

TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. B OSA - TOM. 337

HUMANIORA

D'abord, ensuite, enfin et 0, De plus :
**Organisation textuelle par des séries linéaires
dans les articles de recherche**

Veronika Laippala

TURUN YLIOPISTO
UNIVERSITY OF TURKU
Turku 2011

Département d'études françaises
Université de Turku
Finlande

Thèse dirigée par
Professeur Eija Suomela-Salmi
Département d'études françaises
Université de Turku
Finlande

Pré-rapporteurs:

Professeur Shirley Carter-Thomas
Département langues et sciences humaines
Télécom Ecole de Management / Institut Télécom
Evry, France

Professeur Catherine Schnedecker
UFR Lettres
Université de Strasbourg
France

Rapporteurs critiques:

Professeur Marie-Paule Péry-Woodley
Département de Sciences du Langage
Université de Toulouse – Le Mirail
France

Professeur Catherine Schnedecker
UFR Lettres
Université de Strasbourg
France

ISBN 978-951-29-4788-1 (PRINT)

ISBN 978-951-29- 4790-4 (PDF)

ISSN 0082-6987

Turun yliopisto: Painosalama

A mon père Pekka

AVANT-PROPOS

Je n'aurais jamais pu accomplir cette thèse sans l'aide et le soutien de nombreuses personnes. Tout d'abord, je tiens à remercier mon directeur de thèse, Eija Suomela-Salmi, de m'avoir soutenue tout au long de ce travail et d'avoir toujours trouvé le temps de lire mes manuscrits malgré les délais parfois stricts. Deuxièmement, je remercie mes pré-rapporteurs, Shirley Carter-Thomas (Télécom Ecole de Management) et Catherine Schnedecker (Université de Strasbourg), pour leurs commentaires très pertinents qui m'ont encore aidée à améliorer ce texte. De plus, j'exprime ma gratitude à Juhani Reiman (Lingsoft Inc.) et à Tapio Salakoski (Université de Turku / Sciences informatiques) d'avoir organisé la révision linguistique de cette thèse.

Pendant la préparation de cette thèse, j'ai eu le privilège de travailler avec des gens et équipes de recherche formidables. Mes remerciements vont à Lotta, Milla et Kaiju du Département d'études françaises pour leur amitié et pour nos discussions sur la linguistique, la recherche et la vie... Vous avez rendu le travail quotidien agréable et c'est grâce à vous que j'ai trouvé la patience d'aboutir cette tâche qui semble souvent, comme on le sait, interminable et désespérée. Je veux aussi remercier Ana Maria, Fred, Maarit, Marjut et Mirka de m'avoir toujours aidée et encouragée pendant ces années. J'exprime ma gratitude également à Anne, Maija et Marja-Berit, sans qui j'aurais oublié de manger, d'aller à la maison le soir et de remplir Sole...

La vie est pleine de surprises et de coïncidences, et c'est ainsi que j'ai rencontré mes collègues des Départements de sciences informatiques et de sciences infirmières ainsi que du projet IKITIK. Même si le français, les sciences informatiques et les sciences infirmières semblent avoir peu en commun au premier regard, cette collaboration a influencé énormément ma recherche. Je veux surtout remercier Filip, sans qui l'annotation XML de mon corpus et ainsi toute mon étude n'aurait pas été possible. Grâce à lui et sa patience, mon idée optimiste et risquée de baser toute mon étude et mes analyses sur un corpus annoté et des programmes python s'est réalisée. Mes remerciements vont également (dans l'ordre alphabétique) à Hanna, Heljä, Jenna, Katri, Kirsi, Riitta, Sampo, Samuel, Sanna, Tapio et Timo. Le travail que nous avons fait ensemble m'a été une source d'inspiration extraordinaire et m'a beaucoup apportée, que ce soit sur la définition du participe passé en finnois, sur les scripts Unix ou sur le sens de Norri (non, ce n'est pas un docteur...).

Une autre source d'inspiration et d'encouragement importante m'a été l'école doctorale nationale de linguistique Langnet. Cette communauté m'a permis de travailler dans une atmosphère à la fois professionnelle et conviviale et m'a beaucoup appris du travail de chercheur, de la linguistique ainsi que de la recherche en général. Je tiens à remercier Marja-Liisa Helasvuo et Anna Mauranen pour leur encouragement, aide, commentaires et confiance pendant ces années. Grâce à Langnet et grâce à vous, j'ai compris que l'accomplissement de la thèse est en fait possible et non seulement un but imaginaire, et que moi aussi je pourrais le faire. Aussi nos discussions inoubliables avec Anna O., Lotta

et Päivi, entre autres sur les notions de texte, discours et genre discursif, m'ont donné la force d'accomplir cette thèse et m'ont aidée à mieux comprendre ce que je suis en train de faire (même si *le genre* m'est encore un peu obscur...). De même, mes remerciements vont aux séminaires organisés par l'école doctorale Utuling de l'Université de Turku et Spremi de Åbo Akademi dans lesquels j'ai eu la possibilité de présenter mes idées et travaux en cours.

Encore, une partie importante de ce travail n'aurait pas été possible sans l'aide et la contribution de Stefan Werner (Université de Joensuu) sur les calculs statistiques nécessaires dans ce travail. Merci, Stefan, pour ton aide ! Je veux également remercier Langnet, la Fondation Finlandaise pour la Culture et la Fondation de l'Université de Turku d'avoir financé mes études et mes voyages aux colloques.

Heureusement, et malgré mes heures de travail, j'ai la chance d'avoir une vie également à l'extérieur de l'Université. Je veux ainsi remercier tous mes amis de ne pas m'avoir oubliée même quand j'ai passé tout mon temps au bureau et de m'avoir écoutée même quand je ne parlais que de l'organisation textuelle par des séries linéaires... Vous êtes trop nombreux pour être énumérés, mais vous savez qui vous êtes =)

Enfin, mes remerciements vont à ma famille, ma mère Leena, ma sœur Aleksandra et ma tante Riikka. Merci pour votre compréhension, soutien, aide et patience.... et pour votre optimisme (souvent exagéré, je trouve) pour la réussite de ce projet.

Pour terminer, les mots ne suffisent pour exprimer ma gratitude à Matti qui n'a pas cessé de m'aimer malgré mes heures de travail, absences (physique et mentale) et humeurs, et qui me comprend même quand je ne me comprends pas moi-même. Aussi je tiens à remercier le petit Martti d'avoir enrichi les derniers pas de ce travail et de m'avoir laissée achever ce projet au temps prévu.

<u>LISTE DES ABRÉVIATIONS.....</u>	<u>11</u>
---	------------------

<u>LISTE DES TABLEAUX ET DES FIGURES.....</u>	<u>13</u>
--	------------------

<u>1. INTRODUCTION.....</u>	<u>15</u>
------------------------------------	------------------

1.1 POUR UNE ORGANISATION TEXTUELLE PAR DES SÉRIES LINÉAIRES	15
---	-----------

1.2 OBJECTIFS	17
----------------------------	-----------

1.3 QUESTIONS DE RECHERCHE ET HYPOTHÈSES	19
---	-----------

<u>PARTIE THÉORIQUE</u>	<u>22</u>
--------------------------------------	------------------

<u>2. LINGUISTIQUE DE CORPUS ET LE DISCOURS DE RECHERCHE.....</u>	<u>22</u>
--	------------------

2.1 LINGUISTIQUE DE CORPUS.....	22
--	-----------

2.2 DISCOURS DE RECHERCHE.....	24
---------------------------------------	-----------

2.2.1 POURQUOI LE DISCOURS DE RECHERCHE ?	24
---	----

2.2.2 MÉTADISCOURS ET SÉRIES LINÉAIRES	26
--	----

2.2.3 ARTICLES DE RECHERCHE ET VARIATION INTERDISCIPLINAIRE ET INTERLINGUISTIQUE	28
--	----

2.2.3.1 Articles de recherche en français.....	29
--	----

2.2.3.2 Articles de recherche en sciences naturelles et en sciences humaines et sociales ...	30
--	----

2.2.3.3 Articles de recherche en linguistique.....	32
--	----

2.2.3.4 Articles de recherche en éducation.....	33
---	----

2.2.3.5 Articles de recherche en histoire	34
---	----

2.2.3.6 Bilan : articles de recherche.....	35
--	----

<u>3. FORMATION DU TEXTE</u>	<u>37</u>
---	------------------

3.1 COHÉRENCE.....	37
---------------------------	-----------

3.2 TEXTE VS. DISCOURS	41
-------------------------------------	-----------

3.3 COHÉSION	44
---------------------------	-----------

3.4 INDICES DE COHÉSION.....	45
-------------------------------------	-----------

3.4.1 CONJONCTION, CONNECTEURS ET ORGANISATEURS TEXTUELS	45
--	----

3.4.2 RÉFÉRENCE ET ANAPHORE.....	48
----------------------------------	----

3.4.3 COHÉSION LEXICALE	53
-------------------------------	----

3.4.4 ENCADREMENT DU DISCOURS.....	55
------------------------------------	----

<u>4. ORGANISATION TEXTUELLE.....</u>	<u>58</u>
--	------------------

4.1 SÉQUENTIALITÉ DU DISCOURS (GOUTSOS 1996)	59
4.1.1 SCHÉMA DE GOUTSOS : STRATÉGIES, TECHNIQUES ET ZONES.....	60
4.1.2 TECHNIQUES DE DÉPLACEMENT ET DE CONTINUATION.....	61
4.1.2.1 Techniques de déplacement.....	62
4.1.2.2 Technique de continuation.....	64
4.1.3 CORRESPONDANCES ET CONFLITS ENTRE LE MODÈLE DE GOUTSOS ET LES TRAVAUX SUR LA COHÉRENCE.....	65
4.2 ORGANISATION TEXTUELLE PAR DES SÉRIES LINÉAIRES	67
4.2.1 SÉRIES LINÉAIRES ET LEURS MARQUEURS DANS DES TRAVAUX ANTÉRIEURS.....	67
4.2.2 POUR UNE DÉFINITION DES SÉRIE LINÉAIRES : QUELLES FONCTIONS ET CARACTÉRISTIQUES PRENDRE EN COMPTE ?.....	76
4.2.3 TRAITS DE SURFACE COMME SIGNES DE SÉRIE LINÉAIRE – ATOUTS ET PROBLÈMES ..	79
4.2.4 AUTRES COMPOSANTS D’UNE SÉRIE LINÉAIRE : AMORCE ET CLÔTURE, ÉQUIVALENCE ENTRE LES ITEMS.....	81
4.3 BILAN : ORGANISATION TEXTUELLE	83
4.3.1 NOTRE CONCEPTION DE CERTAINES NOTIONS CENTRALES DANS L’ORGANISATION TEXTUELLE.....	84
4.3.2 NOTRE CONCEPTION DES SÉRIES LINÉAIRES.....	86

5. CORPUS ET MÉTHODE.....89

5.1 CORPUS	89
5.1.1 CORPUS ET CORPUS. QUE SONT-ILS ? COMMENT SONT-ILS ?.....	89
5.1.2 CORPUS DE L’ÉTUDE : CRITÈRES DE COLLECTE ET TAILLE FINALE.....	91
5.2 MÉTHODE	95
5.2.1 DÉMARCHE PRATIQUE.....	95
5.2.2 SCHÉMA D’ANNOTATION.....	100
5.2.2.1 Quelques critères de classification des marqueurs d’items individuels.....	100
5.2.2.2 Classes des composants des séries linéaires.....	102
5.2.2.3 Critères d’annotation des séries linéaires simples.....	106
5.2.2.4 Critères d’annotation des séries enchaînées.....	107
5.2.3 QUELQUES REMARQUES FINALES SUR LA MÉTHODE ET LES CRITÈRES D’ANNOTATION.....	108

PARTIE EMPIRIQUE.....110

6. SÉRIES LINÉAIRES DANS LE CORPUS : FRÉQUENCES, LONGUEURS ET MARQUAGES.....110

6.1 FRÉQUENCES ET FONCTIONS SPÉCIFIQUES	110
6.2 LONGUEURS	120
6.3 MARQUAGES	126
6.3.1 SÉRIES SIMPLES.....	127

6.3.2 SÉRIES ENCHAÎNÉES.....	136
6.3.3 AMORCES ET CLÔTURES	141
6.4 BILAN : SÉRIES LINÉAIRES, SÉRIES VARIÉES	145

7. MARQUAGE ET NON MARQUAGE DES SÉRIES LINÉAIRES : FACTEURS TEXTUELS EXPLICATIFS.....149

7.1 POUR UN GUIDAGE ADÉQUAT DU LECTEUR : LE DEGRÉ D'EXACTITUDE DU MARQUAGE	150
7.1.1 AMORCES ET CLÔTURES	151
7.1.2 MARQUEURS D'ITEM.....	158
7.1.3 BILAN : STRATÉGIES DE MARQUAGE D'ORGANISATION TEXTUELLE EXACT ET VAGUE	162
7.2 LONGUEUR ET MARQUAGE DES SÉRIES LINÉAIRES.....	163
7.2.1 LONGUE ET IMPRÉCISE.....	164
7.2.2 ... ET COURTE ET EXACTE	168
7.2.3 MARQUEURS GLOBAUX ET LOCAUX	172
7.2.4 SÉRIES SECTIONS ET SÉRIES MÉTA – SEGMENTATION DU TEXTE PAR LES TITRES.....	183
7.2.5 BILAN : MARQUAGE DES SÉRIES LINÉAIRES DE LONGUEURS DIFFÉRENTES	190
7.3 IDENTIFICATION DES ITEMS NON MARQUÉS ET INDICES COMPLÉMENTAIRES.....	192
7.3.1 RÉPÉTITION LEXICALE	194
7.3.2 CORRESPONDANCE DES ITEMS AU PRINCIPE D'ÉNUMÉRATION ET CONTIGUÏTÉ SÉMANTIQUE.....	197
7.3.3 AUTRES INDICES COMPLÉMENTAIRES.....	201
7.3.4 FRÉQUENCES DES INDICES COMPLÉMENTAIRES DANS LE CORPUS	203
7.3.5 INDICES COMPLÉMENTAIRES ET TITRES.....	212
7.3.6 BILAN : INDICES COMPLÉMENTAIRES DE SÉRIATION, FACTEURS DE NON MARQUAGE ?	214
7.4 BILAN : FACTEURS DU MARQUAGE ET DU NON MARQUAGE DES SÉRIES LINÉAIRES	216

