait que les auteurs français annoncent les critères d'éligibilité de cette manière, sans recours à *si* donne l'impression d'une position plus affirmative, par opposition aux chercheurs anglophones, où la valeur suspensive ou « non assertée » de la proposition laisse au lecteur la possibilité de remettre en cause la décision prise par les auteurs.

Notre deuxième grande catégorie, les fonctions de réglage, *« refocusing functions »*, est plus représentée dans les articles français. Les auteurs français exploitent plus le potentiel argumentatif des constructions avec *si* que ne le font les auteurs anglophones avec *if.* (Le troisième groupe d'auteurs, les auteurs francophones d'articles en anglais, se trouve de nouveau à mi-chemin entres ces deux groupes). Ces différences trouvent leur origine en partie dans l'utilisation variable des constructions conditionnelles concessives par les trois groupes d'auteurs. Les constructions conditionnelles concessives avec *even if* ou *même si* sont particulièrement utiles dans l'article de recherche afin d'anticiper les objections potentielles des lecteurs.

> (11) Notre pourcentage est plus élevé que dans l'étude multicentrique, <u>même si</u> une comparaison est difficile compte tenu que 13 patientes de notre étude ont été incluses dans cette étude multicentrique.
>
> *(12) Even if noncancer deaths did not differ, the presence of concurrent adverse health conditions, or comorbidity, has been shown to increase cancer mortality.*

En employant une phrase conditionnelle concessive, le chercheur prévient d'éventuelles critiques de la part de ses lecteurs. Cependant, si les trois groupes d'auteurs ont recours aux conditionnelles de ce type, les auteurs français (Fr) les emploient deux fois plus que les auteurs anglais (En). Une comparaison stricte entre les deux langues s'avère néanmoins difficile à cause de la plus grande gamme de valeurs assurée par la conjonction *si*. En français, *si* peut souvent être employé à la place d'une autre conjonction concessive telle que *bien que*, pour signaler une relation concessive pure, non conditionnelle (Paillard 1989). Le corpus d'articles de recherche en français contient plusieurs exemples de ce type :

> (13) La limitation de la dissection du curage axillaire au seul étage I, **si** elle diminue le taux de complications, n'apparaît pas être une alternative satisfaisante ; le taux de faux négatifs observé étant de 10% à 15%.

En anglais de tels emplois sont peu fréquents. Quand la réalité de la proposition P est présupposée, l'anglais préfère employer un autre subordonnant concessif tel que *although* (Ranger 1998). Toutefois, même en écartant les concessives pures comme en (13), les chercheurs français ont largement plus recours aux conditionnelles concessives, et aux conditionnelles du type *« refocusing »* en général, que les auteurs anglais. Ce constat semble indiquer une volonté d'adopter une position argumentative plus prononcée de la part des auteurs français, plutôt qu'une simple différence linguistique liée à la gamme de valeurs plus large de *si* en français.

Les résultats présentés ici ont souligné à plusieurs reprises l'emploi un peu particulier qui est fait des phrases conditionnelles par le troisième groupe d'auteurs (Fse), les auteurs francophones des articles rédigés en anglais. Ces auteurs présentent en effet un profil hybride. D'une part, ils suivent les auteurs anglais en exploitant les fonctions factuelles des phrases conditionnelles. D'autre part cependant, comme dans les articles en français, ils exploitent plus le potentiel argumentatif de *if* que ne le font les auteurs anglophones. Je reviens dans la partie suivante sur d'autres différences dans l'emploi des phrases conditionnelles par les locuteurs anglophones et non anglophones, cette fois-ci dans les communications de congrès.

9.2 Emploi des phrases conditionnelles dans les communications de congrès

Les communications de congrès et les articles de recherche sont des genres proches. Ils sont gérés par les mêmes communautés discursives (la communauté des chercheurs universitaires) et peuvent être qualifiés d'exemples de « *research genres* » (Swales 2004). Une communication de congrès donne souvent lieu à un article scientifique ou vice versa. Toutefois, comme il a déjà été démontré dans les chapitres précédents (par exemple section 7.2) certaines conditions spécifiques aux communications de congrès, et en particulier les contraintes liées à leur mode de transmission et à leur statut épistémologique, entraînent des choix syntaxiques bien particuliers. L'emploi de phrases conditionnelles en est un exemple frappant. Je propose donc dans les paragraphes suivants de passer en revue certaines spécificités du fonctionnement des phrases conditionnelles dans les communications de congrès [THOM 08a et 08b]. Je passe ensuite à une comparaison de l'utilisation faite de cette structure par les conférenciers anglophones et non anglophones [THOM 08b].

Le corpus employé pour ces études comprenait trente communications de congrès faites en anglais dans le domaine de l'oncologie, quinze des conférenciers étant d'origine anglophone et quinze d'origine francophone[4]. Ce corpus de communications orales a été comparé à la partie anglaise du corpus d'articles de recherche en oncologie décrite plus haut (9.1.1). Seules ont été analysées les phrases conditionnelles en *if.*

[4] Les détails sur ces corpus peuvent être consultés en [THOM 08b : 45].

9.2.1 Fréquences, fonctions et formes (Contrastes avec l'article de recherche)

Genre	# textes	# mots	# occ.	Taux *if*/1000m.
Communication de congrès	30	72 106	216	3
dont : En	15	39 454	152	3,85
Fse	15	32 652	64	1,96
Articles de recherche	60	226 298	217	0,96
dont : En	30	111 907	119	1,06
Fse	30	114 391	98	0,85
Total	90	298.404	433	1,45

Tableau 1. Fréquence des phrases conditionnelles : communication de congrès *vs* articles de recherche (tableau adapté de [THOM 08b]).
En= Chercheurs d'origine anglophone ; Fse= Chercheurs d'origine francophone.

Comme le tableau 1 le démontre, les phrases conditionnelles sont trois fois plus fréquentes dans les communications de congrès que dans les articles de recherche. Ces résultats corroborent ceux d'autres études (Ford et Thompson 1986 ; Biber *et al.* 1999 ; Ferguson 2001) qui ont également conclu que les phrases conditionnelles sont plus fréquentes à l'oral qu'à l'écrit. Cependant, il existe également un décalage assez important dans l'emploi des phrases conditionnelles par les conférenciers anglophones et les conférenciers d'origine francophones, différence qui sera abordée dans la partie suivante (9.2.2).

Les traits formels des phrases conditionnelles dans la communication de congrès sont aussi très différents de ceux employés dans l'article de recherche À la différence de l'article de recherche, c'est le positionnement initial de la subordonnée qui est privilégié.

Genre	# Occ.	Position de la subordonnée (%)		
		Initiale	Médiane	Finale
Communication de congrès	216	75,5	11,1	12,6
Article de recherche	217	43,3	3,2	53,4

Tableau 2. Position des subordonnées en *if* (Tableau adapté de [THOM 08b]).

Cette préférence pour le positionnement initial semble très liée aux contraintes cognitives du traitement d'un discours en temps réel. Afin d'éviter des malentendus éventuels, il est souvent important d'établir un terrain en commun,

de savoir qu'il s'agit de P et non pas ~P, avant d'énoncer la proposition en Q. Les orateurs ont besoin aussi de fixer l'attention des auditeurs sur les nouveaux référents qui vont figurer dans le discours ultérieur. Ces fonctions « présentationnelles » des phrases conditionnelles sont manifestes dans plusieurs exemples du corpus des communications de congrès :

> *(14) If we look now at the plasma concentrations measured in the patients versus the days on therapy with the drug, we can see that in no case were plasma levels over 125 nanograms per milliliter.*

La fréquence plus élevée de la position médiane de la subordonnée semble aussi être liée aux contraintes de la communication en direct. Un orateur peut décider d'interrompre une phrase en cours afin d'insérer, comme une quasi-improvisation, une subordonnée en *if* pour expliciter les conditions de l'assertion dans la proposition principale :

> *(15) This tumor is very amenable if it's a carcinoma by biopsy to local excision.*

Ce type d'improvisation serait écarté dans un document écrit.

Concernant les formes verbales, les trois séquences conditionnelles prototypiques sont de nouveau peu exploitées (cf. *supra)*. Bien que légèrement plus fréquentes que dans les articles de recherche, ces séquences représentent seulement 12,5% des occurrences et il n'y a aucun exemple de la conditionnelle « 3 » (si + plus que parfait + conditionnel passé). Il y a de nouveau un éventail d'autres séquences verbales, la plus fréquente étant une séquence comportant deux verbes au présent. Deux autres configurations assez fréquentes dans les communications de congrès sont les conditionnelles tronquées et les conditionnelles sans apodose [THOM 08b : 51].

Les communications de congrès et les articles de recherches n'exploitent pas les mêmes fonctions des phrases conditionnelles. Bien que l'on trouve presque la même proportion de conditionnelles du type *« refocusing »* dans les deux genres, les fonctions factuelles des phrases conditionnelles sont beaucoup moins exploitées dans la communication de congrès (41% contre 61% dans les articles). Dans les articles de recherche, la plupart des conditionnelles

factuelles se trouvent dans la section « Méthodes ». Or, dans les communications de congrès, les explications sur la méthodologie sont généralement courtes : les conférenciers cherchent plus à souligner l'originalité de leurs travaux et/ou de leurs résultats aux dépens d'une présentation détaillée de leur méthodologie (Swales 2004). Pour ce faire, ils ont donc moins recours aux fonctions factuelles. Au contraire, c'est le recours aux fonctions discursives des conditionnelles qui caractérise mieux la communication de congrès. Alors qu'il y avait très peu de phrases conditionnelles discursives dans le corpus d'articles de recherche, les communications de congrès contenaient soixante occurrences (26%) de conditionnelles discursives.

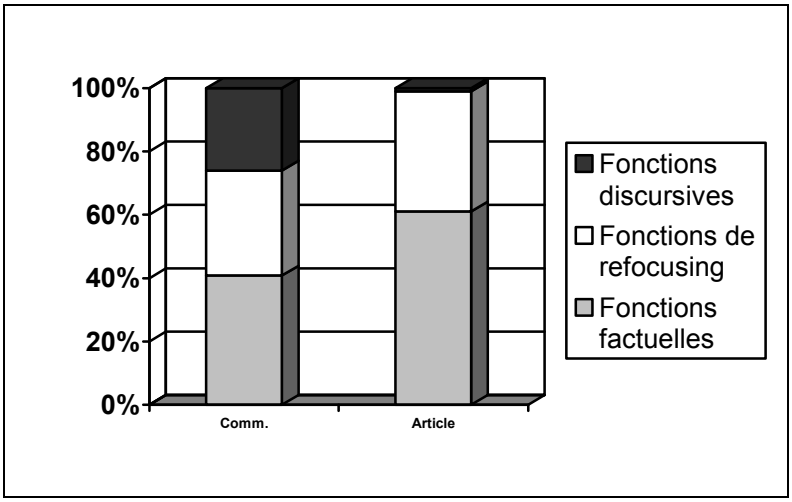


Figure 4. Fonctions des phrases conditionnelles : communication de congrès *vs* articles de recherche. Adapté de [THOM 08a].