8. SÉRIES LINÉAIRES ET VARIATION INTERDISCIPLINAIRE219

8.1 FRÉQUENCES	220
8.2 LONGUEURS	224
8.3 MARQUAGES	229
8.3.1 CATÉGORIES DE SÉRIES DANS LES DISCIPLINES ÉTUDIÉES	230
8.3.2 AMORCES ET CLÔTURES	233
8.4 VARIATION INTERCONTINENTALE ?	237
8.5 BILAN : VARIATION INTERDISCIPLINAIRE ET VARIATION INTERCONTINENTALE .241	

9. POUR CONCLURE.....244

9.1 QUESTIONS, HYPOTHÈSES ET RÉPONSES	244
9.2 DISCUSSION.....	250
<u>RÉFÉRENCES</u>	<u>253</u>
<u>ANNEXE 1 LISTE DES ARTICLES DE RECHERCHE DU CORPUS</u>	<u>264</u>
<u>ANNEXE 2 EXEMPLES DES SOUS CORPUS.....</u>	<u>273</u>
<u>INDEXE</u>	<u>292</u>

LISTE DES ABREVIATIONS

A	Expression anaphorique
AC	Série signalée par la combinaison d'une expression anaphorique et d'un connecteur.
ACL	Série signalée par la combinaison d'une expression anaphorique, d'un connecteur et d'une construction numéro+nom.
AL	Série signalée par la combinaison d'une expression anaphorique et d'une construction numéro+nom.
C	Connecteur
CL	Série signalée par la combinaison d'un connecteur et d'une construction numéro+nom. Construction Num+N Construction numéro+nom, du type <i>le premier exemple</i> ou <i>un autre travail</i> .
Equiv.	Equivalence structurelle des items.
HOMO	Série homogène, composée d'items signalés par le même type de marqueurs.
L	Construction numéro+nom
MIL	Marqueur d'intégration linéaire (cf. Turco et Coltier 1988)
MIX	Série mixte, composée d'items signalés par des marqueurs différents.
MIXAL	Série mixte, signalée par des expressions anaphoriques et par des constructions numéro+nom. Sous-type des séries MIX.
MIXREST	Série mixte, signalée par d'autres marqueurs que des expressions anaphoriques et des constructions numéro+nom.
N	Nombre d'occurrences présentées dans la figure ou tableau.
NM	Série composée d'items non marqués, dans laquelle le nombre d'items est signalé avant ou après par une amorce ou une clôture.
NMet	Série composée d'items non marqués et d'items signalés par des conjonctions de coordination. Le nombre d'items est signalé avant ou après par une amorce ou une clôture.
0=0	La série n'est pas précédée d'une amorce.
0=IMP	Amorce implicite.
0=NUM	Énumération exacte. L'amorce indique le nombre d'items de la série.
0=Q	Amorce-question.
0=VAG	Énumération vague. L'amorce indique que le texte à suivre est segmenté en items, mais n'en indique pas le nombre.
Para.	<i>La paraphrase</i> crée de la cohésion lexicale entre les items et est souvent employée dans les séries métadiscursives afin de présenter les items-sections suivants.
PE	Principe d'énumération
PNM	Série partiellement non marquée, composée d'items non marqués et explicitement signalés.

R=0	La série n'est pas suivie d'une clôture.
R=ACON	La clôture est signalée par un autre connecteur.
R=ANA	La clôture correspond à une reprise anaphorique.
R=CON	La clôture est signalée par un connecteur conclusif / explicatif.
R=NUM	La clôture définit le nombre d'items de la série.
Rep.	<i>La répétition proprement dite</i> correspond à la reprise de mots ou de syntagmes selon le même rôle syntaxique dans tous les items de la série.
Rep.dif.	<i>La répétition des éléments différents</i> établit un lien de répétition entre les items de la série mais seulement entre une partie d'eux.
Rep.par.	<i>La répétition partielle</i> correspond à la répétition de mots individuels dans les items, insuffisante seule pour leur identification.
RST	<i>Rhetorical Structure Theory</i> (Mann et Thompson 1988)
SDRT	<i>Segmented Discourse Representation Theory</i> , la Théorie des représentations discursives segmentées (Asher 1993 ; Asher & Lascarides 2003)
TSC	<i>Text strategic continuity</i> , continuité stratégique du texte.

LISTE DES TABLEAUX ET DES FIGURES

Tableau 1 Sous-corpus de l'étude.....	94
Tableau 2 Amorces et clôtures dans le corpus.	111
Tableau 3 Séries simples, enchaînées et à étiquettes particulières.	113
Tableau 4 Longueurs des séries en nombres de mots, d'items et de séries inférieures.	122
Tableau 5 Abréviations utilisées.....	142
Tableau 6 Sous-corpus examiné dans cette section.....	194
Figure 1 Modèle de Goutsos (1996 : 507).	60
Figure 2 Modèle de thème dérivé (Danes 1974 : 119).	75
Figure 3 Le logiciel Callisto.	96
Figure 4 Exemple de l'analyse <i>python</i> du XML annoté.....	99
Figure 5 Démarche pratique de notre méthode.....	99
Figure 6 Longueurs de toutes les séries linéaires du corpus.....	121
Figure 7 Proportions des catégories des séries linéaires simples.	131
Figure 8 Fréquences des marqueurs et des combinaisons de marqueurs dans les séries simples.	132
Figure 9 Fréquences des catégories de séries enchaînées.....	137
Figure 10 Proportions des catégories de séries simples dans les structures enchaînées.	138
Figure 11 Fréquences des combinaisons de types de marqueurs dans les structures enchaînées.....	140
Figure 12 Proportions des séries enchaînées dans lesquelles ces combinaisons de marqueurs sont présentes.	141
Figure 13 Amorces dans les séries simples et dans les séries enchaînées.	143
Figure 14 Clôtures dans les séries simples et dans les séries enchaînées.	144
Figure 15 Amorces dans les catégories de série.	151
Figure 16 Amorces et clôtures exactes des catégories de série.	153
Figure 17 Degré d'exactitude des marqueurs des séries simples.	159
Figure 18 Degré d'exactitude des marqueurs des séries enchaînées.	160
Figure 19 Longueurs des catégories de série, séries section et méta exclues.....	165
Figure 20 Variation de la longueur (médiane) selon le degré d'exactitude des marqueurs d'item.	166
Figure 21 Variation de la longueur (médiane) selon le degré d'exactitude des amorces et clôtures.....	166
Figure 22 Longueurs des séries NM et NM selon la présence d'une amorce et / ou clôture exacte.	171
Figure 23 Longueurs des items signalés par des marqueurs différents. Le groupe <i>connecteurs</i> inclut également les expressions cadratives.	173
Figure 24 Longueurs des items signalés par des marqueurs différents.	176
Figure 25 Longueurs des items introduits par les connecteurs les plus fréquents.....	178
Figure 26 Longueurs des séries section.....	185
Figure 27 Longueurs des séries méta.	185
Figure 28 Catégories des séries section.....	186
Figure 29 Catégories des séries méta. Les NM et NMet sont toujours combinées sous la catégorie des NM.....	186

Figure 30	Fréquences des marqueurs des séries section et méta (en pourcentage du total de la catégorie).	188
Figure 31	Variation du marquage des séries linéaires selon la longueur de la série.	191
Figure 32	Types de répétition dans les séries (en pourcentage des séries dans lesquelles ce type de répétition est présent). Les NMet sont toujours unies sous le titre NM.	205
Figure 33	Types de contiguïté sémantique des items (en pourcentage des séries dans lesquelles il est présent).	208
Figure 34	Fréquences des traits syntaxiques, de la correspondance des items aux exemples et des autres marqueurs en pourcentage du total.	211
Figure 35	Indices complémentaires dans les séries section et méta.	213
Figure 36	Fréquences des séries dans les sous-corpus, en moyenne du nombre de séries / 1 000 mots.	220
Figure 37	Fréquences des séries avec étiquettes particulières (en pourcentage du total). ..	222
Figure 38	Longueurs des séries des trois disciplines, séries section et méta exclues.	225
Figure 39	Séries section et méta dans les trois domaines. Les fréquences sont en pourcentages du nombre total des séries dans le sous-corpus.	226
Figure 40	Catégories de séries selon les trois disciplines examinées. Les fréquences sont en pourcentages du total.	230
Figure 41	Types d'amorces selon les disciplines dans toutes les séries du corpus en pourcentages du total.	233
Figure 42	Types de clôtures selon les disciplines dans toutes les séries du corpus en pourcentages du total.	234
Figure 43	Fréquences des types d'amorce et clôtures rares dans les domaines (en pourcentage du total).	236
Figure 44	Les séries simples (à gauche) et les enchaînées (à droite) dans les revues françaises et canadiennes dans le sous-corpus éducation.	238
Figure 45	Les séries simples (à gauche) et les enchaînées (à droite) dans les revues françaises et canadiennes dans le sous-corpus histoire.	239
Figure 46	Catégories de séries selon les deux types de revue en éducation.	240
Figure 47	Catégories de séries selon les deux types de revue en histoire.	240
Figure 48	Résumé des caractéristiques des domaines examinés.	242

1. INTRODUCTION

1.1 POUR UNE ORGANISATION TEXTUELLE PAR DES SERIES LINEAIRES

Cette thèse examine les façons dont les phrases individuelles forment un texte en tant qu'ensemble pertinent. Le point de départ de notre problématique est qu'une suite quelconque de phrases ne forme pas un texte et qu'un texte n'est pas une suite quelconque de phrases. Au contraire, un texte est un tout intelligible et organisé : les phrases s'y combinent afin de former un tout compréhensible pour le lecteur.

Contrairement à la phrase, dans laquelle la syntaxe détermine l'agencement des mots, aucune règle ne définit la façon dont les phrases devraient être organisées dans le texte. C'est la sémantique qui règne dans le texte : un texte est un texte s'il a du sens. Pour ce faire, c'est le lecteur qui doit, en s'appuyant sur ses connaissances, pouvoir comprendre la façon dont les phrases sont liées pour former un tout.

Le texte peut également contenir des indices signalant explicitement comment les phrases sont reliées. Ces indices guident le lecteur dans l'interprétation des liens entre les phrases afin qu'il n'ait pas à les déduire lui-même. Ces indices peuvent être, notamment, des connecteurs spécifiant la nature de la relation entre les phrases, *ensuite* par exemple, ou des constructions indiquant la place de la phrase dans une énumération, telle que *la deuxième raison*. Aussi, les expressions qui indiquent l'organisation des phrases à venir, comme *la section suivante consiste en trois parties*, fonctionnent comme des signaux de la façon dont les phrases doivent être interprétées. Un texte peut cependant être clair et bien organisé même sans ces indices explicites. Dans ce cas, c'est le lecteur qui doit travailler davantage afin de situer le texte dans le contexte souhaité par l'auteur.

L'organisation des phrases individuelles comme un tout intelligible ainsi que les façons dont cette organisation est marquée, ou non, sont au centre de cette thèse. Nous traiterons essentiellement les différents marqueurs utilisés pour indiquer cette organisation ainsi que les cas où les marqueurs explicites ne sont pas utilisés. De la sorte, nous chercherons à mieux comprendre ce qui fait un texte et à étudier pourquoi les marqueurs explicites ne sont pas toujours utilisés.

Un texte peut être organisé de plusieurs façons. Nous nous limiterons à l'étude de l'organisation pourvue par les *séries linéaires*. Les séries linéaires segmentent le texte en items qui se suivent en son sein et sont, au moins partiellement, ordonnés. L'ordre ou le nombre de ces items peut être indiqué devant un, voire tous les items qui, ensemble, forment une série linéaire dont la longueur varie de deux phrases à un texte intégral. Ci-dessous, l'Exemple 1 illustre une série linéaire segmentant le texte en trois items, tous introduits par des marqueurs linguistiques signalant l'ordre : le connecteur *tout d'abord*, la construction *le deuxième exemple* et la combinaison du connecteur *enfin* et de la mention *dernier exemple* indiquant qu'il s'agit du dernier item de la liste. De plus, dans cet

exemple, la série est également précédée par une *amorçe* indiquant déjà à l'avance au lecteur le nombre d'items de la série.

Exemple 1 Les parties en gras et les numéros devant les composants sont de nous.

0 Illustrons ceci, à partir de **trois exemples**.

1 **Tout d'abord**, la quantification nous a permis de comparer les différents sous-corpus du point de vue de deux des principales « étiquettes » de significations que nous avons attribuées au Npr Outreau entre la première et la dernière période. [...]

2 **Le deuxième exemple**, déjà abordé sous « l'angle d'attaque » 3 du § 3.3 (scénario et actants), se rapporte aux rangs de fréquences des mots des sous-corpus comparés. [...]

3 **Enfin – dernier exemple** –, nous avons eu l'occasion de signaler, dans le cadre de l'angle d'analyse 2 (se rapportant aux réseaux de Npr), que la référence à Dutroux comme comparant était révélatrice du cadre de référence de Outreau comme fait public. [...]

Lecolle1

Dans les séries linéaires objets de notre thèse, les items sont, en tout ou partie, marqués par des moyens lexicaux signalant énumération, addition ou progression. Prototypiquement, chaque item est introduit par *un marqueur d'intégration linéaire* (désormais MIL, cf. Turco & Coltier 1988) : *Premièrement... Deuxièmement... Enfin...* De plus, les séries linéaires sont souvent définies comme comprenant des items équivalents ou parallèles. Jackiewicz (2005 : 95) constate, par exemple, que « La sériation [...] est un procédé [...] permettant à la fois d'ordonner et de situer sur un axe d'équivalence paradigmatique des portions de texte ». Dans la pratique, pourtant, Virbel (1999) et Péry-Woodley (2000 : 138-139) ont noté que les items peuvent être très hétérogènes et marqués de façons variées.

Outre les MIL, les items de la série peuvent aussi être introduits, par exemple, par les anaphoriques, les constructions numéro+nom et les verbes : *Le premier... Le deuxième exemple... Terminons par...* De plus, la sériation peut être signalée avant la série par une amorçe marquant le début d'une série et, possiblement, le nombre de ses items (cf. l'Exemple 1), ou après la série par une *clôture* indiquant la fin de la série et, possiblement, le nombre d'items de la série : *Ces deux parties... / Donc...* (cf. Tadros 1994 ; Jackiewicz 2005). Enfin, la série peut également inclure des items *non marqués* qui ne sont pas introduits par un marqueur indiquant leur ajout ou leur place dans la série. Ces items peuvent être identifiables, par exemple, grâce à la répétition de mots identiques au début de chaque item ou d'autres indices contextuels. L'Exemple 2 ci-dessous illustre le non marquage. Le premier item est, selon nos critères, non marqué car il n'est pas signalé par un marqueur d'addition ou d'ordre. Nous pouvons toutefois identifier l'item grâce à l'amorçe *quelques départements* et aux expressions spatiales devant chacun des deux items. Indiquant au lecteur la fin de la série, la clôture *dans ces départements* facilite encore la lecture.

¹ Nous indiquons à la fin de chaque exemple le nom du premier auteur de l'article dont l'extrait est tiré. Les informations bibliographiques plus précises se trouvent dans l'annexe.

Exemple 2

- 0 **De nombreux départements** actifs dans le pétitionnement répondent à ce modèle gardois.
 - 1 Dans la Sarthe, 11 800 signatures sont récoltées dans 90 communes, dont seulement 16 % au Mans.
 - 2 Dans le Bas-Rhin **également**, seuls 22 % des 12 252 signatures ont été recueillies à Strasbourg 62.
- Cl. **Dans ces départements**, le pétitionnement se caractérise par la prépondérance de l'initiative villageoise et la dissémination rurale.

Jarrige

1.2 OBJECTIFS

Cette thèse a pour objectif l'étude de l'organisation du texte par des séries linéaires. Ce sont surtout les modes de marquage de cette organisation qui nous intéressent, en particulier la variation des marqueurs utilisés et l'alternance de marquage explicite et d'absence de marquage. Comment les séries linéaires contribuent-elles à l'organisation du texte? Pourquoi sont-elles introduites de la sorte ?

Le marquage de l'organisation textuelle est souvent étudié à travers des marqueurs individuels, tels que *premièrement*, *mais* ou *enfin*. Pourtant, comme le dit Halliday (1977-2003 : 45), c'est uniquement ensemble que les indices de cohésion présents dans le texte facilitent la formation d'une entité intelligible (cf. également Péry-Woodley et Scott 2006 : 12-13 ; Ho Dac et Péry-Woodley 2009 et les résultats rapportés dans Ho Dac 2007). Ainsi, afin d'examiner l'organisation textuelle dans son ensemble, tout type d'indices linguistiques devrait être pris en compte. Cette approche assure une étude plus réaliste de la façon dont les textes sont organisés et dont les différents types de marqueurs fonctionnent ensemble.

Un texte organisé peut donc être perçu comme le produit d'une combinaison d'indices. Péry-Woodley (2000 : 9, 134-135) propose une façon d'observer le texte plus globalement : prendre comme point de départ de l'analyse une fonction discursive, soit une relation qui fonctionne dans l'organisation du texte, et non un marqueur préalablement choisi, susceptible d'être signe de cette fonction. Péry-Woodley constate (*ibid.*) également que l'étude des marqueurs individuels reste insuffisante. Ils ne sont, en effet, pas indispensables pour l'établissement de la fonction : par exemple, une relation de conséquence entre deux phrases n'a pas besoin de marqueur consécutif pour exister. En outre, les marqueurs lexicaux peuvent être plurifonctionnels et par conséquent, leur fonction n'est pas nécessairement figée à leur forme. Cette approche proposée par Péry-Woodley (*ibid.*) - prendre comme point de départ une fonction discursive - est adoptée, entre autres, dans la thèse de Ho Dac (2007). Elle examine les fonctions de la place initiale de la phrase dans l'organisation du discours en tenant compte de tous les types d'indices de surface, connecteurs, adverbiaux et constructions impersonnelles inclus. Notre thèse s'efforce d'adopter une approche similaire : nous examinons les séries linéaires introduites

par tout type de marqueurs repérés dans le texte, fonctionnant comme introducteur d'item d'une série. Ce faisant, nous cherchons, d'une part, à compléter les travaux antérieurs relatifs aux différents types de marqueurs et, d'autre part, à acquérir une vision plus englobante de l'organisation textuelle.

Notre choix - prendre comme point de départ de l'analyse la fonction discursive des séries linéaires et non une forme spécifique - nous a également conduite à prendre en compte dans l'analyse les items non marqués des séries linéaires. Bien que cette résolution risque de soulever des difficultés d'identification, elle assure une étude plus complète de l'organisation textuelle et de son marquage. De plus, dans les études relatives à l'organisation textuelle, les raisons textuelles sous-jacentes du non marquage restent encore à clarifier. Ainsi, Jackiewicz (2005 : 107-109) se contente de noter que les items des séries linéaires peuvent également être non marqués, sans marqueurs explicites. Dans le cadre d'un modèle structural de l'organisation du texte, la Théorie des structures rhétorique (RST – Rhetorical Structure Theory, Mann et Thompson 1988), l'organisation textuelle est décrite *via* la définition de toutes les relations, marquées ou non marquées, entre les phrases et les segments supérieurs. En s'appuyant sur cette même théorie, Taboada (2006) a également calculé des proportions de relations signalées et non signalées. Nous nous efforçons de mener l'analyse plus loin en recherchant des raisons textuelles sous-jacentes du non marquage. Pourquoi tous les items des séries linéaires ne sont-ils pas marqués explicitement ?

Une autre restriction, souvent présente dans des travaux sur l'organisation textuelle, est leur manque de corpus : ils invoquent fréquemment des exemples inventés ou des extraits en nombre limité. Par conséquent, et bien qu'ils fournissent une description adéquate des fonctions possibles, d'un marqueur ou d'un type de marqueur par exemple, ils ne peuvent que difficilement avancer des résultats généralisables sur **l'usage** de ces marqueurs dans le texte.

Notre étude porte sur cet aspect moins considéré de l'organisation textuelle et, plus particulièrement, sur l'usage pratique des séries linéaires. Pour ce faire, nous les examinons dans un corpus et, suivant Biber et al (1998: 1 et passim), « au lieu d'examiner ce qui est en théorie possible, étudions l'usage de la langue dans son contexte naturel » (notre traduction). Ce faisant, nous cherchons, encore, à acquérir une vision plus globale de l'organisation textuelle en tenant compte de la pratique et du contexte naturel de la langue.

Enfin, l'organisation textuelle et son marquage dépendent du contexte et de la situation dans lesquels le texte est rédigé. Un roman policier, dans lequel le lecteur ne doit pas tout savoir dès le début, n'est pas organisé de la même façon, par exemple, qu'un article dans le journal *Le Monde* dont l'intention est d'informer le lecteur. Notre corpus est composé d'articles de recherche dont le but est non seulement d'informer mais également de persuader le lecteur d'accepter leur contenu (cf. section 2.2). L'organisation textuelle joue un rôle important dans ces articles car un texte mal organisé risque d'être peu convaincant.

De plus, comme les disciplines scientifiques se différencient les unes des autres selon divers aspects, l'organisation textuelle dans les articles de recherche varie également selon la discipline. Notre corpus comprend des articles appartenant à trois disciplines : la linguistique, l'histoire et l'éducation. Comme nous l'exposerons plus tard en section 5.1.2, cette répartition nous permet d'examiner les séries linéaires dans un corpus davantage représentatif tout en comparant différentes disciplines.

1.3 QUESTIONS DE RECHERCHE ET HYPOTHESES

Cette thèse se propose d'observer l'organisation textuelle à l'aide des séries linéaires ainsi que la variation des marqueurs et le non marquage constatés dans ces structures. L'approche adoptée tend à examiner l'usage des séries linéaires dans un corpus d'articles de recherche de trois disciplines en sciences humaines et sociales. Nous considérons l'organisation textuelle comme le produit de l'usage de différents types de marqueurs et du non marquage.

Plus spécifiquement, notre problématique est donc double : d'une part, nous examinons l'usage et le fonctionnement des séries linéaires dans l'organisation textuelle ; d'autre part, nous abordons la variation des marqueurs utilisés pour signaler cette organisation, de même que l'alternance du marquage explicite et de l'absence de marquage. Nous recherchons de ce fait une perception globale des différents marqueurs et du rôle du non-marquage. Ces questions sont complétées par l'aspect comparatif entre les différentes disciplines scientifiques examinées.

Les questions de recherche auxquelles nous visons à répondre sont :

1. **Comment les séries linéaires participent-elles à l'organisation textuelle ?**
 - Quels sont les types de segments textuels organisés en séries linéaires ?
 - Quelles formes (structures hiérarchiques) les séries linéaires prennent-elles dans le texte ?
 - Comment les séries linéaires sont-elles introduites dans le texte ? Par quels types de marqueurs ? Le non marquage est-il fréquent ?
2. **Pourquoi les séries linéaires sont-elles signalées comme elles le sont?**
 - Des facteurs textuels sont-ils susceptibles d'expliquer la façon dont la série est marquée ou l'usage d'un marqueur particulier ?
 - Des facteurs textuels sont-ils susceptibles d'expliquer le non marquage d'un ou plusieurs items ?
 - En l'absence d'usage de marqueur explicite, comment l'item est-il identifié ?
3. **L'usage des séries linéaires, varie-t-il selon la discipline scientifique ?**

Comme **hypothèse 1**, nous avançons que les séries linéaires contribuent à l'organisation textuelle à plusieurs niveaux du texte, organisant des segments très courts et très longs. De plus, outre le fait de segmenter le texte en items qui se suivent, nous supposons que les

séries linéaires peuvent également organiser un segment textuel en thèse, antithèse et conclusion, comme le note l'étude de Jackiewicz (2005) : *D'abord... Mais... Bref...* De plus, suivant les travaux de Porhiel (2007) sur les *séries à deux temps*, soit les séries combinant deux séries, il est à supposer que ces structures sont utilisées également dans notre corpus.

Deuxièmement, comme **hypothèse 2**, nous supposons que le marquage des séries est accompli avec une rigueur certaine grâce au rôle important que le marquage de l'organisation textuelle joue dans les articles de recherche et grâce au processus d'édition et de révision que ces textes ont parcouru avant d'être publiés. Par conséquent, l'usage des marqueurs ou des paires de marqueurs dont l'acceptabilité peut être mise en question, tel que *d'une part sans d'autre part*, est assurément peu fréquent. De même, il est probable que les amorces incorrectes, telles que *Cet article consiste en trois parties* alors que ce n'est pas le cas en réalité, sont très rares.

Or, quoique le marquage de l'organisation textuelle soit rigoureux, il serait irréaliste de penser que tout soit indiqué explicitement dans les articles de recherche : un tel texte serait très pesant à lire. Ainsi, l'usage abondant des séries prototypiques, dans lesquelles tous les items sont signalés par des MIL, alourdirait le texte et nous semble peu probable. Une autre question liée au marquage porte sur l'équivalence entre les items des séries linéaires, notamment proposée par Jackiewicz (2005 : 95) mais contestée par Virbel (1999) et Péry-Woodley (2000 : 138-139). Dans la pratique, le déséquilibre peut se manifester, entre autres, par l'usage de marqueurs différents dans une même série, par des items non marqués ou par une variation importante de longueur entre les items (cf. Luc 2001 et la section 4.2.4). Bien qu'il soit difficile d'estimer la proportion des séries équivalentes ou déséquilibrées, le rôle important du marquage de l'organisation textuelle dans les articles de recherche et le processus d'édition et de révision qu'ils ont parcouru ont sans doute également un effet sur celle-ci.