La catégorie de fonctions discursives regroupe les fonctions qui sont liées à la gestion du discours. Comme on l'a constaté plus haut, ces fonctions sont peu exploitées dans l'article de recherche. La structure très formatée de l'article de recherche dispense souvent les auteurs d'indications métadiscursives supplémentaires. Or dans la communication de congrès la situation est tout autre. Dans un monologue dense, tel que la communication de congrès, les marqueurs organisationnels sont nécessaires afin de segmenter le flot discursif continu en de petites portions plus digestes et de signaler la structure informationnelle désirée. Comme le rappelle Péry-Woodley (2000 : 136),

« structurer revient toujours à segmenter », c'est-à-dire à la fois à diviser et à rassembler. Les conditionnelles du type discursif permettent aux orateurs de signaler explicitement le thème/topique, tout en créant des segments plus faciles à traiter pour les auditeurs. En voici un exemple :

(16) If we now go on to discuss the other problem, that is the local recurrence problem, this is a serious problem.

Une autre caractéristique de la communication de congrès qui favorise ce recours aux fonctions discursives est liée à la gestion de ressources visuelles. Les conférenciers doivent constamment veiller à intégrer les informations visuelles projetées à l'écran dans la trame linguistique de leur communication verbale. L'emploi d'une conditionnelle discursive permet d'attirer poliment l'attention des auditeurs sur une image ou un schéma que le conférencier serait en train de montrer :

(17) If you look at the number of patients that were treated, there (…)

Comme il a été souligné dans le chapitre précédent (8.2), l'emploi de phrases conditionnelles de ce type contribue également à créer un rapport direct avec les auditeurs présents dans la salle de conférence.

9.2.2 Comparaisons : conférenciers anglophones/conférenciers non anglophones

Parmi les résultats que l'on peut retenir de la comparaison sur l'emploi des phrases conditionnelles chez les deux groupes de conférenciers, deux points me semblent particulièrement significatifs. Le premier concerne leur fréquence (cf. tableau 1 ci-dessus). Si globalement les phrases conditionnelles sont trois fois plus fréquentes dans les communications de congrès que dans les articles de recherche, c'est principalement parce que les communications des conférenciers anglophones contenaient deux fois plus de phrases conditionnelles que celles des confrères non anglophones : 3,8/1000 dans les communications des chercheurs anglophones contre seulement 1,9/1000 dans les communications des chercheurs d'origine française. Autrement dit, les chercheurs francophones semblent beaucoup moins différencier les deux situations de communication que leurs confrères anglophones. La transition d'une syntaxe écrite vers une syntaxe orale est moins nette. Ces constats

permettent d'étayer certaines hypothèses déjà avancées dans mes travaux précédents sur l'emploi d'autres structures syntaxiques telles les phrases pseudo-clivées, les inversions ou les passifs [THOM 01 ; THOM 05d]. Dans tous ces cas, les structures caractéristiques du discours scientifique écrit sont réutilisées dans les communications des conférenciers non anglophones, tandis que les structures caractéristiques de la communication scientifique orale sont beaucoup moins en évidence.

Un deuxième point important qui ressort de cette comparaison entre les deux groupes d'auteurs concerne l'emploi de phrases conditionnelles discursives. Les conférenciers anglophones exploitent beaucoup plus les fonctions discursives des phrases conditionnelles que ne le font les conférenciers non anglophones. Il y avait quarante-sept occurrences (31%) dans le premier groupe contre seulement neuf occurrences (14%) dans le deuxième groupe. Bien que les conférenciers d'origine francophone emploient les fonctions factuelles et les fonctions du type *« refocusing »*, ils semblent peu familiers avec cet emploi pragmatique des phrases conditionnelles, sans doute parce qu'ils n'ont jamais utilisé ces fonctions dans leur apprentissage de l'anglais. Or comme on vient de le voir, les conditionnelles du type discursif peuvent jouer un rôle important dans l'organisation d'une communication, permettant aux orateurs de structurer leur discours et ainsi de faciliter son traitement par les auditeurs. L'incorporation de certaines explications sur de telles fonctions dans un enseignement de la syntaxe plus axé sur le genre pourrait donc être très utile pour les apprenants (cf. section 9.4 *infra*).

9.3 Emploi des phrases conditionnelles dans les éditoriaux spécialisés

Le troisième genre abordé dans mes travaux sur les phrases conditionnelles est l'éditorial spécialisé médical. Les éditoriaux spécialisés constituent un genre bien particulier, partageant certains traits du discours scientifique et d'autres du discours journalistique. Ils ont pour objectif de soulever des questions propres à la vie interne de la communauté, mais aussi d'attirer l'attention sur des faits ou des polémiques qui peuvent toucher toute la société Les éditorialistes des

revues médicales scientifiques travaillent ainsi à l'interface de la recherche scientifique et de l'opinion publique. À la différence des éditorialistes de la presse généraliste, ils ne sont pas généralement des journalistes mais des chercheurs et des membres de la communauté spécialisée concernée. De son côté, le lectorat des éditoriaux médicaux est assez hétérogène. Il comprend des chercheurs spécialistes, mais également des généralistes et parfois aussi la communauté scientifique dans son ensemble au sens large. À travers une comparaison avec l'article de recherche, il semblait donc intéressant d'essayer de cerner le rôle des phrases conditionnelles dans ce genre un peu hybride

Le corpus employé pour cette étude comportait soixante-quatorze éditoriaux de deux revues médicales de recherche, *The Lancet Oncology et International Journal of Radiation Oncology.* Ces éditoriaux étaient comparés à trente articles de recherche, en partie extraits des mêmes revues [THOM 09b]

9.3.1 Fréquences, fonctions et formes (Contrastes avec l'article de recherche)

Genre	# textes	# mots	# occ.	Taux *if*/1000m.
Editoriaux	74	95 143	171	1,8
Articles	30	111 907	119	1,06

Tableau 3. Fréquence des phrases conditionnelles : communication de congrès *vs* articles de recherche (tableau adapté de [THOM 09b]).

Comme le tableau 3 le démontre, les phrases conditionnelles sont plus fréquentes dans les éditoriaux médicaux que dans les articles de recherche. Ces résultats corroborent ceux de Ferguson (2001), qui a trouvé un taux similaire dans ses propres travaux sur l'emploi des conditionnelles dans le discours médical. Autrement dit, les éditorialistes semblent avoir plus besoin des phrases conditionnelles que les auteurs des articles de recherche

Il existe également une grande différence entre les traits formels et fonctionnels des phrases conditionnelles dans le cas des éditoriaux spécialisés et dans le cas des articles de recherche Dans le cas des éditoriaux spécialisés, les traits correspondent beaucoup plus à ce qui est généralement considéré comme les traits prototypiques des conditionnelles. Les conditionnelles hypothétiques sont

fréquentes et les trois types « canoniques » de séquences verbales sont par exemple beaucoup plus employés que dans les articles de recherche.

	Séquence 1	Séquence 2	Séquence 3	Séquences traditionnelles (1, 2, 3)
Éditoriaux	18	14	5	37 (21,6%)
Articles	2	4	1	7 (5,9%)

Tableau 4. Séquences verbales « canoniques » : comparaison éditoriaux /articles.

Les conditionnelles du type 1 (présent + futur) sont souvent employées dans les éditoriaux pour émettre des prévisions concernant l'environnement sociopolitique de la médecine :

> *(18) If the lack of charity regulation in Scotland <u>is allowed</u> to persist, future scandals… <u>will inevitably occur</u>.*

Une telle affirmation serait difficile à envisager dans l'article de recherche, car dans une enquête clinique sur les origines d'une maladie ou de son diagnostic la certitude absolue est difficile à assurer. En revanche, dans le domaine social, l'éditorialiste peut plus facilement se permettre de faire des pronostics assez fermes.

La plupart des conditionnelles du type 2 et tous les exemples du type 3 dans les éditoriaux sont des conditionnelles contrefactuelles, et leur emploi semble également largement déterminé par les objectifs argumentatifs du genre. Ces types d'hypothèse permettent aux locuteurs d'envisager les conséquences éventuelles des actions ou faits qui ne se sont pas produits dans le passé ou qui ne sont pas actuellement valides (Akatsuka and Strauss 2000). Ils permettent ainsi aux éditorialistes de critiquer les actions entreprises (ou non) par d'autres chercheurs.

> *(19) In the context of previous trials in solid tumors (…), one needs to consider the question of what the likely outcome would have been if the primary end point of this study had been survival, and the study had been continued despite the emergence of a statistically significant difference in progression-free survival.*

Dans cet exemple, l'éditorialiste émet des critiques sur une étude qui a été arrêtée trop vite pour fournir des informations valables sur la guérison des malades. Il est évident que, dans un article de recherche, un chercheur ne mettrait pas en question ses propres travaux de cette manière.

Les autres séquences verbales exploitées par les auteurs des articles et les éditorialistes sont également très différentes. Dans les articles de recherche, la majorité des séquences comportent un verbe au passé. En revanche, dans les éditoriaux, les formes du passé sont rares. L'argumentation est solidement ancrée dans le présent, avec 50% des séquences contenant un présent simple dans l'une ou les deux propositions de la phrase conditionnelle.

Les fonctions jouées par les phrases conditionnelles sont distribuées aussi d'une manière très différente dans les deux genres. Dans les éditoriaux, les fonctions factuelles sont rarement exploitées : 10% des occurrences contre 61% dans les articles. Les éditorialistes ne rendent pas compte de recherches primaires et ont par conséquent peu besoin des fonctions factuelles des conditionnelles qui servent dans les articles de recherche essentiellement à présenter les décisions de méthodologie et de traitement choisies par les chercheurs. En revanche, c'est surtout le recours aux fonctions de *« refocusing »* des conditionnelles qui caractérise ce corpus d'éditoriaux spécialisés : 87% des occurrences. Les phrases conditionnelles se trouvent dans les contextes argumentatifs où elles fournissent un certain potentiel de manœuvre aux éditorialistes pour traiter des questions polémiques et pour émettre des hypothèses sur les conséquences ou les suites à donner à la recherche examinée.

Le profil d'emploi des phrases conditionnelles dans les éditoriaux est donc globalement très différent de celui rencontré dans les articles de recherche. L'une des seules similarités semble concerner le positionnement de la subordonnée dans les deux genres. Bien que le positionnement initial de la protase ait été légèrement plus fréquent dans les éditoriaux (55%), ce pourcentage reste beaucoup plus près de celui relevé dans les articles de recherche (43%) que celui par exemple des communications de congrès (75%). Le mode de transmission semble jouer ici un rôle déterminant. Cependant, ces différences potentielles dans le fonctionnement des phrases conditionnelles

entre l'écrit et l'oral font rarement l'objet d'une présentation dans les grammaires pédagogiques (cf. *infra*).