Une autre hypothèse liée au marquage des items, soit **l'hypothèse 3**, concerne le fonctionnement des différents types de marqueurs. De fait, les résultats de plusieurs travaux antérieurs montrent que leur capacité organisatrice varie. Entre autres selon Schnedecker (2000 : 22), le point de référence des expressions anaphoriques *le premier – le second* se trouve en général dans la proximité. De façon analogue, les séries introduites par ces marqueurs devraient être relativement courtes. Pour les connecteurs, Ho Duc (2007 : 227) avance que ces marqueurs organiseraient le plus souvent des segments d'un paragraphe tandis que Jackiewicz et Minel (2003 : 7) parlent de deux paragraphes. La capacité des connecteurs semble donc également plutôt limitée quoique plus vaste que celle des anaphoriques.

Enfin, notre dernière **hypothèse 4** porte sur la fréquence inférieure des séries linéaires dans les articles du domaine de l'histoire en regard des deux autres disciplines. Comme nous l'exposons dans la section suivante, le marquage de l'organisation textuelle varie selon la

discipline d'appartenance de l'article. L'histoire se caractérise ainsi par l'usage des structures et du style narratifs (cf. par exemple Holmes 1997 ; Silver & Bondi 2004 ; Bondi 2009a) ce qui peut s'associer avec la rareté relative du marquage d'organisation textuelle. Par conséquent, nous estimons que les séries linéaires sont moins fréquentes dans ce sous corpus que dans les articles de linguistique et d'éducation.

PARTIE THÉORIQUE

2. LINGUISTIQUE DE CORPUS ET LE DISCOURS DE RECHERCHE

Nous approchons les séries linéaires d'un point de vue fonctionnel. Autrement dit, nous les examinons à partir de leur fonction dans le texte. Ensuite, notre approche est textuelle : nous nous intéressons à la relation entre une forme linguistique et sa fonction textuelle et non entre une forme linguistique et un processus cognitif sous-jacent à son usage, par exemple. De surcroît, notre approche intègre deux autres aspects déterminants que sont la linguistique de corpus et le discours de recherche que nous avons d'ailleurs choisi comme genre de notre étude. Cette section traite ces deux aspects. Nous commencerons par la justification de l'usage d'un corpus et la discussion des atouts d'une approche quantitative, et puis, la seconde partie de la section abordera le discours de recherche. Nous présentons d'une façon générale le discours de recherche et les articles de recherche et menons une réflexion sur le métadiscours, un phénomène textuel proche des séries linéaires et souvent étudié dans les articles de recherche. Enfin, la section débat des articles de recherche et de leur variation selon la langue et la discipline en question.

2.1 LINGUISTIQUE DE CORPUS

En définissant nos objectifs pour cette thèse, nous avons indiqué vouloir examiner les séries linéaires par rapport à leur usage en nous concentrant sur leur mode d'emploi dans la pratique. Pour ce faire, nous adoptons une approche quantitative et fondons notre étude sur un corpus².

Traditionnellement les études quantitatives sur l'organisation textuelle et son marquage sont considérées comme encore plutôt rares et les études portent souvent sur des exemples inventés par le chercheur ou sur un nombre limité d'attestations. Biber *et al.* (1998 : 106-107) et Péry-Woodley (2005 : 202, cf. aussi Leech *et al.* 1997 ; Péry-Woodley et Scott 2006 : 14) soulignent que l'analyse quantitative de l'organisation textuelle est compliquée, avant tout, par une difficulté d'identification automatique et un manque d'outils computationnels nécessaires. Contrairement aux mots ou morphèmes individuels, elle peut se manifester par des indices très variés difficiles à énumérer à l'avance de manière exhaustive. Le non marquage est une autre éventualité : tout n'est pas signalé explicitement par un indice linguistique. En outre, les indices linguistiques, comme les connecteurs, sont souvent polyfonctionnels et peuvent remplir d'autres fonctions que l'indication de l'organisation textuelle. Par conséquent, un traitement manuel, facilement laborieux et exigeant des outils spécialisés, est en général nécessaire pour le repérage de ces indices.

² Par *corpus* nous comprenons une collection de textes avec un volume de données important permettant une étude quantitative (cf. 5.1.1).

Deuxièmement, la rareté des études quantitatives sur l'organisation textuelle peut également être imputée au fait que cette approche n'a simplement pas été considérée comme utile dans des travaux antérieurs. Dans la tradition générativiste, par exemple, l'intuition du chercheur est considérée comme une source d'information suffisante. Un corpus de soutien n'est donc pas jugé nécessaire (cf. Chomsky 1957 ; la discussion dans McEnery & Wilson 1996 : 4-11 ; Condamines 2003 : 13-15).

Nous adoptons une approche quantitative surtout dans le but d'examiner l'usage de la langue et non ses possibilités théoriques. Nous estimons que l'intuition ne suffit pas à expliquer, par exemple, les raisons sous-jacentes du non marquage et les différentes fonctions textuelles des marqueurs. L'intuition peut également se fourvoyer (cf. aussi Reppen *et al.* 2002 : VII et Condamines 2005 : 16). Dans l'approche quantitative, en revanche, les résultats sont plus indépendants de l'analyste et ainsi plus objectifs que s'ils se fondent sur l'introspection ou seulement quelques exemples. En outre, la généralisation des résultats au-delà du corpus d'étude et la comparaison des résultats sont des atouts souvent associés à un corpus vaste (voir Biber 1995 : 32 ; McEnery & Wilson 1996 : 62, 77-81 ; Biber *et al.* 1998 : 106 ; Péry-Woodley 2005 : 185). Pour autant, nous tenons à souligner que notre approche demeure strictement complémentaire de l'approche qualitative et que la présente étude serait inenvisageable sans les connaissances issues des travaux fondés sur des corpus restreints. Notre but est de compléter ces travaux par l'examen de l'emploi, dans la pratique, du mode d'agencement du texte et de son marquage.

Pour surmonter les écueils liés à la difficulté d'identification des marqueurs d'organisation textuelle, la méthode conventionnelle porte sur leur marquage manuel dans le corpus puis leur examen à partir de ce marquage (entre autres Taboada 2006 ; Prasad *et al.* 2008 ; Ho Dac *et al.* 2009). Selon cette méthode, les connecteurs tout comme les contenus sémantiques des relations non marquées sont indiqués dans le texte par les annotateurs humains. Optionnellement, le repérage automatique des phénomènes textuels examinés et les techniques automatiques d'accès au contenu textuel offrent une méthode alternative à l'annotation manuelle. Ces techniques sont cependant surtout utilisées à d'autres fins que l'examen de l'organisation textuelle (Nazarenko 2005, mais aussi Teufel 1998 pour un exemple opposé).

Enfin, pour le repérage des séries linéaires dans le corpus, nous suivons le modèle adopté entre autres par Taboada (2006), Prasad *et al.* (2008) et Ho Dac *et al.* (2009) : nous annotons les séries linéaires manuellement dans le texte. Notre analyse partant du principe que la fonction discursive n'est pas une forme ni une liste de formes préalablement rédigée, le repérage automatique des séries linéaires et surtout des items non marqués serait une tâche trop exigeante. L'annotation manuelle a pourtant aussi quelques points faibles. Le plus important, peut-être, concerne l'objectivité des résultats tirés d'annotations faites par des humains. Comme nous l'expliquons en section 5.2, nous nous efforçons de

minimiser cette subjectivité par l'identification automatique des marqueurs explicites et par l'analyse des cas difficiles par plusieurs personnes.

2.2 DISCOURS DE RECHERCHE

Cette thèse fait partie intégrante du projet *Perspectives croisées sur le discours de recherche* du département d'études françaises de l'Université de Turku. Le projet vise à étudier les stratégies organisationnelles, cognitives et textuelles dans trois disciplines de recherche, soit en linguistique, l'éducation et l'histoire. Par *discours de recherche*, nous entendons, d'une manière plutôt restreinte, la communication entre les spécialistes dans un milieu universitaire sur des sujets professionnels. Toutefois, nous en excluons la communication d'enseignement de cette pratique aux étudiants, à l'école comme à l'université.

2.2.1 POURQUOI LE DISCOURS DE RECHERCHE ?

Le discours de recherche a fait l'objet, surtout depuis les années 1990, d'un nombre croissant de travaux. Cet intérêt peut s'expliquer par la nouvelle conception de ce domaine. Contrairement à son présupposé traditionnel d'objectivité et de neutralité, les travaux plus récents s'accordent sur le fait que le discours de recherche vise, outre le fait d'informer, à convaincre et influencer le lecteur tout en se faisant accepter dans la communauté scientifique. Markkanen et Schröder (1997 : 9), par exemple, notent que « scientific texts are not only content-oriented and informative but also aim at convincing and influencing their audience » (« les textes scientifiques ne sont pas uniquement orientés sur le contenu et l'information mais tendent également à convaincre et influencer leur public ») (cf. aussi Hyland 1998a : 438-439). Le discours de recherche n'est non plus considéré comme neutre. En revanche, les marques d'atténuation, c'est-à-dire les façons dont l'auteur exprime son incertitude ou des précautions envers son discours, sont un aspect très étudié dans ce domaine (cf. Meyer 1997 ; Hyland 1998b ; Vold 2008).

Cas fréquent dans les différents domaines de la linguistique, la majorité des études menées sur le discours de recherche traite de l'anglais. Citons, notamment, en exemple le travail de Hyland sur les marques d'atténuation (1998b, 2005b) et sur le métadiscours qui guide le lecteur dans l'interprétation du texte (Hyland 1998a, 2005a). Le travail de Swales (1990 et 2004, etc.), pour sa part, traite la description des articles de recherche comme un genre discursif et sa macrostructure ainsi que les pas rhétoriques propres à chaque section de l'article. Les études contrastant les pratiques de l'anglais et d'une autre langue, par exemple pour la pratique rhétorique des natifs et des non natifs (Mauranen 1993, 2009) ou l'amorce des introductions dans les articles de recherche (Silver & Bondi 2004 ; Bondi 2009a), constituent également un groupe important de travaux.

En revanche, les études sur le discours de recherche en français sont relativement rares. Nous pouvons noter, en exception, les travaux réalisés dans le cadre du projet KIAP à

l'université de Bergen sur la manifestation de l'identité culturelle. Comparant les articles de recherche en français, anglais et norvégien, d'une part, et en linguistique, médecine et économie de l'autre, ce projet opte pour une double perspective contrastive (Fløttum *et al.* 2006 : 1-2). En son sein, la modalité épistémique (Vold 2008), le métadiscours (Dahl 2004 et Fløttum *et al.* 2006) et l'usage du pronom *on* (Gjesdahl 2008), notamment, ont constitué des objets d'intérêt. D'autres travaux sur le discours de recherche en français incluent la thèse de Poudat (2006) sur l'analyse des genres, la thèse de Rinck (2006) sur les modes de « subjectivité représentée » dans l'article scientifique, et le travail de Vassileva (2000) sur la présence de l'auteur en français, anglais, russe, allemand et bulgare.

Dans les travaux relatifs au discours de recherche, les articles de recherche ont reçu beaucoup d'attention (pour le genre de l'article de recherche, cf. Swales 1990 et 2004). Le rôle central des articles au sein de la communauté de recherche est l'une des raisons de cet intérêt ; comme le dit Eva Vold dans sa thèse (2008 : 21), il s'agit du moyen central de communication pour les chercheurs afin d'échanger des résultats (cf. aussi Hyland 2005a : 89-90). De plus, contrairement aux communications orales et aux ouvrages entiers, par exemple, les articles de recherche offrent un accès relativement facile, même au format électronique.

Les articles de recherche forment pour nous un objet d'étude opportun à divers titres. Une première raison concerne le rôle central que l'organisation textuelle et son marquage revêtent dans ce genre. Comme nous venons de le constater, le but des articles de recherche est non seulement d'informer le lecteur mais également de le convaincre des résultats présentés. Hyland (1998a : 438-439) a également avancé que, pour convaincre le lecteur, il est primordial de le prendre en compte lors de la rédaction. En ce sens, l'organisation textuelle doit être également claire et suffisamment explicite puisqu'un texte mal organisé ne peut que difficilement être convaincant. Cela étant, nous devons toutefois convenir que tous les articles de recherche ne correspondent sans doute pas à cette présomption et que ce n'est pas nécessairement le but de certains articles. Les textes rédigés par des chercheurs très renommés, par exemple, peuvent être extrêmement lus et faire l'objet de multiples interprétations bien qu'ils ne soient ni clairs ni faciles à lire.

Le processus d'édition et de révision affecte également l'organisation textuelle et, pour le moins, son marquage dans les articles de recherche. Au lieu d'être rédigé par une seule personne, l'article de recherche peut être modifié par l'éditeur et les évaluateurs anonymes de la revue. Comme l'indique Swales (1990 : 93), l'article publié est plutôt un produit de la négociation menée entre l'auteur et l'éditeur afin de préserver la qualité et les autres normes de la revue. Pour les séries linéaires, nous pouvons supposer que ce processus résulte en une réflexion sur les façons dont l'article est organisé et dont cette organisation est signalée. En pratique, elle pourrait notamment être visible dans l'absence de marqueurs à l'acceptabilité questionable, tels que *d'une part* sans *d'autre part*.

La nature soutenue des articles de recherche et le rôle central que l'organisation textuelle y occupe font de ces textes un objet d'étude intéressant, surtout pour les séries linéaires. De fait, et comme nous l'indiquons en définissant les séries linéaires, ces structures sont souvent associées à une certaine équivalence entre les items (cf. par exemple Jackiewicz 2005). Toutefois, dans la pratique, Virbel (1999) et Péry-Woodley (2000 : 138-139), entre autres, constatent que les items peuvent tout de même être très différents et marqués de façons variées. Pour notre part, nous estimons très intéressant l'examen de la fréquence du non marquage des items et des structures déséquilibrées également dans les articles de recherche qui, comme nous venons de l'exposer, sont souvent associés à la clarté et la précision ainsi qu'à une organisation textuelle rigoureuse.

Enfin, nous justifions notre choix des articles de recherche comme corpus par une dernière raison qui est d'ordre pratique : leur longueur est idéale pour une étude des séries linéaires. Ils sont tout à la fois suffisamment longs pour permettre l'observation du fonctionnement des séries linéaires dans des segments vastes, organisés en sections et sous-sections, et suffisamment courts pour rendre possible une étude quantitative reposant sur un corpus composé d'un nombre relativement important de textes entiers. De plus, Internet et les versions électroniques des articles de recherche qu'il contient assurent un accès relativement commode à un nombre important de textes au format électronique : un aspect facilitant l'étude quantitative que nous souhaitons mener.

2.2.2 METADISOURS ET SERIES LINEAIRES

Parmi les travaux sur les articles de recherche, les plus proches de notre problématique sont ceux traitant du métadiscours (par exemple Hyland 1998a et 2005a ; Dahl 2004 ; Lores Sanz *et al.* 2006). Le métadiscours est souvent (Hyland et Tse 2004 : 156 ; Hyland 1998 : 437 ; Dahl 2004 : 1811) défini comme les aspects du texte par lesquels l'auteur prend en compte le lecteur et est lui-même présent dans le texte. Cette définition repose sur la conception de l'écriture comme un engagement social entre l'auteur et le lecteur. Traditionnellement, Hyland (1998a : 442-444) attribue deux fonctions au métadiscours : le *métadiscours textuel* inclut les indices exprimant explicitement le mode d'interprétation des propositions afin de correspondre aux intentions de l'auteur. Cette catégorie contient des connecteurs tels que *finalement, mais, ainsi* ainsi que des expressions référant ailleurs dans le texte telles que *comme nous l'avons déjà indiqué* et *voir section 2*. La plupart des séries linéaires serait également classifiée dans cette catégorie. La deuxième catégorie, le *métadiscours interpersonnel*, indique, en revanche, au lecteur l'attitude de l'auteur vis-à-vis de l'information propositionnelle du texte et du lecteur lui-même. Ce type de métadiscours peut se manifester, par exemple, par les marques d'attitude ou de personne tels que *nous sommes d'accord, étonnamment* et *nous, à moi*, etc. (Pour d'autres classifications, cf. Vande Kopple 1985 ; Crismore *et al.* 1993 ; Dahl 2004).

Au lieu d'une distinction nette entre les fonctions textuelle et interpersonnelle du métadiscours, les travaux plus récents de Hyland (2005a) et de Hyland et Tse (2004)

soulignent l'aspect interpersonnel de tout le métadiscours. Cette nouvelle vue correspond peut-être mieux à la nôtre puisqu'elle privilégie l'importance du marquage de l'organisation textuelle dans la construction du texte même et ne la traite pas comme une partie séparée. Selon cette approche, outre le métadiscours interpersonnel, aussi le métadiscours textuel reflète la façon dont le lecteur perçoit et comprend le texte. Par conséquent, ce type de métadiscours affecte également le contenu sémantique du texte. Pour mieux s'ajuster à cette conception du texte, Hyland (2005a : 48-49) et Hyland et Tse (2004) proposent un nouveau schéma des deux fonctions du métadiscours. La fonction *interactive* est liée à la prise en compte du lecteur et de ses besoins. Elle tend à expliciter l'organisation du texte et les interprétations souhaitées par l'auteur, soit à « guider le lecteur à travers le texte » (Hyland 2005a : 49). Les indices de ce type de métadiscours correspondent à ceux du métadiscours textuel du paragraphe précédent. La deuxième fonction *interactionnelle*, en revanche, est liée à la façon dont l'auteur poursuit l'interaction et commente son message. L'auteur cherche à exprimer explicitement son opinion envers le texte et à engager le lecteur dans le texte. Cette fonction du métadiscours peut être manifestée par le même type d'indices que le métadiscours interpersonnel du paragraphe précédent.

En définitive, un aspect du métadiscours qui le distingue de notre problématique est que les séries linéaires n'en font pas toutes partie : au lieu de se référer au texte, les marques de cohésion et autres indices signalant les items des séries linéaires peuvent également faire référence aux événements du monde extérieur (Martin 1992 : 178-183 ; Martin et Rose 2003 : 120-140 ; cf. aussi Adam 1990 : 160³). Dans les travaux sur le métadiscours, cette distinction entre les connecteurs *internes* et *externes* se fait naturellement : les connecteurs *internes* peuvent être inclus dans le métadiscours, tandis que les connecteurs *externes* en sont exclus (cf. aussi Hyland et Tse 2004 ; Hyland 2005a : 45-48). Ainsi, dans l'Exemple 3 suivant, le connecteur *enfin* se réfère aux événements du monde extérieur énumérés et ne peut pas être analysé comme métadiscours.

Exemple 3

- 1 Par exemple, c'est elle qui, en 1932, demande la création du système de cartes.
- 2 Deux ans plus tard, le bureau de placement, qui avait fermé ses portes en 1931 en raison de ses coûts trop élevés et des cas d'abus découverts, est également réorganisé à la suite d'une requête de l'Association des manufacturiers.
- 3 **Enfin**, en 1940, lorsque le conseil de ville voudra réformer la distribution des pièces d'identité par le centre de placement, il ne manquera pas de consulter l'Association des manufacturiers.

Roux-Pratte

Dans les travaux relatifs à la cohésion et l'organisation textuelle, la distinction entre les marqueurs internes et externes ne survient que rarement (citons comme exceptions les

³ Dans la distinction des deux directions de référence, Adam (1990 : 160) adopte une perspective alternative et sépare l'organisation réalisée par la *référence*, c'est-à-dire par un mode d'organisation du réel, et l'organisation construite par le *discours* qui est une trace de la mise en texte.

travaux de Martin 1992 et Martin et Rose 2003 ; Adam 1990) ; nous pouvons supposer que ce choix est influencé par la contribution des deux types marqueurs à la cohérence. De plus, comme le constate Martin (1992 : 183), certains cas rendent parfois difficile la distinction entre ces deux types qui adoptent aussi les mêmes formes lexicales. Nous incluons dans l'analyse les deux types de marqueurs mais faisons également la distinction entre ceux-ci (cf. section 6.1).

Deuxièmement, un autre aspect du métadiscours l'éloigne de notre problématique : il est obligatoirement explicitement marqué dans le texte tandis que nous nous intéressons également aux séries linéaires avec des items non marqués. Comme l'énonce Hyland (2005a : 28), l'intérêt envers l'étude du métadiscours réside dans sa présence dans le texte et ses fonctions, la principale étant de guider le lecteur dans la lecture. Celle-ci exige un usage explicite : le métadiscours non marqué ne peut pas guider le lecteur et serait ainsi peu intéressant. Dans l'étude de l'organisation textuelle, en revanche, il est également généralement admis, entre autres, que les connecteurs sont uniquement des signes de surface de l'organisation textuelle sous-jacente. L'alternance des marquages explicite et implicite ainsi que le lien entre une relation implicite et les marqueurs individuels utilisés pour la signaler ont également fait l'objet de nombreuses études (entre autres Mann et Thompson 1988 ; Taboada 2006 ; Vergez-Couret 2009).

Vu ces deux aspects distinctifs entre le métadiscours et les séries linéaires, notre problématique porte plus sur l'étude de l'organisation textuelle et moins directement sur l'étude du métadiscours. Tout de même, nous estimons pouvoir également contribuer à la recherche sur le discours de recherche, les articles de recherche voire le métadiscours. En menant une étude détaillée sur le marquage (et non marquage) de l'organisation textuelle, nous pensons pouvoir fournir des informations sur son usage plus exactes que celles proposées par les études antérieures sur le métadiscours. Par exemple, outre les fréquences des séries linéaires, nous pouvons également analyser leur longueur et les marqueurs utilisés.

2.2.3 ARTICLES DE RECHERCHE ET VARIATION INTERDISCIPLINAIRE ET INTERLINGUISTIQUE

La variation interdisciplinaire et interlinguistique du discours de recherche a inspiré de nombreuses études. L'hypothèse principale fondant ces travaux est l'influence de ces deux propriétés au texte ; comme l'indiquent Dahl (2004 : 1808) et Fløttum *et al.* (2006 : 16 et passim) (voir aussi Becher & Trowler 2001), le chercheur rédigeant un article de recherche est influencé par les pratiques de sa langue maternelle ainsi que par celles de sa discipline et de sa communauté scientifique d'appartenance. Par conséquent, les articles contiennent des traces de ces cultures et diffèrent selon la langue et la discipline. Des différences existent, entre autres, dans la présence de l'auteur, la fréquence des marques d'incertitude qu'il emploie et le marquage de l'organisation textuelle.

Dans cette étude, outre l'examen de l'organisation textuelle et du métadiscours en général, nous adoptons également une perspective comparative : comme nous l'indiquons, notre corpus se compose d'articles de recherche de trois disciplines : la linguistique, l'histoire et l'éducation. Grâce à cette répartition, non seulement nous élargissons la représentativité de notre étude au-delà d'une seule discipline mais nous comparons aussi les domaines étudiés.

Dans les sections suivantes, nous chercherons à décrire les articles de recherche formant notre corpus, en tenant surtout compte des particularités qu'ils portent par rapport à la langue utilisée (le français) et aux disciplines représentées. Ce faisant, nous nous efforcerons de prévoir les différences potentielles de ces disciplines dans l'organisation textuelle et son marquage, tout en prenant en compte le fait qu'un article de recherche typique n'existe probablement pas (cf. Fløttum 2006). Nous commencerons par présenter les caractéristiques des articles de recherche rédigés en français et poursuivrons par la discussion d'ordre général de la distinction entre les sciences dures et les sciences humaines et sociales, sans oublier les raisons sous-jacentes des différences entre ces deux groupes. Enfin, nous terminerons par l'examen plus détaillé de chaque discipline choisie dans notre corpus.

2.2.3.1 Articles de recherche en français

Les articles de notre corpus sont, évidemment, rédigés en français. Bien que la langue de rédaction de l'article semble influencer sa forme et son contenu dans une moindre mesure que sa discipline d'appartenance, quelques différences notables entre les articles de recherche rédigés en différentes langues existent, essentiellement rapportées grâce au projet KIAP (Fløttum *et al.* 2006). Notre problématique est notamment influencée par une différence interlinguistique concernant la distinction proposée par Hinds (1987) entre les langues dans lesquelles la réussite de communication repose surtout sur l'auteur et celles pour lesquelles cette tâche revient au lecteur. En anglais, ces deux rôles sont dénommés « writer responsible » (« auteur responsable ») et « reader responsible » (« lecteur responsable »). Autrement dit, selon Hinds (1987 : 143), dans les langues « writer responsible », l'auteur doit répondre de la réussite du passage de son message. L'échec de la transmission de l'information et de la communication est dû à un manque de clarté de l'auteur et non à un défaut d'attention du lecteur. Dans les langues « reader responsible », en revanche, la charge de la responsabilité est inversée. Dans la pratique, la différence entre ces deux types de langue est manifeste par exemple dans la fréquence du métadiscours, utilisé par l'auteur pour guider le lecteur dans l'interprétation du texte et pour expliciter, entre autres, l'attitude de l'auteur vis-à-vis des phénomènes discutés. Dans les langues recourant moins au métadiscours, le lecteur doit s'efforcer davantage de comprendre ces aspects du texte.

Selon Dahl (2004 : 1820-1821), le français dans les articles de recherche se caractérise surtout comme « reader responsible » : l'interprétation du texte incombe au lecteur qui doit s'efforcer de comprendre le message. L'auteur n'est pas responsable du succès de la

communication. En ce sens, Dahl (2004 : 1820) relève que la responsabilité du lecteur se reflète dans l'usage du métadiscours : il représente en français uniquement la moitié de celui utilisé en anglais et en norvégien, des langues caractérisées comme « writer responsible ». La différence semble donc essentielle.