9.3.2 L'étude de quatre variantes

Jusqu'ici nous avons surtout porté notre attention sur le schéma de base des phrases conditionnelles, la forme *If P, Q.* Toutefois, de la même manière que les formes verbales impactent le sens des phrases conditionnelles, il existe aussi plusieurs variantes aux schémas de base dont chacun ajoute une nuance particulière à la signification [THOM 09b] (Dancygier 1998[5]). Dans la section qui suit, je propose de brièvement résumer le rôle de quatre de ces variantes dans les éditoriaux : [*If P, then Q*], [*Q only if P*], [*even if P, Q*] ou [*Q even if P*] et [*if P, Q?*].

Ces variantes sont très fréquemment rencontrées dans les éditoriaux, représentant 26% des occurrences par rapport à seulement 7,5% des occurrences dans les articles de recherches. Le rajout de ces lexèmes, *then*, *only* et *even* et l'emploi d'une conditionnelle à l'apodose interrogative peuvent être extrêmement révélateurs quant aux stratégies rhétoriques des éditorialistes.

Variante	Éditoriaux	Articles
If P, then Q	12	4
Q only if P	7	0
Even if	11	5
If P, Q?	14	0
Total	44 26%	9 7,5%

Tableau 5. Quatre variantes de *If P, Q* [THOM 09b]

[5] Dancygier (1998 : 6) identifie un certain nombre de facteurs susceptibles d'impacter sur l'interprétation d'une construction conditionnelle, dont le rajout de lexèmes, les formes verbales, la « structure » (ordre des propositions) et l'intonation.

If P, then Q

Comme B. Dancygier (1998) et Dancygier et E. Sweetser (1997) l'ont montré, l'ajout de *then* dans l'apodose de la phrase conditionnelle permet de renforcer la relation de dépendance entre les propositions P et Q et de contraindre ainsi l'interprétation du récepteur. La variante avec *if P, then Q* est plus catégorique que la simple avec *if P, Q.* Les éditorialistes sont censés prendre position sur des questions d'actualité dans le monde médical et l'emploi de cette forme leur confère un certain statut d'autorité :

> *(20) If the provision of such services is not kept in touch with the increasing number of cancer survivors then the importance of quality of life in cancer will no longer be sustainable (ED 11).*

En revanche, chez les auteurs d'articles de recherche où une telle manifestation d'autorité pourrait apparaître présomptueuse, on trouve beaucoup moins d'occurrences de cette variante.

Q only if P

Cette variante est seulement employée dans les éditoriaux, car elle implique de nouveau un positionnement qui pourrait être perçu comme trop autoritaire dans le contexte d'un article de recherche. Grâce à la signification restrictive de *only,* la proposition en P est présentée comme la seule et unique condition pour l'assertion en Q.

> *(21) The project will succeed only if there are enough adequately trained radiologists to operate the equipment and interpret the results.*

Cette structure confère à l'éditorialiste un statut de pouvoir, l'autorisant à faire des déclarations définitives qui écartent toute autre option que le lecteur pourrait être tenté d'envisager.

Even if P, Q ou *Q even if P*

L'addition de *even* impose une interprétation concessive sur la phrase conditionnelle. Comme il a été démontré dans les sections précédentes (section 9.1.3), les conditionnelles de ce type peuvent jouer un rôle important dans l'argumentation dans la mesure où elles permettent de prévenir les critiques éventuelles de la part des lecteurs. Toutefois, comme mes travaux contrastifs

anglais/français l'ont montré [THOM 07], tout en étant fréquemment employées par les auteurs d'articles de recherche en français, les conditionnelles concessives sont peu sollicitées par les chercheurs anglais. En revanche, les éditorialistes anglais s'en servent plus fréquemment afin de combattre des arguments concurrents et de défendre leur propres positions [THOM 09b] :

> *(22) Even if a ban on smoking in public places only reduced the number of passive smokers presenting with tobacco-related diseases, a substantial saving would be made, both in human and economic terms.*

If P, Q?

Les conditionnelles à l'apodose interrogative apparaissent seulement dans le corpus des éditoriaux. Ces constructions ont tout d'abord une fonction interactionnelle importante. Le lecteur est invité à prendre part à la discussion. Cependant, en même temps, l'emploi de cette forme permet à l'éditorialiste d'exercer un certain contrôle sur l'interprétation du contenu. Les questions posées sont celles choisies par lui et auxquelles il a une réponse à suggérer. Il peut ainsi conduire les lecteurs vers l'interprétation souhaitée.

> *(23) If "fit elderly" patients can tolerate aggressive multimodality therapy, does this mean that all older patients should be treated this way? The answer is no.*

Dans l'exemple (23), en répondant à sa propre question, il adopte une attitude clairement didactique. Dans d'autres cas, l'intention de l'éditorialiste apparaît explicitement critique :

> *(24) After the serious limitations of this analysis are taken into careful consideration, the question remains: If a graft-versus-lymphoma effect does exist, why was evidence of it not observed in this analysis?*

En (24), l'emploi de la conditionnelle hypothétique dans la proposition P permet d'abord de mettre en doute les résultats de la recherche examinée, avant de les combattre avec force dans la forme interrogative en Q. Dans l'article de recherche, où rédacteur et lecteur sont beaucoup plus sur un pied d'égalité, de telles questions seraient considérées sans doute comme méprisantes.

Les recherches présentées dans cette section ont démontré que les phrases conditionnelles sont employées très différemment dans les articles de

recherche et dans les éditoriaux spécialisés. L'examen du fonctionnement de ces quatre variantes de la construction conditionnelle a été en particulier très révélateur des stratégies rhétoriques spécifiques adoptées par les éditorialistes. Alors que dans les articles de recherche l'argumentation est surtout établie sur des bases empiriques, dans les éditoriaux les opinions sont souvent exprimées d'une façon explicite et personnelle. Les éditorialistes sont censés prendre position par rapport aux polémiques dans le monde médical. L'emploi des variantes examinées, tout comme par exemple l'emploi de conditionnelles contrefactuelles, leur permet d'adapter leurs stratégies rhétoriques aux besoins du genre, afin de faire des conjectures, d'avertir, de critiquer, ou d'émettre des hypothèses contraires.

9.4 De la théorie à la pratique

Les études présentées dans les sections précédentes ont montré que les phrases conditionnelles fonctionnent très différemment dans ces trois genres. Leurs fréquences, leurs formes et leurs fonctions présentent un profil très différent en fonction du genre examiné. Dans deux précédents travaux [THOM 08a et 08b], nous avons cherché à confronter la réalité de leur usage dans ces trois genres professionnels avec les présentations des phrases conditionnelles faites dans la littérature linguistique sur les conditionnelles et dans les grammaires pédagogiques destinées aux apprenants de l'anglais.

Ces recherches révèlent qu'il existe un grand décalage entre ces usages dans le discours spécialisé et les présentations des phrases conditionnelles dans les cours de langues[6]. Dans les grammaires pédagogiques et les manuels destinés aux apprenants, l'accent est surtout placé sur la morphologie verbale. On distingue habituellement trois séquences verbales canoniques dans lesquelles les formes verbales sont associées à différents degrés de probabilité. Le positionnement initial de la subordonnée est généralement présenté comme la position par défaut, et on trouve peu d'attention accordée aux fonctions

[6] Si mes remarques portent surtout sur l'enseignement de l'anglais comme langue étrangère, d'après mes observations la situation en FLE (Français langue étrangère) est très similaire.

discursives des conditionnelles ou à leur rôle potentiel dans l'argumentation. Or si les chercheurs non anglophones se fient à ces types de description qu'ils auront sans doute rencontrés dans leur apprentissage de l'anglais, ils risquent d'être mal préparés pour à la fois comprendre et employer les phrases conditionnelles dans leur vie professionnelle de chercheur.

Nous avons formulé ces remarques critiques de la façon suivante en [THOM 08a] :

> *Restricting conditional meaning to degrees of hypotheticality, isolating only a very limited number of tense sequences, and ignoring discipline-specific forms of argumentation gives an impoverished picture of discourse practice. Any NNS*[7] *researchers relying too strongly on the above sequencing and explanations are therefore likely to experience a great discrepancy between their previous linguistic exposure to this structure and its uses and values in the highly focussed, purposeful disciplinary genres. This can be expected to lead to difficulties for them both in decoding actual usage – reading research journals or listening to research communications in their field – and in producing discourse that is rhetorically appropriate* [THOM 08a :]

L'exemple des conditionnelles habituellement étiquetées comme des irréelles du passé ou des conditionnelles contrefactuelles permet de bien illustrer ce décalage potentiel entre les réalités du discours spécialisé et les exemples des manuels. Ces conditionnelles occupent habituellement une place importante dans les présentations pédagogiques traditionnelles des conditionnelles. Or comme notre recherche l'a démontré, les contrefactuelles sont quasiment absentes dans les articles de recherche et dans les communications de congrès. L'exemple suivant est l'un des rares exemples que j'ai relevé dans un article de recherche en français :

> *Une cœlioscopie en fin d'intervention aurait permis d'éviter ce cancer, voire d'éviter sa survenue,* ***si*** *la patiente n'avait pas eu de contre-indication à cette voie d'abord (obésité sévère).*

L'absence des exemples de ce type peut s'expliquer par les objectifs très particuliers de ces deux exemples de *« research genres »* où les chercheurs essaient de donner l'impression que « tout ce qui a pu être fait a été fait ». Si

[7] Non native-speaker.

les chercheurs accordaient trop d'importance aux actions alternatives, leurs arguments se trouveraient affaiblis. Les approches ou les actions alternatives qui mettraient en doute leurs propres conclusions ne sont donc pas mentionnées. En revanche dans les éditoriaux, les conditionnelles contrefactuelles sont bien en évidence car les critiques des actions des autres font partie du mandat des éditorialistes. Enseigner les formes des conditionnelles sans prendre en compte leurs fonctions dans le discours semble peu utile.

Une autre caractéristique des phrases conditionnelles qui souligne ce décalage entre ce qui est prédit dans la littérature et l'usage dans le discours spécialisé est la question du positionnement de la subordonnée conditionnelle au sein de la construction conditionnelle. La position initiale est souvent présentée comme la position par défaut. Or d'après nos données le choix de positionnement serait surtout lié au mode de transmission et ne serait pas un attribut de la construction elle-même.

Les typologies des phrases conditionnelles proposées dans la littérature linguistique apparaissent de la même façon peu adaptées à l'étude de ces discours spécialisés. Une difficulté majeure avec les classifications proposées par Sweetser (1990) ou Athanasiadou et Dirven (1997) réside dans le fait que ces typologies sont fondées sur des exemples décontextualisés et par conséquent semblent trop peu détaillées pour la description du discours spécialisé. Comme Ford et Thompson l'ont signalé à juste titre :

> *Baseline data on what types of conditional occur and how they relate to their discourse contexts are essential if we hope to explain how conditionals are used rather than how we* ***think*** *they are used. (1986 : 354)*

Sweetser (1990) dégage par exemple trois catégories de conditionnelles : les conditionnelles de contenu, les conditionnelles épistémiques et les conditionnelles illocutoires. Toutefois, tandis que les conditionnelles épistémiques et conditionnelles illocutoires sont largement absentes dans nos corpus spécialisés, les conditionnelles *de contenu* semblent en revanche trop peu nuancées pour être utiles dans la description. C'est pour cette raison que

nous avons proposé notre propre classification fondée sur une approche de genre (cf. section 9.1.2 *supra*).