Outre le métadiscours, le projet KIAP a également relevé d'autres indices textuels dont le statut varie selon la langue de rédaction. Selon Fløttum *et al.* (2006 : 74), l'usage des pronoms personnels de la première personne comme sujets d'une phrase, par exemple, apparaît plus fréquent en anglais et en norvégien qu'en français. Le français tente de les éviter. De plus, Fløttum *et al.* (2006 : 265) indiquent qu'en norvégien il est avéré que l'usage des négations et des conjonctions adversatives est plus fréquent qu'en français et anglais. Cet élément pourrait suggérer que le discours de recherche norvégien est plus ouvertement polyphonique que celui des deux autres langues. Enfin, pour la modalité épistémique, ou le marquage d'incertitude et de certitude, Vold (2008) a démontré que le français utilise moins des expressions de ce genre que les deux autres langues.

Selon ces résultats, le français apparaît, en définitive, comme une langue « reader responsible », recourant donc moins au métadiscours. De plus, la fréquence relativement basse des indices explicites du lecteur (marques de modalité épistémique, usage de la première personne et indices de polyphonie) et de l'interaction entre le lecteur et l'auteur (surtout le métadiscours) propose que cette langue tente d'éviter des indices explicites de l'auteur et de son interaction avec le lecteur dans le texte. Enfin, la nature « reader responsible » du français ne signifie pourtant pas que le métadiscours, et par conséquent le marquage de l'organisation textuelle, soient absents des articles de recherche en français. En effet, cette propriété est relative et dépend également de l'objet de comparaison. Suomela-Salmi (1992), par exemple, constate que l'auteur exprime plus explicitement la façon dont le texte est organisé en français qu'en finnois.

2.2.3.2 Articles de recherche en sciences naturelles et en sciences humaines et sociales

Les différences entre les disciplines sont souvent illustrées à travers une division des domaines entre sciences « dures » et sciences « molles », c'est-à-dire entre les sciences naturelles et les sciences sociales et humaines. Les différences entre ces deux domaines sont si fondamentales qu'elles ont des conséquences inévitables pour la forme et le contenu d'un article. Les objets de recherche et les communautés scientifiques de ces domaines, par exemple, diffèrent grandement, ce qui influence le mode de rédaction des articles et les attentes des lecteurs à leur égard.

Tout d'abord, comme l'indique Hyland (2005b : 187-188), les pratiques rhétoriques de l'auteur sont intimement liées au but et à la nature des résultats de la discipline. Or, ils ne sont pas nécessairement identiques dans tous les domaines. Les sciences naturelles cherchent à produire de l'information fondée sur la croissance cumulative des résultats. Pour les fournir, le chercheur peut s'appuyer sur des procédures communément acceptées

et connues dans la communauté. Dans les sciences humaines et sociales, en revanche, l'auteur n'a jamais la certitude que l'audience accepte l'approche méthodologique adoptée et doit argumenter pour ses choix.

Plus loin, Hyland (2005b : 187-188, cf. aussi Hyland 2000 : 33 ; Dahl 2005 : 294-296 ; Fløttum *et al.* 2006 : 20-21) relève que la nature différente des résultats est manifeste dans les modes de rédaction des articles : les sciences naturelles rapportent plutôt les résultats de tests effectués, tandis que les sciences sociales et humaines interprètent les résultats et les construisent dans le texte. Afin de les faire accepter, l'auteur doit également argumenter plus ouvertement. Les articles sont donc plus subjectifs et la présence de l'auteur plus explicite. Pour cette raison, les articles de recherche en sciences humaines et sociales contiennent également davantage d'indices d'incertitude et de certitude. Pour décrire cette différence, Hyland (2000 : 33) avance que les chercheurs en sciences humaines et sociales se proposent comme but de « découvrir la vérité au lieu de la créer » (« discovering the truth, not making it »).

L'objet de la recherche et la nature des résultats distinguent également les sciences humaines et les sciences sociales. Cependant, cette distinction n'est pas aussi nette que celle entre les sciences naturelles et celles-ci. Fløttum *et al.* (2006 : 20-21) notent que les sciences sociales sont situées entre les sciences humaines et les sciences naturelles : certaines parties d'articles peuvent, par exemple, reposer sur des calculs exacts et des résultats mesurables, tandis que d'autres sont plus subjectives. En sciences humaines, en revanche, l'interprétation subjective des résultats est le mode d'analyse principal. Cela ne signifie évidemment pas que la recherche en sciences humaines se fonde uniquement sur la réflexion personnelle du chercheur. Pourtant, le chercheur y est plus libre pour créer des relations entre des faits constatés.

La communauté scientifique de rédaction de l'article n'est pas non plus sans conséquences. Selon Gunnarsson (2003, cité in Dahl 2005 : 294), les communautés scientifiques fortes ont des conventions plus strictes que les communautés dispersées. La structuration du texte selon le modèle traditionnel Introduction-Méthode-Résultat-Et-Discussion (IMRAD, Swales 1990) est l'exemple d'une telle convention, qui est aussi le signe d'une communauté forte. Le projet KIAP (Dahl 2004 : 1819-20) concluait que les articles de médecine suivent ce modèle beaucoup plus strictement que, par exemple, les articles en linguistique. La même tendance est notée par Breivega (2003 : 12) qui constate que les textes des sciences humaines sont beaucoup plus hétérogènes que ceux des sciences naturelles. Poursuivant, elle relève que la personnalité de l'auteur s'y apprécie d'une manière différente : le style personnel de l'auteur peut être visible.

Les articles en sciences humaines et sociales semblent donc se caractériser par une certaine liberté, issue de la nature de leur objet de recherche et des traditions de leur communauté scientifique. Comme ils sont ouvertement plus argumentatifs que leurs équivalents en sciences dures, l'auteur y est plus explicitement présent pour soutenir son opinion.

L'auteur bénéficie également du droit d'écrire son article plus librement que ses collègues des sciences naturelles, contraints de suivre plus strictement les conventions de leur communauté.

Notre corpus consiste donc en articles de linguistique, d'histoire et d'éducation. Toutes trois peuvent être considérées comme des sciences « molles » dont nous venons de souligner certaines caractéristiques. Elles ne forment pourtant pas un champ homogène. Tandis que l'histoire est souvent conçue comme un prototype des sciences humaines, la linguistique a été nommément située à l'intersection des sciences humaines et sociales (Fløttum *et al.* 2006 : 20-21). En revanche, les sciences de l'éducation peuvent être incluses aussi bien dans les sciences humaines que dans les sciences sociales.

Evidemment, l'inclusion dans le corpus d'une science « dure », probablement très différent des sciences « molles », nous permettrait plus certainement de contester des différences entre les domaines examinés. Au lieu de viser à démontrer des différences aussi claires que possibles, notre choix des disciplines étudiées se base pourtant sur d'autres critères.

Nous avons choisi les articles en éducation et en histoire surtout en raison de la rareté des études en ces domaines. Nous relevons une lacune nette des travaux sur ces disciplines, surtout en français. De plus, comme nous allons l'exposer, l'organisation textuelle des articles en histoire est censée (Holmes 1997 ; Silver & Bondi 2004 ; Bondi 2009a) différer de celle des autres domaines, ce qui les rend particulièrement intéressants à nos yeux. Enfin, les articles en linguistique ont été choisis en premier lieu pour garder un point de repère en regard des autres travaux.

À la suite, nous poursuivrons par l'examen plus en détail de chacune de ces disciplines. Nous commencerons par la linguistique et continuerons par l'éducation et l'histoire.

2.2.3.3 Articles de recherche en linguistique

Selon Fløttum *et al.* (2006 : 21), la linguistique n'est pas nécessairement une discipline prototypique des sciences humaines : son argumentation ne repose pas uniquement sur le raisonnement et l'argumentation de l'auteur, et les résultats peuvent également contenir des informations mesurables. En outre, le processus de recherche peut comporter des régularités et des outils méthodologiques plus communs dans les sciences sociales.

Toutefois, la linguistique affiche également beaucoup de propriétés l'ancrant fermement dans les sciences humaines. Ce sont surtout le rôle important de l'argumentation et la visibilité explicite de l'auteur dans les articles, principalement provoqués par la nature subjective des résultats, qui en constituent des signes. Plusieurs études du projet KIAP discutent de ces propriétés. Dahl (2004 : 1819-1820) indique que l'auteur en linguistique se caractérise comme *l'auteur-acteur* visible dans le texte, essentiellement du fait que les résultats sont en large part construits *via* l'argumentation. Poursuivant, elle énonce que la

visibilité explicite de l'auteur se manifeste également dans l'usage plus important du métadiscours. En outre, Fløttum *et al.* (2006 : 263) notent que la nature argumentative des articles en linguistique se manifeste dans la pratique, par exemple, par l'usage accru des négations et des conjonctions adversatives et par la fréquence plus élevée des pronoms de la première personne. Les verbes utilisés pour se référer aux sources bibliographiques démontrent également la visibilité explicite de l'auteur.

Le rôle et la nature de la communauté scientifique forment une autre propriété de la linguistique typique des sciences humaines. Comme nous l'indiquons, la normalisation du format de l'article de recherche est signe d'une communauté scientifique forte, au sein de laquelle l'auteur doit suivre les conventions du domaine plus strictement que dans les communautés moins fortes. Les sciences humaines, comme la linguistique, se caractérisent par des conventions plutôt libres. Dahl (2004 : 1819-1820), entre autres, indique que la linguistique est dépourvue de format d'article de recherche normalisé. De là des organisations hétérogènes ainsi qu'un usage fréquent du métadiscours pour guider le lecteur. Comme le lecteur ne peut pas connaître à l'avance le mode d'organisation de l'article, il a besoin d'une aide accrue dans sa lecture.

2.2.3.4 Articles de recherche en éducation

Selon Rinne *et al.* (2004 : 45), l'éducation peut faire partie soit des sciences sociales, soit des sciences humaines. Beke et Bolívar (2009), en revanche, la considèrent comme partie intégrante des sciences humaines. Pourtant, pour le moins comparée aux sciences humaines prototypiques, l'éducation affiche des caractéristiques susceptibles de la lier également aux sciences sociales. Rinne *et al.* (2004 : 27-29) indiquent, par exemple, que la recherche en éducation repose sur l'empirisme et que les méthodes statistiques sont souvent appliquées dans le cadre de la recherche bien qu'elles ne conviennent pas à tous les objets d'étude. Quelle que soit la place exacte de l'éducation parmi les sciences humaines et sociales, cette description peut également nous servir pour la description du domaine de l'éducation.

D'abord, les résultats des sciences de l'éducation affichant des traits des deux types, soit des éléments mesurables ainsi que des éléments plus subjectifs, nous pouvons penser que l'argumentation et la visibilité explicite de l'auteur dans ces articles se situent à équidistance des articles des sciences dures et des sciences humaines. Nous pouvons également supposer que les conventions relatives, par exemple, à la structuration du texte y sont plus libres que dans les sciences naturelles mais plus strictes que dans les sciences humaines prototypiques.

La nature plutôt libre des conventions des sciences de l'éducation est soutenue par une remarque de Beke (2005) dans son article sur le métadiscours interpersonnel dans les articles en éducation en espagnol. Elle y note que son usage varie selon l'auteur. Par conséquent, les conventions concernant la structuration du discours et l'usage du

métadiscours ne peuvent pas y être très strictes. De plus, Beke (2005) compare l'usage du métadiscours de son corpus à un corpus d'articles en anglais issus de disciplines variées étudié par Hyland (1999). Beke conclut que son corpus comprend en moyenne une expression métadiscursive pour 80 mots, alors que la fréquence est d'une expression pour 28 mots dans le corpus de Hyland. À notre avis, le résultat est pourtant difficile à généraliser pour décrire l'usage de tout type de métadiscours en éducation. En effet, le corpus de Hyland intègre huit disciplines parmi lesquelles les pratiques peuvent varier de façon notable (Hyland 1999 : 103). Par ailleurs, les études de Hyland (1999) et Beke (2005) s'intéressent uniquement au métadiscours interpersonnel, susceptible de couvrir aussi bien la majorité que la minorité du total du métadiscours. Par conséquent, ils ne permettent pas de déduire la fréquence totale du métadiscours dans le domaine.

Un deuxième aspect des sciences de l'éducation, étudié dans de rares travaux sur le sujet, concerne les marques de modalité utilisées. Dans un travail contrastif, Beke & Bolivar (2009) examinent les marques de modalité en éducation. Leurs résultats sont en fait plutôt surprenants : les marques de modalité sont plus fréquentes en éducation qu'en linguistique. Ce résultat proposerait que l'éducation, sous cet aspect, appartienne davantage aux sciences humaines que la linguistique. Or, la différence entre les deux disciplines étant minimale, la différence ou la similitude entre elles est à étudier plus en détail.

En somme, la description des articles de l'éducation n'est pas évidente, bien que certains travaux aient été menés en l'espèce. Des études ultérieures, et de préférence comparatives, seront nécessaires pour pouvoir situer correctement l'éducation en regard des autres disciplines. La présente étude tente également de contribuer à cette recherche.

2.2.3.5 Articles de recherche en histoire

Nous avons déjà mentionné que l'organisation textuelle des articles en histoire peut différer de celle des autres domaines. D'autres propriétés de l'histoire lui confèrent également une place à part en regard des autres disciplines. Commençons, pourtant, par les similarités présentées par ce domaine.

Tout d'abord, typiquement pour les sciences humaines, l'interprétation du lecteur constitue un aspect important de la recherche en histoire. Suivant Hunston (1993 : 132), nous pouvons même avancer que l'interprétation humaine constitue toute la discipline⁴. Pourtant, comme l'affirme Bondi (2009a : 87), elle n'est guère suffisante. Il convient également d'argumenter pour l'interprétation choisie : « history is not just account and interpretation of events, but also dialogic argumentation of the interpretation put forward » (« l'histoire n'est pas simplement un compte-rendu et une interprétation d'événements mais également une argumentation diagnostique de l'interprétation avancée ») (cf. aussi Martin 2003). Le rôle important de l'argumentation a aussi été prouvé par Hunston (1993) qui

⁴ Le fait de renoncer au but d'une description objective a provoqué beaucoup de discussion chez les historiens (cf. par exemple Carr 1961).

dans son étude contrastive entre la linguistique, la biochimie et l'histoire montre que l'histoire est la plus argumentative des disciplines étudiées.

Ensuite, une autre propriété de l'histoire typique des sciences humaines concerne le niveau de normalisation des articles. Comme constaté à plusieurs reprises, les sciences humaines se caractérisent par un niveau de normalisation plutôt faible. C'est également un constat de Bondi (2009a : 91) qui, en examinant les introductions et la façon dont les auteurs commencent l'article en histoire et en économie, note qu'en économie les ouvertures sont beaucoup plus normalisées et la structure plus explicitement signalée qu'en histoire. Poursuivant, elle relève la difficulté de la distinction des différents types d'ouvertures dans les articles d'histoire, du fait de leur nature dispersée.

Or, comme nous l'avons dit, l'histoire en tant que domaine affiche des propriétés qui la distinguent des autres, même au sein des sciences humaines. Dans cette discipline, le but de la recherche depuis le XIX siècle est censément de présenter « ce qui s'est réellement passé » (Carr 1961). Bien que cet objectif et son mode d'interprétation soient encore objets de discussion de nos jours (cf. Kalela 2000 ; Day 2008), nous pouvons relever ici une différence par rapport aux autres domaines. Tandis que plusieurs disciplines s'efforcent de généraliser des phénomènes ou d'autres actions à l'extérieur de l'objet d'étude limité, l'histoire cherche à présenter des périodes ou événements individuels du passé tels qu'ils étaient et à argumenter pour cette interprétation. Leur généralisation ou abstraction à concerner d'autres événements ou périodes ne ferait pas de sens, comme le note Holmes (1997 : 328).

Ce but - présenter des événements tels qu'ils se sont déroulés - est également visible dans l'organisation des textes des articles d'histoire : ils sont souvent « set in the narrative mode » (« en mode narratif ») (Bondi 2009a : 98) et accaparent également d'autres propriétés typiques des textes narratifs (voir aussi Bondi 2009b). Les articles en histoire examinés par Bondi (2009a) et Silver et Bondi (2004) commencent, entre autres, souvent par des descriptions d'un événement du passé pour lesquelles les voix des sources sont dramatisées par un discours rapporté. En outre, Holmes (1997 : 328) constate que les articles en histoire privilégient la structure *Introduction – Argument principal ou narratif – Discussion ou conclusion* et non pas le modèle IMRAD. Enfin, en se référant notamment aux travaux d'Eggins *et al.* (1993), Coffin (1997, 2006) et Martin (2003), Bondi (2009b : 163) souligne également l'importance des éléments temporels dans ce domaine.

2.2.3.6 Bilan : articles de recherche

Dans cette sous-section, nous avons eu pour but la description des façons dont les articles de recherche varient selon la langue et la discipline ainsi que la présentation des trois disciplines représentées par notre corpus. Très peu d'études offrant une vue holistique des particularités de ces disciplines, c'est une tâche que nous avons souhaité compléter.

Toutefois, tous les aspects de ces descriptions ne sont peut-être pas très importants en regard de notre problématique dans cette étude.

Les disciplines choisies dans notre corpus ont toutes leurs propres caractéristiques, susceptibles d'influencer l'organisation textuelle et son marquage. La linguistique et l'éducation présentent pourtant plusieurs similitudes : nous pouvons avancer qu'elles se situent à l'intersection des sciences humaines et sociales et qu'elles fondent leurs résultats tantôt sur des faits objectifs et mesurables, tantôt sur l'argumentation et le raisonnement de l'auteur. Les façons d'utiliser les marques de modalité dans ces deux domaines semblent également similaires. Par conséquent, nous présumons que leurs modes d'organisation et de marquage du texte sont aussi semblables. Il est pourtant intéressant de déterminer si c'est bien le cas : les résultats concrets, surtout sur les conventions de l'éducation, étant rares, il n'est pas possible de prévoir des résultats exacts.

Selon les travaux antérieurs, des trois disciplines décrites, c'est l'histoire qui diffère des autres le plus clairement. Le rôle inévitable de l'interprétation subjective des résultats peut être considéré comme un signe de son appartenance aux sciences humaines prototypiques. De plus, la nature de ses résultats et sa manière d'argumenter comportent des traits qui la distinguent même des autres sciences humaines : contrairement à celles-ci, l'histoire ne vise pas à généraliser des résultats au-delà de l'objet d'étude, ni à en déduire des abstractions. Les articles d'histoire ont également une tendance à dramatiser les voix de leurs sources ainsi qu'à utiliser des structures narratives dans le texte. Ces éléments affectent sans doute l'organisation du texte et son marquage. Le style narratif, surtout, suggérerait que les séries linéaires seraient moins fréquentes dans ces articles que dans ceux des deux autres disciplines.

3. FORMATION DU TEXTE

Un texte ne peut pas se composer de phrases individuelles et les phrases individuelles agencées les unes après les autres ne peuvent pas former de texte. Un texte est un ensemble, une unité complexe dont nous pouvons observer la formation selon divers points de vue et à laquelle de nombreux éléments linguistiques et extra-linguistiques contribuent.

Dans cette section, nous examinerons le texte ainsi que la façon dont les phrases individuelles peuvent le former et se transformer en une unité intelligible. Nous commencerons par étudier la cohérence, c'est-à-dire le principe cognitif selon lequel toutes les séquences d'énoncés, phrases incluses, sont interprétées comme une unité. Puis, nous nous proposerons de définir les deux concepts de texte et de discours, souvent susceptibles d'être confondus du fait du nombre de définitions attribuées à ces deux termes. Ensuite, la section 3.3 abordera la cohésion qui englobe les ressources linguistiques de l'organisation textuelle, soit les marqueurs linguistiques utilisés pour relier les phrases entre elles. Enfin, nous terminerons cette section par l'examen détaillé des différents types de marqueur de cohésion importants pour le marquage des séries linéaires, notamment les connecteurs, les expressions anaphoriques, la cohésion lexicale et l'encadrement du discours.

3.1 COHERENCE

La façon dont les phrases individuelles forment une unité intelligible est une question centrale en linguistique textuelle qui ne cesse d'intéresser les chercheurs. Comme l'indiquait déjà Charolles (2006 : 39), il est généralement accepté que le processus d'interprétation des phrases comme une unité est soumis au principe de *cohérence* qui s'applique à toutes les séquences d'énoncés. C'est donc ce processus qu'il faut explorer afin de pouvoir comprendre ce qu'est un texte et comment il est établi.

La cohérence est discutée par différents travaux selon diverses perspectives. Parmi toutes ces approches⁵, deux conviennent le mieux à notre étude. Charolles (1995, 1997, 2006) examine la cohérence comme un principe d'interprétation du texte, principalement du point de vue cognitif (voir aussi de Beaugrande et Dressler 1981). Halliday (1985), Halliday et Hasan (1976) et Mann et Thompson (1988 ; 2001), en revanche, l'approchent selon une perspective textuelle et l'étudient essentiellement à partir de l'unité du texte et des relations sémantiques entre les segments textuels. À la suite, nous aborderons plus en détail l'approche de Charolles pour passer ensuite aux travaux fondés sur l'aspect textuel de la cohérence.

Selon Charolles (1997 : 2, 2006 : 39), c'est donc grâce à la cohérence que les phrases se suivant d'une façon linéaire dans le texte sont comprises comme une unité. Afin

⁵ Pour d'autres travaux et approches, cf. *inter alia* Van Dijk (1972) ; Brown et Yule (1983) ; Sanders & Spooren (1999) ; Jaubert (2005), et aussi la discussion dans Carter-Thomas (2000 : 47).

d'interpréter les phrases, le destinataire cherche à établir des relations entre elles. Comme l'illustre l'extrait suivant, Charolles propose que ce processus d'interprétation repose sur des connaissances linguistiques et extralinguistiques du destinataire et non sur des indices linguistiques du texte.

La reconnaissance de ce qui rend un discours cohérent implique non seulement l'interprétation des éventuelles marques de cohésion qu'il comporte mais encore [...] la mise en œuvre d'opérations inférentielles [...] portant conjointement sur le contenu donné discursif, la situation dans laquelle il est communiqué et les connaissances d'arrière-plan des sujets.

Charolles (1995 : 133)

Dans l'extrait, Charolles souligne la nature sémantique et pragmatique de la cohérence : c'est le contexte du discours et la situation de communication qui guident le lecteur. Dans un autre article moins récent, Charolles (1983) accentue également l'importance de la compréhension de l'intention de l'émetteur dans le texte. Charolles (1983 : 72-73) affirme que, pour comprendre la façon dont une séquence d'énoncés ou de phrases forme une entité cohérente, il faut connaître la raison sous-jacente de leur union. Ainsi, pour appréhender la façon dont les phrases sont reliées dans un texte, il faudrait savoir où et pourquoi le texte est produit. En introduction, nous avons mentionné qu'un roman policier n'est pas organisé de la même façon qu'un article dans le journal *Le Monde*. Dans les deux cas, malgré leurs organisations différentes, le lecteur peut concevoir les textes comme étant cohérents. Pourtant, si l'article du journal *Le Monde* était organisé comme le roman policier, il serait sûrement considéré comme mal organisé. L'interprétation du lecteur dépend donc du contexte qui, en l'occurrence, est le genre discursif du texte.

Le rôle du destinataire est un autre aspect important de la cohérence pour Charolles. Il propose (1983 : 95) qu'un texte n'est pas fondamentalement cohérent ou incohérent mais qu'il est interprété comme tel par le destinataire. Ainsi, il entend que la cohérence dépend de la capacité du destinataire à interpréter les indices présents dans le texte. Par conséquent, selon Charolles (1983 : 81, 1995 : 133), la cohérence ne serait pas un trait physique du discours, ni un aspect structural du texte, mais un principe cognitif guidant le lecteur dans l'interprétation, simplement « a matter of point of view » (« une question de point de vue ») (1983 : 81).

Le rôle du destinataire et l'aspect subjectif de la cohérence sont également discutés par d'autres chercheurs. Hoey, par exemple, constate que « coherence is a facet of the reader's evaluation of a text » (« la cohérence est une facette de l'évaluation du texte par le lecteur ») et qu'elle est donc subjective (1991 : 12). Pourtant, interprétée littéralement, cette perspective se révèle très problématique en ce qu'elle sépare le texte de sa cohérence et, suivant cette logique, tous les « textes » pourraient être interprétés comme cohérents, bien qu'apparaissant illogiques à première vue.

Le fait d'accorder un rôle important au destinataire dans la création de la cohérence a provoqué beaucoup de critiques parce que cette approche risque de minimiser l'importance des indices linguistiques du texte même. Surtout, la présomption par le lecteur du texte comme étant cohérent, présente par exemple dans Brown et Yule (1983 : 223 et passim), a été souvent critiquée. Selon cette supposition, non dépourvue de lien avec la maxime de pertinence chez Grice (1975), le destinataire présume que le « texte » forme une unité intelligible et cherche à tout prix à l'interpréter sous cette lumière. Une remarque souvent avancée contre cette minimisation du rôle des indices linguistiques est l'extrait suivant, annoncé par de Beaugrande et Dressler (1981 : 35; cf. aussi Péry-Woodley 2000 : 13).

Recent debates over the role of the reader point up to the dangers of assuming that text receivers can do whatever they like with a presentation. If that notion was accurate, textual communication would be quite unreliable, perhaps even solipsistic. There must be definitive, though not absolute, controls on the variations among modes of utilising a text by different receivers.

De Beaugrande et Dressler (1981 : 35)

Comme le disent de Beaugrande et Dressler, le lecteur ne peut pas faire ce qu'il veut avec le texte. Bien que le produit du processus d'interprétation puisse varier selon le destinataire, l'étude du texte est impossible si le point de départ n'est pas le texte même. L'interprétation du texte ne peut pas dépendre uniquement du destinataire. Comme le disent de Beaugrande et Dressler (ci-dessus) et Péry-Woodley (2000 : 13), il faut, de quelque sorte, restreindre les interprétations possibles, avoir « des moyens de contrôler les variations d'interprétations ».