Enseigner les conditionnelles de façon très générale en s'appuyant sur des exemples décontextualisés, sans tenir compte des besoins spécifiques des genres, ne permet pas de bien préparer les apprenants aux réalités des discours spécialisés qu'ils peuvent rencontrer dans la vie professionnelle. Dans l'enseignement de structures courantes telles que les conditionnelles, une approche plus liée au genre s'avère fort utile. Une telle approche passe par une familiarisation avec les spécificités des configurations syntaxiques que les apprenants sont susceptibles de rencontrer en lisant, en écrivant ou en parlant le discours de leurs communautés spécialisées.

Texte et contexte

Texte et contexte : pour une approche fonctionnelle et empirique, le titre donné à ce mémoire reflète les grandes orientations adoptées dans mes travaux de recherche depuis une dizaine d'années. Pour conclure, je propose de brièvement revenir sur ces options théoriques et méthodologiques avant d'ébaucher quelques aspects applicatifs de mes recherches.

10.1 Options méthodologiques

Textes et contextes sont irrévocablement imbriqués. Si dans mes travaux j'accorde une place importante à la notion de texte et aux phénomènes de cohésion et de cohérence textuelles, il est tout aussi important pour moi de prendre en compte les contextes énonciatifs. Les formes textuelles examinées sont mises en relation avec des situations d'énonciation précises. Il est surtout question de l'interaction entre texte et contexte et de leur relation d'interdépendance. Mon objectif est d'examiner l'impact de certains choix linguistiques sur l'interprétation du texte en cours et en même temps d'évaluer l'influence des facteurs contextuels sur les choix qui sont faits.

Ce sont donc les approches fonctionnelles qui correspondent les mieux à ces objectifs. Comme le précise, entre autre S. Dik (1997 : 4), une approche fonctionnelle doit rendre compte de deux types de systèmes de règles, celles du système de la langue utilisée, et celles souvent moins conscientes liées au contexte de communication et aux règles pragmatiques qui le gouvernent Autrement dit, l'un des défis principaux de ces approches est d'expliquer les fonctionnement de certains choix de structures et de formulations syntaxiques par rapport aux situations dans lesquelles elles sont employées.

> *Linguistic expressions can be understood properly only when they are considered as functioning in settings, the properties of which are co-determined by the contextual and situational information available to speakers and addressees.* (1997 : 6)

Dans les travaux présentés ici, il a surtout été question du rôle structurant de la syntaxe et des instructions qu'elle véhicule au sein de certaines situations

communicatives déterminées. L'hypothèse qui sous-tend mes travaux est que l'organisation textuelle laisse forcément des traces à la surface des textes – des traces qui peuvent aider le récepteur dans la reconstruction du discours d'origine. Or l'identification (ou l'utilisation) de ces traces passe par une prise en compte de la situation linguistique qui a engendré le texte.

Dans la démarche proposée ici, le travail empirique sur corpus joue un rôle central et difficilement remplaçable. La description des traces ou instructions à la surface textuelle ne prend tout son sens que dans la mesure où elles sont rapportées au discours par lequel elles sont produites. Il devient alors impossible de se contenter d'exemples attestés. Il est essentiel dans ma conception d'un corpus d'avoir accès, non seulement aux textes intégraux, mais aussi de pouvoir situer très précisément le contexte extralinguistique des produits textuels sous étude. Le choix et la constitution même du corpus font donc partie intégrante de la méthodologie.

Mes travaux sont pour la plupart fondés sur des corpus très ciblés de petite taille, constitués au sein de genres scientifiques bien déterminés. Le choix de tels corpus permet de résoudre nombre d'incompatibilités potentielles entre les objectifs de la linguistique textuelle et la linguistique de corpus (cf. Péry-Woodley 2005 ; Biber *et al.* 2007). Un des reproches fréquemment fait par les linguistes travaillant sur les grands corpus à ceux qui s'intéressent aux notions liées à la structuration informationnelle ou à la cohésion textuelle, concerne la quantification des résultats et les difficultés à généraliser ces résultats à d'autres textes. Or la notion de genre aide en grande partie à résoudre ce problème. Se focaliser sur certaines régularités de la langue au sein de genres aux contours bien définis, c'est se donner les possibilités d'approcher de plus près de ce qui crée les spécificités linguistiques des usages adoptés. Si l'identification de marques de cohésion et de cohérence textuelles passe par une prise en compte de la situation linguistique qui a engendré le texte, ces marques peuvent être à un certain degré prévisibles à l'intérieur d'un genre bien déterminé.

L'interaction entre texte et contexte et leur relation d'interdépendance trouvent un terrain d'analyse particulièrement fructueux dans l'étude de différents genres textuels du milieu de la recherche scientifique institutionnalisé. Les pratiques sociales récurrentes dans ces milieux font que ces genres manifestent souvent une configuration stable d'acteurs, d'objectifs, de rôles et de *patterns* d'organisation textuelles. Il devient possible ainsi de dessiner les contours des genres avec précision, et de garantir de cette manière les conditions optimales pour mettre à jour les corrélats entre des traits linguistiques et des fonctions discursives

10.2 Aspects applicatifs

Si mes recherches sur les textes spécialisés impliquent nécessairement une analyse syntaxique fine, ces analyses ne sont pas entreprises dans une visée seulement descriptive. Des considérations d'ordre applicatif sous-tendent aussi la majorité de mes travaux. Une première application concerne la pédagogie et, plus précisément, l'enseignement ainsi que l'évaluation de l'expression écrite dans les cours de langues. La tendance assez répandue chez les apprenants d'une langue étrangère à vouloir s'occuper en priorité de la précision phrastique est souvent faite au détriment de l'organisation à un niveau textuel plus global. Or une sensibilisation à la structuration informationnelle des énoncés permet de franchir le passage de la phrase au texte et ainsi d'encourager les apprenants à devenir plus attentifs à certains défauts dans leur rédaction à un niveau interphrastique. Une prise en compte de cette dimension informationnelle aide aussi les enseignants à faire des corrections plus efficaces. À mon avis, l'appareil analytique de la structuration informationnelle présente des atouts qui ne sont toujours pas suffisamment exploités dans les méthodes didactiques.

Plus spécifiquement, mon intérêt grandissant pour l'analyse de textes scientifiques me permet d'envisager un certain nombre d'applications didactiques destinés aux chercheurs non anglophones et/ou des chercheurs novices dans l'organisation de leurs textes spécialisés en anglais (articles de recherche ; communication de congrès ; *abstracts* ; thèses, rapports). Ces

applications comprennent, par exemple, une sensibilisation à l'organisation rhétorique de ces textes, aussi bien qu'une sensibilisation aux marqueurs de la structuration informationnelle les plus sollicités dans les genres examinés. Texte et contexte sont ici étroitement entremêlés. Afin de comprendre pourquoi l'article de recherche scientifique privilégie l'emploi de certains types de configurations syntaxiques (telles les passifs et l'extraposition avec *it*), tandis que la communication de congrès en privilégie d'autres (telles les phrases pseudo-clivées ou les phrases conditionnelles), il est nécessaire de replacer ces structures dans leur contexte situationnel complet. La notion de genre nous aide à fixer ce contexte, à identifier les principaux enjeux de la communication et les stratégies nécessaires pour atteindre ces objectifs. Les structures syntaxiques dont vont avoir besoin des étudiants pour mener à bien les stratégies organisationnelles et rhétoriques appropriées vont dépendre de ce contexte et des contraintes qui y sont associées. Comme Enkvist l'a souligné :

> *The syntactic arrangment of the sentence is not independent but subservient to a text strategy.* (1991 :16)

Tout en enseignant les formulations linguistiques les plus usitées, il me semble aussi très important de montrer leur motivation contextuelle en situant cet enseignement au sein d'une approche de genre.

Un autre aspect applicatif de mes recherches que j'aimerais approfondir à l'avenir concerne les possibilités de modéliser certains aspects de la structuration rhétorique des textes scientifiques afin de faciliter la recherche d'informations ou l'acquisition des connaissances. Les articles de recherche scientifiques contiennent, par exemple, certaines étapes rhétoriques quasi obligatoires qui se manifestent par des configurations récurrentes de formulations syntaxiques ou lexicales. On sait par exemple que le *move* lors de l'introduction dans lequel les auteurs « établissent leur niche » (Swales 1990 ; chapitre 6 et chapitre 8 *supra*) sera souvent signalé par un connecteur adversatif, tel *however,* et que ce *move* sera habituellement lui-même suivi de près par un *move* signalent les objectifs de l'article *(In this paper, we propose (…)).* Les modèles de *moves* de Swales (1990) ont déjà fourni le point de

départ pour des schémas d'annotation en vue de la production des résumés automatiques (Teufel et Moens 1999*).* D'une façon analogue, F. Ibekwe-SanJuan (2005) propose d'extraire les indices de nouveauté dans les articles scientifiques afin de contribuer aux systèmes de veille scientifique et technologique. Or le repérage et l'analyse de tels indices de l'organisation rhétorique exigent non seulement une analyse linguistique très fine, mais aussi une très bonne connaissance des différences disciplinaires (de leurs traditions et de leurs épistémologies) ainsi que du type de publication et de son lectorat cible. La renommée d'une revue et l'étendue de sa diffusion auront aussi, par exemple, des répercussions sur les stratégies rhétoriques adoptées par les auteurs (Swales 2004). Seule une approche résolument multidisciplinaire, comportant des approches linguistique et informatique, mais aussi psychologique, épistémologique et sociologique sera à même de proposer des solutions vraiment performantes dans ce domaine

Au début de ce mémoire, j'ai évoqué la question de la nature et de la fiabilité des moyens de « contrôle » linguistique sur l'interprétation des récepteurs. Comme il a été démontré au fil de ces pages, ces indices linguistiques peuvent être assujettis à maintes variations. Tout est affaire de probabilité. Je crois néanmoins que ce sont les approches linguistiques incorporant une réflexion sur le genre qui nous permettront d'avancer.

Bibliographie

Note liminaire :
Afin de ne pas alourdir le texte par les nombreuses autoréférences rendues nécessaires par la nature de l'exercice, cette note de synthèse adopte un double système de référencement bibliographique.
Les références à mes propres travaux (en tant qu'auteur ou co-auteur) sont faites de la façon suivante : [THOM XX] et renvoient à la liste des travaux personnels ci-dessous.

Toutes les autres références (externes) figurent dans la bibliographie générale en fin de volume sous une forme conventionnelle : (Auteur(s), année).

1. Références aux travaux personnels

[THOM 94] Carter-Thomas, S. 1994. Langue de spécialité : cohésion, culture et cohérence. *ASp*, la revue du GERAS, Vol 5/6, 61-68.

[THOM 98] Carter-Thomas, S. 1998. Erreurs locales et erreurs globales : une contribution à l'analyse textuelle de l'anglais scientifique. *Actes de la 7ème journée ERLA/GLAT :* 267-280.

[THOM 99a] Carter-Thomas, S. 1999. Thematic networks and text type. *ASp*, la revue du GERAS, Vol. 23/26, 139-148.

[THOM 99b] Carter-Thomas, S. 1999. La stratégie thématique : son importance dans l'analyse textuelle. *Actes de la 5ème Journée de la Formation Doctorale*, Département de Linguistique Générale et Appliquée, Université René-Descartes/Paris V : 49-64.

[THOM 99c] Carter-Thomas, S. 1999. L'organisation thématique et ses conséquences sur la clarté d'un texte. In P. Parlebas (dir.) *Le Corps et le Langage*, L'Harmattan : 121-137.

[THOM 00] Carter-Thomas, S. 2000. *La cohérence textuelle – pour une nouvelle pédagogie de l'écrit*, Paris, L'Harmattan, coll. Langue et parole, 400p.

[THOM 01] Carter-Thomas, S., Rowley-Jolivet, E. 2001. Syntactic differences in oral and written scientific discourse: the role of information structure. *ASp,* la revue du GERAS, Vol. 31, 19-37.

[THOM 02] Carter-Thomas, S. 2002. Theme and information structure in French and English: A contrastive study of jounalistic clefts. *14th Euro-International Systemic Workshop, Issues in Language Description: Rethinking Systemic Functional Theory*, Université de Lisbonne, Portugal, 24-27 juillet 2002.
http://halshs.archives-ouvertes.fr/docs/00/40/36/89/PDF/Lisbonrevised.pdf

7[THOM 03] Carter-Thomas, S., Rowley-Jolivet, E. 2003. Analysing the scientific conference presentation (CP). A methodological overview of a multimodal genre. *ASp*, la revue du GERAS, Vol. 39-40, 59-72.

[THOM 04a] Tersis N., Carter-Thomas, S. 2004. Stability and flexibility of word order in East Greenlandic Tunumiisut. *Colloque: Crisis and stability*, Inalco, décembre 2004.
http://halshs.archives-ouvertes.fr/docs/00/28/55/92/PDF/stabart5sct.pdf

[THOM 04b] Carter-Thomas, S. 2004. Information Structure in practice: A discussion of some potential applications for discourse analysts and language teachers. *20th Scandinavian Conference of Linguistics,* Helsinki, 7-9 janvier 2004. (Résumé et exemplier).

[THOM 04c] Carter-Thomas, S. 2004. Specialised syntax for specialised texts? An examination of the preferred syntactic patterns in 2 agnate scientific genres: Proceedings articles and conference presentations. *Actes du colloque GLAT 2004 :* 11-21.

[THOM 04d] Carter-Thomas, S., Rowley-Jolivet, E. 2004. Structure informationnelle et genre : le cas de la communication de congrès. In M.M.J Fernandez-Vest, S. Carter-Thomas (éds), *Structure informationnelle et particules énonciatives – essai de typologie.* Paris : L'Harmattan : 137-167.

[THOM 05a] Rowley-Jolivet E, Carter-Thomas, S. 2005. Scientific conference Englishes: Epistemic and language community variations. In G. Cortese, A. Duszak (eds), *Identity, Community, Discourse: English in Intercultural Settings.* Berne, Peter Lang: 295-320.

[THOM 05b] Rowley-Jolivet E., Carter-Thomas, S. 2005. The rhetoric of conference presentation introductions: context, argument and interaction. *International Journal of Applied Linguistics*, Vol 15 (1), 45-71.

[THOM 05c] Tersis N., Carter-Thomas, S. 2005. Integrating syntax and pragmatics: Word order and transitivity variations in Tunumiisut. *International Journal of American Linguistics*, Vol. 71(4), 473-500.

[THOM 05d] Rowley-Jolivet E., Carter-Thomas, S. 2005. Genre awareness and rhetorical appropriacy: manipulation of information structure by NS and NNS scientists in the international conference setting. *English for Specific purposes,* Vol. 24 (1), 41-64.

[THOM 05e] Carter-Thomas, S. 2005. L'anglais à l'épreuve de la pluralité : Quelles implications pour la communication scientifique ? Communication faite au colloque : *L'avenir des langues et des sciences humaines dans les Grandes Ecoles*. INT, Evry, 13 mai 2005.
http://halshs.archives-ouvertes.fr/docs/00/40/57/15/PDF/Dix_ansGC.pdf

[THOM 07] Carter-Thomas, S. 2007. The 'iffiness' of medical discourse: cross-linguistic analysis of English 'if' and French 'si' in research articles. In K. Fløttum (ed.), *Language and Discipline Perspectives on Academic Discourse*, Cambridge Scholars Press: 150-176.

[THOM 08a] Carter-Thomas, S., Rowley-Jolivet, E. 2008. If-conditionals in medical discourse: From theory to disciplinary practice. *Journal of English for Academic Purposes*, Vol. 7(3), 191-205.

[THOM 08b] Rowley-Jolivet E., Carter-Thomas, S. 2008. When practice belies 'theory': Form, function and frequency of if-conditionals in specialised discourse. *ASp*, la revue du GERAS, Vol. 53/54, 39-61.

[THOM 08c] Sarda L., Carter-Thomas, S. 2008. Rôle informationnel et textuel des adverbiaux détachés : Exemples anglais et français en *sur* et *on*. Actes du Colloque international *Le Texte : modèles, méthodes, perspectives*, Université Babeş-Bolyai, Cluj-Napoca, Roumanie, 25-27 septembre 2008. http://halshs.archives-ouvertes.fr/docs/00/38/49/51/PDF/Soumission_Sarda_Carter-Thomas.pdf

[THOM 08d] Carter-Thomas, S. 2008. Teaching coherence through genre. In D. Alamargot, J. Bouchand, E. Lambert, V. Millogo, C. Beaudet (eds), Proceedings of the International Conference : *De la France au Québec – l'Ecriture dans tous ces états*. IUFM de Poitou-Charentes, 12-15 novembre 2008. http://www.poitou-charentes.iufm.fr/IMG/pdf/CarterThomas.pdf

[THOM 09a] Carter-Thomas, S. 2009. The French c'est-cleft: Function and frequency. *La linguistique systémique fonctionnelle et la langue française*, (textes réunis par D. Banks, S.Eason, J.Ormrod), Paris, L'Harmattan : 127-157.

[THOM 09b] Carter-Thomas, S., Rowley-Jolivet, E. Sous presse. Multiple agendas? A corpus-based analysis of if-conditionals in medical research editorials. *Les cahiers de CORPUS*, Numéro coordonné par L. Rouveyrol, M. Petit, Presses du CNRS.

2. Références bibliographiques

ACHARD, P. 1995. Formation discursive, dialogisme et sociologie. *Langages* 117 : 82-95.

ACHARD-BAYLE, G. 2005. Si polysémique et Si polyphonique. In L. Perrin (dir.), *Le sens et ses voix. Dialogisme et polyphonie dans la langue et le discours*. Coll. Recherches linguistiques 28, Université de Metz : 407-434.

ADAM, J.-M. 1985. *Le texte narratif.* Paris : Nathan.

ADAM, J.-M. 1987. Types de séquences textuelles élémentaires. *Pratiques* 56 : 54-80.

ADAM, J-M. 1990. *Éléments de linguistique textuelle*. Liège : Mardaga.

ADAM, J-M. 1997. *Les textes : types et prototypes. Récit, description, argumentation, explication et dialogue*. Paris : Nathan. Coll. Fac. linguistique.

ADAM, J-M. 1998. Si hypothétique : la fiction dans la langue et dans le discours scientifique. In C. Coquio, R. Salado (éds.), *Fictions et connaissance.* Paris : L'harmattan : 55-77.

ADAM, J-M., HEIDMANN, U. 2004. Des genres à la généricité. L'exemple des contes (Perrault et les Grimm). *Langages* 153 : 62-72.

AKATSUKA, N.M., STRAUSS, S. 2000. Counterfactual reasoning and desirability. In E. Couper-Kuhlen, B. Kortmann (eds.), *Cause – Condition – Concession – Contrast*. Berlin: Mouton de Gruyter: 205-234.

ATHANASIADOU, A., DIRVEN, R. 1997. Conditionality, hypotheticality, counterfactuality. In A. Athanasiadou and R. Dirven (eds.), *On Conditionals Again.* Amsterdam: Benjamins: 61-96.

ATKINS, S., CLEAR, J., OSTLER, N. 1992. Corpus design criteria. *Literary and Linguistic Computing* 7(1): 1-16.

BANKS, D. 1994. *Writ in water. Aspects of the scientific journal article*. ERLA : Université de Bretagne occidentale.

BANKS, D. 1995. There is a cleft in your sentence: Less common clause structures in scientific writing. *ASp,* la revue du Geras 7-10 : 3-12.

BANKS, D. 1999. Decoding the information structure of journalistic clefts. *Interface : Journal of Applied Lingusitics* 14:1: 3-24.

BANKS, D. 2002. Systemic functional linguistics as a model for text analysis, *ASp,* la Revue du Geras 35/36 : 23-35.

BANKS, D. 2004. Anglophone systemicists and French enunciativists: shall the twain never meet? *Language Sciences* 26:4: 391-410.

BANKS, D. 2005. *Introduction à la linguistique systémique fonctionnelle de l'anglais*. Paris : L'Harmattan.

BARBER, C.L. 1962. Some measurable characteristics of modern scientific prose. *Contributions to English Syntax and Philology*. Göteborg: Acta Universitatis gothoburgensis: 21-43.

BAZERMAN, C. 1988. *Shaping Written knowledge: The Genre and Activity of the Experimental Article in Science.* Madison: University of Wisconsin Press.

BEACCO, J-C., MOIRAND, S. 1995. Autour des discours de transmission de connaissances. *Langages* 117 : 32-53.

BEAUGRANDE DE, R., DRESSLER, W. 1981. *Introduction to Text Linguistics*, London: Longman.

BERLAND-DELEPINE, S. 1989. *La Grammaire anglaise de l'étudiant*. Gap : Ophrys.

BERRY, M. 1996. What is Theme? – A(nother) personal view. In M. Berry, C. Butler, R. Fawcett, G. Huang (eds.), *Meaning and Form: Systemic Functional Interpretations*, Norwood, N.J.: Ablex. 1-65.

BIBER, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

BIBER, D. 1993. Using register-diversified corpora for general language studies. *Computational Linguistics* 19: 219-241.

BIBER D., JOHANSSON, S., LEECH, G., CONRAD, S., FINEGAN, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Ltd.

BIBER, D., CONNOR U., UPTON T. A. 2007. *Discourse on the Move*. Amsterdam: John Benjamins

BLOOR, T., BLOOR, M. 1995. *The Functional Analysis of English: a Hallidayan Approach.* London: Arnold.

BRANCA-ROSOFF, S. 1999. Types, modes et genres : entre langue et discours. *Langage et société* (87) : 5-24.

BRANCA-ROSOFF, S., FLEURY, S., LEFEUVRE, F., PIRES, M. 2009. *Discours sur la ville. Corpus de français parlé parisien des années 2000 (CFPP2000)* http://ed268.univ-paris3.fr/CFPP2000/

BREIVEGA, K.R., DAHL, T., FLØTTUM, K. 2002. Traces of self and others in research articles. A comparative pilot study of English, French and Norwegian research articles in medicine, economics and linguistics. *International Journal of Applied Linguistics* 12 (2): 218–239.

BROWN, G., YULE, G. 1983. *Discourse Analysis*, Cambridge: Cambridge University Press.

BUTLER, C.S. 2003. *Structure and Function. Part 2: From Clause to Discourse and Beyond*. Amsterdam: John Benjamins.

BUTLER, C.S. 2005. Functional approaches to language, in C.S. Butler, M.L.A. Gómez-González, S. Doval-Suárez (eds.), *The Dynamics of Language Use: Functional and Contrastive Perspectives*. Amsterdam/Philadelphia: John Benjamins: 3-17.

CAFFAREL, A. 2000. Interpreting French theme as a bi-layered structure: Discourse implications. In E. Ventola (ed.), *Discourse and community: Doing Functional Linguistics.* Tubingen: Gunter Narr.

CAMICIOTTOLI, B., CRAWFORD. 2004. Interactive discourse structuring in L2 guest lectures: some insights from a comparative corpus-based study. *Journal of English for Academic Purposes*, Vol. 3 (1): 39-54.

CANAGARAJAH, A.S. 1999. *Resisting linguistic imperialism in English teaching.* Oxford: Oxford University Press.

CHAFE, W.L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and points of view. In C.N. Li (ed), *Subject and Topic*. New York: Academic Press.

CHAFE, W.L. 1987. Cognitive constraints on information flow. In R. Tomlin (ed.), *Coherence and Grounding in Discourse,* Amsterdam: John Benjamins.

CHAROLLES, M. 1978. Introduction au problème de la cohérence des textes, *Langue Française* 38 : 7-41.

CHAROLLES, M. 1988. Les études sur la cohérence, la cohésion et la connexité textuelles depuis la fin des années 1960, *Modèles Linguistiques* 10 (2) : 45-66.

CHAROLLES, M. 1993. Les plans d'organisation du discours et leurs interactions. In S. Moirand *et al*. (éds.), *Parcours linguistiques de discours spécialisés*. Berne : Peter Lang : 301-314.

CHAROLLES, M. 1995. Cohésion, cohérence et pertinence du discours, *Travaux de Linguistique* 29 : 125-151.

CHAROLLES, M. 1997. *L'encadrement du discours*, document Landisco, Nancy.

CHAROLLES, M. 2003. De la topicalité des adverbiaux détachés en tête de phrase. In M. Charolles, S. Prévost (éds.), Adverbiaux et topiques, *Travaux de Linguistique* 47 : 11-51.

CHAROLLES, M., COMBETTES, B. 1999. Contribution pour une histoire récente de l'analyse du discours. *Langue française* 121 : 76-116.

CHAROLLES, M., PERY-WOODLEY, M-P. 2005. Les adverbiaux cadratifs : introduction. In M. Charolles, M-P. Péry-Woodley (éds.), Les adverbiaux cadratifs, *Langue Française* 148 : 3-8.

CLYNE, M. 1998. Cultural differences in the organization of academic texts: English and German. In J. Cheshire, P. Trudgill (eds.), *The Sociolinguistics Reader*, Vol. 2. London: Arnold: 315-347.

COLLINS, P. 1991. *Cleft and Pseudo-cleft Constructions in English*. London/New York: Routledge.

COMBETTES, B. 1977. Ordre des éléments dans la phrase et linguistique du texte. *Pratiques* 13 : 81-101.

COMBETTES, B. 1978. Thématisation et progression thématique dans les récits d'enfants. *Langue Française* 38 : 74-86.

COMBETTES, B. 1983. *Pour une grammaire textuelle. La progression thématique*. De Boeck-Duculot.

COMBETTES, B., TOMASSONE, R. 1988 (2e tirage 1991). *Le texte informatif*. Bruxelles : De Boeck-Wesmael.

COMRIE, B. 1986. Conditionals: A typology. In E.C. Traugott, A. ter Meulen, J.S. Reilly, C.A. Ferguson (eds.), *On Conditionals*. Cambridge: Cambridge University Press: 77-99.

COMRIE, B. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. 2nd ed. Oxford: Blackwell Publishers.

CONDAMINES, A. 2003. *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques*, Habilitation à diriger les recherches en sciences du langage, Université de Toulouse-Le Mirail.

CONDAMINES, A. 2005. Sémantique et corpus : quelles rencontres possibles ? Introduction. In A. Condamines (ed.) : *Sémantique et Corpus*. Hermes : 17-38.

CONNOR, U. 1984. A study of cohesion and coherence in English as a second language students' writing. *Papers in Linguistics* 17: 301-316.

CORNISH, F. 2006. Relations de cohérence en discours : critères de reconnaissance, caractérisation et articulation cohésion-cohérence. *Corela*, numéro spécial. Accessible en ligne sur :htttp://edel.univ-poitiers.fr/corela/personne.php?id=1281 et type=auteur. (Consulté le 13.6.2009.)

COUTO, J., LUNDQUIST, L., MINEL, J-L. 2005. Naviguer dans les textes pour apprendre. In *TALN 2005.*

CROFT, W. 2003. *Typology and Universals,* 2nd edition. Cambridge: Cambridge University Press.

CROMPTON, P. 2006. The effect of position on the discourse scope of adverbials, *Text and Talk* 26-3: 245-279.

DAHL, T. 2004a. Absent doctors, shy economists and polemic linguists? Writer manifestation in academic texts. *Synaps* 14: 1-4. Instituttserie, Institutt for fagspråk og interkulturell kommuinkasjon, NHH

DAHL, T. 2004b. Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics* 36 (10) 1807–1825.

DANCYGIER, B. 1998. *Conditionals and Predictions: Time, Knowledge and Causation in Conditional Constructions.* Cambridge University Press.

DANCYGIER B., SWEETSER, E. 1997. Then in conditional constructions. *Cognitive Linguistics* 2: 109-136.

DANEŠ, F. 1964. A Three-Level Approach to Syntax. *Travaux linguistiques de Prague*, vol. 1: 225-240.

DANEŠ, F. 1970. One instant of Prague School Methodology: Functional Analysis of Utterance and Text, in P. Garvin (ed.), *Method and Theory in Linguistics*. Le Hague: Mouton: 132–146.

DANEŠ, F. 1974. Functional sentence perspective and the organisation of the text, in F. Daneš (ed), *Papers in Functional Sentence Perspective,* Prague: Academia.

DANES, F. 1978. De la structure sémantique et thématique du message, in *Linguistique et Sémiologie, Textlinguistik*, Université de Lyon, coll. Travaux du Centre de recherches linguistiques et sémiologiques de Lyon : 177-201.

DAVISON, A. 1984. Syntactic markedness and the definition of Sentence Topic. *Language* 60: 4 : 797-846.

DAY, R.A. 1995. *How to Write and Publish a Scientific Paper*, 5th edition, Cambridge: Cambridge University Press.

DIK, S.C. 1997. *The Theory of Functional Grammar 1: The Structure of the clause*, 2e édition révisée, édité par K. Hengeveld. Berlin: Mouton de Gruyter.

DOWNING, A. 1991. An alternative approach to theme: a systemic-functional perspective. *Word,* Vol. 42 (2) : 119-143.

DOWNING, A. 2001. Thematic Progression as a Functional Resource in Analysing Texts.
CLAC :http://www.ucm.es/info/circulo/no5/downing.htm.(Consulté le 9/6/09.)

DUBOIS, B-L. 1980. Genre and structure of biomedical speeches. *Forum Linguisticum* Vol. V: 140-168.

DUBOIS, B-L. 1987. A reformulation of thematic progression typology. *Text,* 7(2): 89-116.

DUCROT, O. 1972. *Dire et ne pas Dire*, Paris : Herman.

DUCROT, O., TODOROV, T. 1972. *Dictionnaire encyclopédique des sciences du langage*. Paris : Le Seuil.

DUDLEY-EVANS, T. 1987. Genre Analysis and ESP, *ELR Journal*, Vol. 1 : 1-9.

DUDLEY-EVANS, T. 1993. Subject specificity in ESP: How much does the teacher need to know of the subject? *ASp,* la revue du Geras,1 : 1-19.

DUDLEY-EVANS, T. 1994. Genre analysis: An approach to text analysis for ESP. In M. Coulthard (ed.), *Advances in Written Text Analysis*, London: Routledge: 219-228.

DUDLEY-EVANS, T. 1995. Genre models for the teaching of academic writing to Second Language Speakers. In *TESOL France, Journal*, vol. 2: 181-192.

EHRLICH, S. 1988. Cohesive devices and discourse competence. *World Englishes,* Vol. 7 (2): 111-118.

ENKVIST, N.E. 1976. Notes on valency semantic scope and thematic perspective as parameters of adverbial placement in English. In N.E. Enkvist, V. Kohonen (eds.), *Reports on Text Linguistics: Approaches to Word Order.* Publications of the research institute of the Åbo Akademi Foundation: Åbo: 51-74.

Enkvist, N.E. 1978. Coherence, pseudo-coherence, and non-coherence. In J-O. Östman, (eds.) *Reports on text linguistics: Semantics and cohesion*. Åbo: Åbo Akademi: 109–128.

ENKVIST, N.E. 1984. Contrastive linguistics and text linguistics. In Fisiak (ed.) *Contrastive linguistics: Prospects and problems*. Mouton de Gruyter: 45-67.

ENKVIST, N.E. 1985. Text linguistics for the Applier: An Orientation. In U. Connor, R.B. Kaplan (eds.), *Writing across Languages: Analysis of L2 Texts,* Reading, MA: Addison-Wesley*:* 23-42.

ENKVIST, N.E. 1989. Connexity, interpretability, universes of discourse, and text worlds. In S. Allén (ed.) *Possible Worlds in Humanities, Arts and Sciences.* Berlin/New-York: Walter de Gruyter: 162-186.

ENKVIST, N.E. 1991. Discourse strategies and discourse types. In E. Ventola (ed.), *Functional and Systemic Linguistics: Approaches and Uses.* Berlin: Mouton de Gruyter.

ERTESCHIK-SHIR, N. 1999. Focus structure and scope. *Grammar of Focus*. G. Rebuschi, L. Tuller (eds.). Amsterdam/Philadelphia: John Benjamins: 119-150.

FAWCETT, R.P., HUANG, G.W. 1995. A functional analysis of the enhanced theme construction in English. *Interface: Journal of Applied Linguistics* 10/1: 113-44.

FERGUSON, G. 2001. If you pop over there: a corpus-based study of conditionals in medical discourse. *English for Specific Purposes* 20/1: 61-82.

FERNANDEZ-VEST, M.M.J. 1994. *Les Particules énonciatives*. Paris : PUF.

FERNANDEZ-VEST, M.M.J. CARTER-THOMAS, S. (éds.). 2004. *Structure informationnelle et particules énonciatives – essai de typologie*. Paris : L'Harmattan.

FIRBAS, J. 1964. On defining the theme in functional sentence analysis. *Travaux linguistiques de Prague* vol.1 : 267-280.

FIRBAS, J. 1986. On the dynamics of written communication in the light of the theory of functional sentence perspective. In Cooper and Greenbaum (eds.), *Studying Writing: Linguistic Approaches*. Beverley Hills, CA: Sage: 40-71.

FIRBAS, J. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.

FLØTTUM, K. 2005. The self and the others – polyphonic visibility in research articles. *International Journal of Applied Linguistics* 15: 29-44.

FLØTTUM, K., DAHL, T., KINN, T., GJESDAL, A.M., VOLD, E.T. 2007. Cultural Identities and Academic Voices. In K. Fløttum (ed.), *Language and Discipline Perspectives on Academic Discourse*. Cambridge: Cambridge Scholars Publishing: 14-39.

FORD, C.E., THOMPSON, S.A. 1986. Conditionals in discourse: A text-based study from English. In E. Traugott *et al.* (eds.), *On Conditionals*. Cambridge: Cambridge University Press: 353-372.

FORTESCUE, M. 1993. Eskimo Word Order Variation and its Contact-induced Perturbation. *Journal of Linguistics* 29 (2): 267-289.

FRADIN, B. 1990. Approche des constructions à détachement : inventaire. *Revue Romane* 25 : 1, 3-34.

FRADIN, B., CADIOT, P. 1988. Présentation : Une crise en thème ? *Langue française* 78 : 3-8.

FREEDMAN, A., MEDWAY, P. 1994. Locating genre studies: Antecedents and prospects. In A. Freedman, P. Medway (eds.): *Genre and the New Rhetoric.* London: 1-20.

FRIES, P.H. 1983. On the status of theme in English: arguments from discourse. In J. Petöfi, E. Sozer (eds.), *Micro and Macro Connexity of Texts*. Hamburg: Buske.

FRIES, P.H. 1994. On Theme, rheme and discourse goals. In Coulthard (ed.), *Advances in Written Text Analysis*. London: Routledge.

FRIES, P.H. 1995. Themes, methods of development, and texts In R. Hasan, P.H. Fries, *On Subject and Theme. A Discourse Functional Perspective*. Amsterdam/Philadelphia: Benjamins: 317-359.

FUCHS, C., FOURNIER, N. 2003. Du rôle cadratif des compléments localisant initiaux selon la position du sujet nominal. *Travaux de Linguistique* 47 : 79-110.

FUCHS, C. 2006. La place du sujet nominal en français : de la syntaxe à l'énonciation. In F. Hrubaru et A. Velicu (éds.) *Enonciation et syntaxe.* Cluj : Echninox : 9-25.

GALMICHE, M. 1992. Au carrefour des malentendus : le thème. *L'information Grammaticale* 54 : 3-10.

GHADESSY, M., HENRY, A., ROSEBERRY, R.L. 2001. *Small Corpus Studies and ELT*. John Benjamins.

GIVÓN, T. 1979. *On Understanding Grammar*, New York/San Francisco/London: Academic Press.

GIVÓN, T. 1988. The pragmatics of word order: Predictability, importance and attention. In M. Hammond, E.A. Moravcsik, J. Wirth (eds.), *Studies in Syntactic Typology*.

GREENBERG, J.H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J.H. Greenberg (ed), *Universals of Language*, Cambridge MA: 75-113.

GROBET, A. 2002. *L'identification des topiques dans les dialogues*. De-Boeck-Duculot.

HABERT, B., NAZARENKO, A., SALEM, A. 1997. *Les linguistiques de corpus*. Paris : Colin.

HAGEGE, C. 1978. Du thème au thème en passant par le sujet : pour une théorie cyclique. *La linguistique*, 14 (2) : 3-38.

HAGEGE, C. 1982. *La structure des langues.* Paris : Presses Universitaires de France.

HALLIDAY, M.A.K. 1967. Notes on transitivity and theme in English, Part 2. *Journal of Linguistics* 3(2): 199-244.

HALLIDAY, M.A.K. 1970. Language structure and language function In J. Lyons (ed.), *New Horizons in Linguistics*. Penguin Books: 140-165.

HALLIDAY, M.A.K. 1994. *An introduction to Functional Grammar.* London: Arnold

HALLIDAY, M.A.K, HASAN, R. 1976. *Cohesion in English.* London: Longman.

HARTNETT, C.G. 1986. Static and dynamic cohesion: signals of thinking in writing. In B. Couture (ed.), *Functional Approaches to Writing. Research Perspectives*, London: Frances Pinter.

HARWEG, R. 1973. Text grammar and literary texts. *Poetics* 9.

IBEKWE-SANJUAN F. 2005. Annotation d'indices de nouveautés dans les écrits scientifiques et techniques. Colloque *Indice, Index, Indexation*, Université de Lille 3, 3-4 novembre : 261-275.

HINDS, J. 1987. Reader versus writer responsibility: A new typology. In U. Connor, R. Kaplan (eds.), *Writing across Languages: Analysis of L2 Text*. Addison-Wesley, Reading, MA: 141–152.

HYLAND, K. 1999. Disciplinary discourses: writer stance in research articles. In C.N. Candlin, K. Hyland. *Writing: Texts, Processes and Practices*. London: Longman: 99-122.

HYLAND, K. 2000. *Disciplinary Discourses: Social Interactions in Academic Writing.* London: Longman.

JACQUES, M.-P. 2005. Pourquoi une linguistique de corpus ? In G. Williams (ed.), *La Linguistique de corpus*. Presses Universitaires de Rennes : 21-30.

JAKOBSON, R. 1963. *Essais de linguistique générale*. Paris : Minuit.

JESPERSEN, O. 1937. *Analytic Syntax*. London : Allen and Unwin.

JOHANSSON, S. 1998. On the role of corpora in cross-linguistic research. In S. Johansson, S. Oksefjell (eds.) *Corpora and Cross-linguistic Research. Theory, Method and Case Studies*. Amsterdam/Atlanta: Rodopi: 3-24.

JOHANSSON, M. 2001. Clefts in contrast: a contrastive study of *it* clefts and *wh* clefts in English and Swedish texts and translations *Linguistics* 39-3: 547-582.

KASSAI, G. 1976. À propos de la linguistique du texte, *La Linguistique* vol. 12 (2) : 119-128.

KATZ, J.J. 1980. Chomsky on meaning. *Language* 56: 1-42.

KATZ, S. 1997. *The syntactic and pragmatic properties of the c'est-cleft construction*, Doctoral dissertation, University of Texas.

KERBRAT-ORRECHIONI, C. 1996. Texte et contexte. P. Schmoll (ed), *Scolia* 6, Contextes : 39-60.

KLINGLER, D. 2003. Spécificité du dispositif créé par le marqueur wa en japonais comparaison avec le français. *Travaux de linguistique* 2(47).

LAMBRECHT, K. 1987. On the status of SVO sentences in French discourse. In R. Tomlin (ed.), Coherence *and Grounding in Discourse*. Amsterdam: Benjamins: 217-262.

LAMBRECHT, K. 1994. *Information structure and sentence form*, Cambridge: Cambridge University Press.

LAMBRECHT, K. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39-3: 463-516.

LAMBRECHT, K. 2004. Un système pour l'analyse de la structure informationnelle des phrases. L'exemple des constructions clivées, In M.M.J. Fernandez-Vest, S. Carter-Thomas (éds.). *Structure informationnelle et particules énonciatives – essai de typologie.* Paris : L'Harmattan.

LAPAIRE, J.R., ROTGE, W. 1992. *Réussir le commentaire grammatical de textes*. Ellipses.

LAUTAMATTI, L. 1987. Observations on the development of simplified discourse. In U. Connor, R.B. Kaplan (eds.), *Writing across Languages: Analysis of L2 Texts*, Reading, MA., Addison-Wesley: 87-114.

LEECH, G. 1992. Corpora and theories of linguistic performance. In J. Svartvik (ed.), *Directions in corpus linguistics,* Berlin/New York: Mouton de Gruyter: 105-122.

LE GOFFIC, P. 1993. *Grammaire de la phrase française*. Paris : Hachette.

LERAT, P. 1995. *Les langues spécialisées*, Paris : PUF.

LI, C.N., THOMPSON, S.A. 1976. Subject and topic: A new typology of language. In C.N. Li (ed), *Subject and Topic*. New York: Academic Press.

LONGACRE, R.E. 1976. *An analysis of speech notions.* Lisse: Peter de Ridder Press.

LUNDQUIST, L. 1980. *La cohérence textuelle : syntaxe, sémantique, pragmatique*. Copenhague : Nyt Nordisk Forlag.

LUNDQUIST, L. 1998. Cadres et espaces mentaux dans la structuration cognitive des connaissances – et des discours – juridiques. In Y. Gambier (Ed.), *Discours professionnels en français*. Bern : Peter Lang : 123-149.

LUNDQUIST, L. 1999. Le factum Textus : fait de grammaire, fait de linguistique ou fait de cognition. *Langue française* 121 : 56-75.

LUNDQUIST, L. 2005. Noms, verbes et anaphores (in)fidèles. Pourquoi les Danois sont plus fidèles que les Français. *Langue française* 145 : 73-91.

LUNDQUIST, L. 2008. *Navigating in Foreign Language Texts*. Copenhague: Samfunds Litteratur.

MAGNET, A. 2000. Etude de l'article scientifique de recherche en anglais : discours sur la méthode. *L'anglais de spécialité en France*, Collection ASp. Geras éditeur.

MAINGUENEAU, D. 1991. *L'analyse du discours : introduction aux lectures de l'archive*. Paris : Hachette.

MAINGUENEAU, D. 1996. *Les termes clés de l'analyse du discours.* Paris : Seuil.

MARCHELLO-NIZIA, C. 1999. Corpus diachroniques. *Revue française de linguistique appliquée* IV(1) :31-39.

MARTIN, J. 1997. Du bon usage du corpus dans la recherche sur le discours spécifique*. ASp,* la revue du Geras, 15/18 : 75-84.

MARTIN, R.R. 1992. *English Text; System and Structure*. Philadelphia: John Benjamins

MAURANEN, A. 1996. Discourse competence - Evidence from thematic development in native and non-native texts. In E. Ventola, A. Mauranen (eds.), *Academic Writing: Intercultural and Textual Issues.* Amsterdam/Philadelphia: John Benjamins: 195-231.

MCCABE, A. 1999. *Theme and Thematic Patterns in Spanish and English History Texts.* Unpublished Ph.D. Thesis. Aston University, Birmingham. U.K.

MCCABE, A. 2004. Thematic progression patterns and text types in history textbooks. In D. Banks (Ed.) *Text and Texture : Systemic Functional viewpoints on the nature and structure of text*. Paris : L'Harmattan.

MCENERY, T., WILSON, A. 1996. *Corpus Linguistics*, Edinburgh: Edinburgh University Press.

MELLET, S. 2002. Corpus et recherches linguistiques : introduction, *Corpus* 1 : 5-13.

MILLER, C. R. 1984. Genre as social action. *Quarterly Journal of Speech 70*: 151-67; reprinted in A. Freedman, P. Medway (eds.); 1994: 23-42.

MITHUN, M. 1992. Is basic word order universal? In D. Payne (ed.). *Pragmatics of Word Order Flexibility*. Amsterdam/ Philadelphia: John Benjamins.

MOLINIER, C. 1996. Constructions en c'est : une classification générale. *Cahiers de Grammaire* 21 : 75-94.

MONDADA, L. 1994, *Verbalisation de l'espace et fabrication du savoir. Approche linguistique de la construction des objets de discours*. Thèse de doctorat, Université de Lausanne.

NWOGU, K.N. 1990. Discourse variation in medical texts: schema, theme and cohesion in professional and journalistic accounts, *Monographs in Systemic Linguistics*, vol. 2, University of Nottingham.

NWOGU, K.N., BLOOR, T. 1991. Thematic profession in professional and medical texts. In E. Ventola (ed.), *Functional and Systemic Linguistics: Approaches and Uses*. Berlin: Mouton de Gruyter: 369-384.

PAILLARD, M. 1989. Les chemins de la concession : quelques contrastes entre anglais et français. *Travaux linguistiques du Cerlico* 6, *Subordination*. Presses universitaires de Rennes.

PERY-WOODLEY, M-P. 1993. *Les écrits dans l'apprentissage*. Paris : Hachette.

PERY-WOODLEY, M-P. 1995. Quels corpus pour quels traitements automatiques ? *TAL* 36 (1-2) : 213-232.

PERY-WOODLEY, M-P. 2000. Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle. Mémoire d'HDR, *Carnets de grammaire* 8 (juillet). Université de Toulouse-Le Mirail : ERSS.

PERY-WOODLEY, M-P. 2001. Modes d'organisation et de signalisation dans des textes procéduraux. *Langages* 141 : 28-46.

PERY-WOODLEY, M-P. 2005. Discours, corpus, traitements automatiques. In A. Condamines (éd.) *Sémantique et corpus*. Paris : Hermès : 177-210.

PETIT, M. 1994. Le fond et le forme. *ASp,* la revue du Geras 5/6 : 1-10.

PETÖFI, P.S. 1973. Text-grammars, Text-theory and the theory of literature. *Poetics* 7: 37-76.

POHRIEL, S. 2005 Les marqueurs de thématisation : des thèmes phrastiques et textuels. *Travaux de linguistique* 2, 51 : 59-88.

PREVOST, S. 1998. La notion de thème : flou terminologique et conceptuel. *Cahiers de Praxématique* 30 : 13-35.

PREVOST, S. 2003a. Les compléments spatiaux : du topique au focus en passant par les cadres. *Travaux de Linguistique* 47 : 51-68.

PREVOST, S. 2003b. Détachement et topicalisation : des niveaux d'analyse différents, *Cahiers de Praxématique* 40 : 97-126.

PREVOST, S. 2005. Exploitation d'un corpus de français médiéval : enjeux, spécificités et apports, in A. Condamines (éd) *Sémantique et corpus*. Paris : Hermès/Lavoisier (Série « Traité IC2 » ; Cognition et traitement de l'information) : 147-176.

PRINCE, E.F. 1981. Towards a taxonomy of Given-New information. In P. Cole (ed.), *Radical Pragmatics*. New York: Academic Press : 223-255.

QUIRK, R., GREENBAUM, S., LEECH, G., SVARTVIK, J. 1985. *A Comprehensive Grammar of the English Language.* London: Longman.

RANGER, G. 1998. *Les constructions concessives en anglais : une approche énonciative*, Paris, Ophrys, coll. Cahiers de recherche.

RASTIER, F. 1989. *Sens et Textualité*. Paris : Hachette.

RASTIER, F. 2005. Enjeux épistémologiques de la linguistique du corpus. In G. Williams (ed.), *La Linguistique du corpus,* Presses universitaires de Rennes : 31-46.

REBOUL, A., MOESCHLER, J. 1998. *La pragmatique aujourd'hui. Une nouvelle science de la communication*. Paris : Seuil.

REINHART, T. 1980. Conditions for text coherence. *Poetics Today* vol. 1: 161-180

RESCHE, C. 2001. Réflexions sur la frontière entre langue générale et langue spécialisée. In Mémet, M. Petit (dir.) *L'Anglais de spécialité en France, Mélanges en l'honneur de Michel Perrin.* Collection ASp, Geras Editeur, Université Victor-Segalen/Bordeaux 2 : 37-46.

ROWLEY-JOLIVET, E. 1998. *La communication scientifique orale. Etude des caractéristiques linguistiques et discursives d'un genre*. Thèse de doctorat. Université de Bordeaux 2.

ROWLEY-JOLIVET, E. 2002. Visual discourse in scientific conference papers. A genre-based study. *English for Specific Purposes* 21(1): 19-40.

RUCK, H. 1980. *Linguistique textuelle et enseignement du français*. Paris : Crédif, Hatier.

SALAGER-MEYER, F. 1994. Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes* 13/2: 149-170

SARDA, L. 2008. Quantitative analysis of the meaning of prepositional adverbials according to their syntactic position: an insight into the degree of grammaticalization of prepositions. Paper given at *NRG4 – New Reflections on Grammaticalization*. Leuven.

SEIDLHOFER, B., WIDDOWSON, H. 1999. Coherence in summary: The contexts of appropriate discourse. In W. Bublitz *et al.* (eds.), *Coherence in Spoken and Written Disourse. How to Create it and How to Describe it.* Philadelphia: John Benjamins Publishing Company: 205-219.

SGALL, P. 1974. On the nature of topic and focus. In *Style and Text: Studies presented to Nils Enkvist*: 409-415.

SINCLAIR, J. 1996. *EAGLES, Preliminary Recommendations on Corpus Typology.* http://www.ilc.cnr.it/EAGLES96/corpustyp/node5.html

SLAKTA, D. 1975. L'ordre du texte. *Études de linguistique appliquée* 19 : 30-42.

SPANKIE, G.M. 1983. *English in Use*. Walton-on-Thames: Nelson.

SWALES, J.M. 1981. Aspects of article introductions. *ESP Research Reports* 1. Birmingham: University of Aston, Language Studies Unit.

SWALES, J.M. 1990. *Genre Analysis.* Cambridge: Cambridge University Press.

SWALES, J.M. 2004. *Research Genres*. Cambridge: Cambridge University Press.

SWEESTER, E. 1990. *From Etymology to Pragmatics*. Cambridge: Cambridge University Press

TABOADA, M. 1995. *Theme Markedness in English and Spanish: A Systemic-Functional Approach*. Manuscript. Universidad Complutense de Madrid.

TARONE, E., DWYER, S., GILLETTE, S., ICKE, V. 1998. On the use of the passive and active voice in astrophysics journal papers: with extensions to other languages and other fields *English for Specific Purposes 17/1*: 113-132.

TERSIS, N. 2000. Economie structurelle et complexité syntagmatique, Groenland oriental. In N .Tersis, M. Therrien (éds.), *Les langues eskaléoutes* : 249-268.

TEUFEL S., MOENS, M. 1999. Discourse-level argumentation in scientific articles: human and automatic annotation. In *Towards Standards and Tools for Discourse Tagging*. ACL. Workshop: 84-93.

THOMPSON, G. 1996. *Introducing Functional Grammar*. London: Arnold.

THOMPSON, S.E. 1994. Frameworks and contexts: A genre-based approach to analysing lecture introductions. *English for Specific Purposes* 13(2): 171-186.

TODOROV, T. 1978. *Les genres du discours*. Paris : Seuil.

TOGNINI-BONELLI, E. 2001. *Corpus Linguistics at Work*. John Benjamins.

TRAVNICEK, F. 1962. O Tak Zvanim Aktualnim Cleneni Vetním [On so-called functional sentence perspective]. Slovo a slovesnost (Prague) 22 (1962): 163-171. Traduit in M. Vasconcellos, *Theme and Focus: Cross-Language Comparison via Translations from Extended Discourse.* Unpublished Doctoral Dissertation. Georgetown University, 1985.

TURK C., KIRKMAN, J. 1982. *Effective Writing*. London: Spon.

VAN DIJK, T. A. 1972. *Some Aspects of Text Grammars*. The Hague: Mouton

VANDE KOPPLE, W.J. 1986. Given and new information and some aspects of the structures, semantics and pragmatics of written texts. In Cooper, Greenbaum (eds.), *Studying Writing: Linguistic Approaches*. Beverley Hills, CA.: Sage.

VANDE KOPPLE, W.J. 1994. Some characteristics and functions of grammatical subjects in scientific discourse, *Written Communication*. Sage Publications Inc: 534-564.

VASSILEVA, I. 2002. Speaker-audience interaction: the case of Bulgarians presenting in English. In E. Ventola *et al.*, *The Language of Conferencing*. Frankfurt am Main: Peter Lang: 255-276.

VENTOLA, E. 1994. Finnish writers academic English. Problems with reference and theme. *Functions of Language*, 1, 2: 1-33.

VENTOLA, E., MAURANEN, A. 1991. Non-native writing and native revising of scientific articles. In E. Ventola (ed.), *Functional and systemic linguistics: approaches and uses*. Berlin: Mouton de Gruyter: 457-490.

VIGNER, G. 1979. *Lire : du texte au sens.* Didactique des Langues étrangères. Paris : CLE international.

VIRTANEN, T. 2004. Point of departure: Cognitive aspects of sentence-initial adverbials. T. Virtanen (ed.) *Approaches to Cognition through Text and Discourse*. (Trends in Linguistics. Studies and Monographs 147.) Berlin/New York: Mouton de Gruyter: 79-97.

WEIL, H. 1991. *De l'ordre de mots dans les langues anciennes comparées aux langues modernes.* Paris : Didier érudition. (Thèse de lettres, 1844 ; reproduction en fac. Sim. de la 3e éd. de Paris, F. Vieweg, 1879).

WEINERT, R., MILLER, J. 1996. Cleft constructions in spoken language. *Journal of Pragmatics 25*: 173-206.

Williams, G., Burnard, L. 2005. À La Recherche de SARA : L'exploration d'un corpus scientifique balisé selon la TEI avec le logiciel SARA. In G. Williams (Ed.), *La linguistique de corpus*. Presses universitaires de Rennes : 247-258.