La séparation du texte de sa cohérence et l'importance du rôle du lecteur, présents surtout dans les travaux des années 1980, peuvent être perçus comme une réaction extrême aux travaux sur l'unité du texte issus des années 1970 avec un point de départ très différent. Ces travaux sur la « grammaire de texte » examinent la cohérence selon le « bon usage » et reposent sur l'idée « qu'il existe un critère de bonne formation à l'échelle du texte de la même façon qu'il en existe un au niveau de la phrase » (Charolles & Combettes 1999 : 84). Influencés par la tradition générative, ces travaux songent à décrire le texte par des modèles formels, composés de parties inférieures, et préexistants au texte. Ainsi, la fonction de la cohérence dans l'interprétation du texte a été définie comme étant similaire à celle de la syntaxe dans la phrase (cf. p.ex. Van Dijk 1972 ; Petöfi 1971 ; pour une discussion, cf. Charolles & Combettes 1999). Elle définirait comment la suite d'éléments doit être interprétée et comment produire une suite « bien formée », correspondant à une « grammaire de texte ».

De même que la séparation complète du texte de sa cohérence, l'idée d'une grammaire de texte est de nos jours rejetée. Charolles & Combettes (1999), entre autres, considèrent qu'un modèle structural préconstruit ne peut organiser le texte comme la syntaxe organise la phrase. Afin qu'une linguistique textuelle soit possible, comme nous venons de le dire, toute séquence d'énoncés ne peut pas être cohérente mais la cohérence ne peut pas non

plus être prédéterminée. Quoique la cohérence doive, à notre avis, avoir un aspect textuel, elle n'est pas non plus une propriété du texte mais dépend, notamment, du destinataire, de ses connaissances et de la situation de communication (cf. aussi Hoey 1991 : 11-12).

En définitive, la réflexion sur la cohérence de Charolles porte donc surtout sur les aspects pragmatiques et cognitifs de la cohérence. L'aspect textuel de la cohérence ou le fait de définir la cohérence à partir du texte ne lui sont pas centraux. Pour cette raison, les travaux de Halliday (1985), de Halliday et Hasan (1976, 1989) et de Mann et Thompson (1988) complètent bien ceux de Charolles en nous offrant une approche plus globale de la cohérence. Contrairement à Charolles, ces travaux partent du texte et de son unité sémantique : selon Halliday & Hasan (1976 : 2), le texte est une unité sémantique dont les parties sont liées par des relations de sens. La cohérence est une condition inhérente au texte par définition : c'est grâce à la cohérence que « the text hangs together » (« le texte tient tout seul ») (Halliday & Hasan 1989 : 48).

De même que Charolles ci-dessus, Halliday et Hasan (1976 : 23, 1989 : 48) soulignent l'importance du contexte dans la création de la cohérence⁶. Ils le comprennent de deux façons : le contexte de situation, d'une part, et le contexte du texte entourant, de l'autre. Un texte doit être cohérent selon ces deux dimensions pour former une unité. Chaque partie du texte est « at once both text and context » (« à la fois texte et contexte ») (Halliday et Hasan 1989 : 48). Le texte est donc un continuum cohérent de segments qui fournissent un contexte d'interprétation aux parties suivantes et qui, simultanément, font déjà partie intégrante d'un tout.

L'importance du contexte, dans toutes ses dimensions, est présente également dans l'extrait suivant dans lequel Halliday souligne la complexité de la cohérence. Si les segments textuels ne se fournissent pas de contexte (sémantique) approprié, le texte ne peut pas non plus être cohérent.

For a text to be coherent, it must be cohesive ; but it must be more besides. [...] it must be semantically appropriate, with lexicogrammatical realizations to match (i.e. it must make sense) ; and it must have structure.

Halliday (1985 : 318)

Selon Halliday, les segments du texte doivent donc former un contexte approprié pour que le texte puisse avoir du sens et être cohérent. De plus, un texte cohérent doit également avoir une structure. Pourtant, il est important de noter qu'il s'agit d'une structure de nature sémantique et non de règles formelles comme celles de la syntaxe.

Un aspect central de la structure sémantique du texte est les liens sémantiques entre les segments formant l'entité textuel. Ces liens, dénommés relations de cohérence, sont en

⁶ Pour l'importance du contexte dans la création du texte (ou discours) comme un ensemble, voir aussi l'approche de Winter (1982) et Hoey (1983).

particulier au centre de la Théorie de la Structure Rhétorique (RST, Mann & Thompson 1988 ; 2001 ; Taboada et Mann 2006). La RST est une méthode d'exploration de la structure du texte dont l'objectif est de découper le texte en unités puis d'assigner un rôle à chacune de ces unités (Mann & Thompson 1988, 2001). Les rôles sont surtout définis grâce à la réunion des unités linguistiques par des liens sémantiques qui définissent la façon dont elles sont liées. Ces liens portent des étiquettes selon leur contenu sémantique, notamment : *élaboration, concession ou motivation*. Elles s'établissent récursivement de sorte que les unités reliées forment ensemble d'autres unités qui peuvent à nouveau se relier avec d'autres. Ainsi, un texte est une structure hiérarchique composée d'unités et de sous unités.

La RST offre également une perspective textuelle à l'examen de la cohérence, et la définit à partir de la segmentation du texte en unités reliées. Selon Mann & Thompson (2001 : 18), un texte cohérent consiste en parties ayant « un rôle évident et discernable dans le texte et rien [...] ne donne à penser qu'il manque des parties ». Bien qu'ils admettent que cette définition est incomplète du fait d'ignorer des aspects de la compréhension des textes (tels que la structure d'information), nous la trouvons simultanément utile grâce à sa simplicité : un texte cohérent est une unité sémantique dans laquelle nous trouvons tout le nécessaire mais sans superflu pour son interprétation. De plus, la perspective textuelle de la RST de la cohérence convient particulièrement à la présente étude en ce sens que nous aussi souhaitons conserver un regard textuel sur l'organisation textuelle et la cohérence.

3.2 TEXTE VS. DISCOURS

Dans la section précédente, en définissant la cohérence, nous déterminions également ce que nous entendons par *texte*. Nous le considérons donc comme une unité sémantique créée grâce aux relations de cohérence construites entre les éléments du texte par le destinataire.

Dans la section précédente, nous avons aussi ajouté que l'interprétation des relations de cohérence repose sur les connaissances linguistiques et extralinguistiques du lecteur. Or, selon cette définition, le texte peut se confondre avec le concept de *discours*. Par exemple, Charolles (1997 : 2) note seulement que « le discours commence dès que l'on est sorti de ce système [la syntaxe] dont la portée est relativement restreinte ». Par ailleurs, d'autres définitions semblent considérer les deux concepts comme égaux (cf. Chafe 1992 ; McCabe 1998). Le discours peut également être compris comme se référant uniquement à la langue parlée⁷ ou à des genres discursifs : discours de recherche, politique ou journalistique. Afin de distinguer ces deux concepts, nous en clarifierons à la suite les différences et similitudes⁸.

⁷ Voir Brown et Yule 1983 : 6-7, 190, et passim), qui utilisent *le texte* pour se référer « to the verbal record of a communicative act » (« à l'enregistrement verbal d'un acte de communication ») (1983 : 6).

⁸ Voir aussi la discussion dans Carter-Thomas (2000 : 27-30).

Pour Halliday et Hasan (1989 : 48 ; aussi 1976 : 1-2, Halliday 1985 : 318), comme nous l'avons dit, un texte est une unité sémantique qui « hangs together ». Le discours, en revanche, est le processus multidimensionnel dont le texte est le produit et qui, pourtant, possède la même complexité. La distinction se fait donc entre le processus et le produit.

Cornish (1999 : 32-35, aussi 1990 et 2006 et Rastier 1989 : 37) et Péry-Woodley (2000 : 13-14), en revanche, comprennent le texte comme une trace physique du discours, qui alors est la représentation mentale et hiérarchique des séquences d'énoncés perçues et interprétés par le destinataire. Plus précisément, le texte est compris comme « la trace – des signes verbaux et signaux non verbaux – d'un discours ancré dans un contexte » (Péry-Woodley 2000 : 13). De plus, Cornish (1999 : 35, 2006) souligne la sous-détermination des textes : les textes peuvent donner lieu à plusieurs discours selon le contexte. Le discours est donc pour lui « une affaire (re)constructive et [...] hautement probabiliste », une construction mentale et, d'une certaine façon, interprétative qui inclut le texte mais aussi d'autres dimensions.

La hiérarchisation des concepts du texte et du discours est également présente dans la façon dont le discours et le texte sont distingués dans le travail de Adam (1990 : 23 ; cf. aussi Jaubert 2002) ainsi que dans des travaux sur l'anglais (par exemple Virtanen 1992 : 3 ; Enkvist 1984 : 65 ; 1989 : 371-372). Selon cette perspective classique, le discours est le texte plus les conditions de production, tandis que le texte est le discours moins les conditions de production. Le texte est donc aussi un objet inférieur au discours. Pour Adam (*ibid.*), le discours se caractérise comme un acte de discours concret ancré dans la situation. En revanche, le texte « est un objet abstrait résultant de la soustraction du contexte opérée sur l'objet concret (discours) » (Adam 1990 : 23). À cet égard, Adam (*ibid.*) affirme que le discours ne peut pas être étudié uniquement par des méthodes linguistiques. En effet, les conditions de sa production, c'est-à-dire le lieu, les participants et le temps, en font partie intégrante. Ainsi, pour l'examiner dans tous ces aspects, il faut également faire appel à d'autres disciplines, telles que l'histoire et la sociologie. Cette perspective associe donc également le discours à des genres discursifs.

En définitive, la distinction entre le texte et le discours peut survenir de multiples façons. La place du contexte dans le schéma apparaît comme une différence fondamentale dans les définitions que nous venons de discuter : tandis que le texte est, pour Adam (1990), un objet abstrait déchiré de son contexte de production, pour Cornish (1999) et Péry-Woodley (2000) (et aussi Halliday et Hasan 1989 ; Halliday 1985 : xxii-xxiii), il est au contraire ancré dans la situation et concret, la trace physique du discours. Pour Cornish et Péry-Woodley, en revanche, c'est le discours qui demeure plus abstrait en tant que représentation mentale des séquences interprétées.

Malgré ces distinctions, il faut cependant admettre que les deux concepts du texte et du discours sont également utilisés dans des contextes très similaires dans lesquels les

différences discutées ne sont pas évidentes. Dans des travaux plus récents, Adam (1999, 2005) propose également de ne plus dissocier les deux termes. De plus, plutôt que de démontrer simplement l'objet de l'étude, le choix de l'un ou de l'autre terme dans un travail semble souvent être lié à la culture et à la tradition adoptées dans la communauté scientifique de déroulement du travail. Ainsi, dans les pays anglo-saxons, les études sur la linguistique textuelle, dans le sens de *text linguistics* (De Beaugrande et Dressler 1983, Brown et Yule 1983, Coulthard 1994), proposent souvent le texte comme objet d'étude, tandis que les recherches similaires francophones ont tendance à parler de l'étude du discours (voir Ho Dac 2007, Péry-Woodley 2000, Charolles 1997, etc). Le terme discours est également appliqué dans plusieurs travaux anglophones sur l'organisation du discours (Virtanen 1992, Goutsos 1996). La linguistique textuelle francophone, finalement, souligne les aspects argumentatif et énonciatif du texte et s'oriente souvent sur l'analyse des textes littéraires (cf. *inter alia* Adam 1999, Monte 2008, articles dans Jaubert 2005).

Enfin, il reste encore une façon de définir le texte utile dans notre présente étude. Cette définition, proposée par Halliday et Hasan (1976), tend à appréhender le texte par rapport à son fonctionnement. Le texte est perçu comme « a unit of language in use » (« une unité de langage d'usage ») (Halliday & Hasan 1976 : 1) qui a « une tâche à remplir » (Halliday & Hasan 1976 : 1-2, 1989 : 10). Conséquence importante de cette hypothèse : la longueur n'est pas un critère dans la formation du texte. Un seul mot, tel que *ouverte*, peut construire un texte autant qu'il a une fonction, c'est-à-dire une tâche à remplir. Comme l'indique Péry-Woodley (2000 : 14), le fait d'inclure cet aspect fonctionnel dans la définition du texte distingue notre approche de la linguistique de l'interphrastique : le texte peut ne comporter qu'un seul mot.

Pour conclure cette discussion, nous comprenons, dans cette étude, le texte comme une unité à trois dimensions. Premièrement, nous le voyons comme unité de langue utilisée dans un contexte dont l'interprétation repose sur les relations de cohérence construites par le destinataire. Dans un deuxième temps, suivant Halliday & Hasan, Cornish et Péry-Woodley, nous le considérons aussi comme la trace textuelle d'un processus plus complexe englobant également des dimensions extralinguistiques. Cette trace textuelle forme notre objet d'étude. Ainsi, et bien que nous y fassions référence de temps à autre, les processus sous-jacents de cette trace sont secondaires à nos yeux et, par principe, échappent à notre intention. Finalement, nous entendons aussi par texte une unité fonctionnelle, soit une unité linguistique ayant une fonction, une tâche à remplir (Halliday & Hasan 1976 : 1-2, 1989 : 10). Pour nous, un texte peut donc se composer d'un seul mot, du moment qu'il se caractérise par une fonction.

3.3 COHESION

En examinant la cohérence, nous avons avancé qu'elle ne reposait pas sur des indices linguistiques ; par exemple dans le cadre de la RST, une relation de *motivation* (cf. ci-dessus) peut relier deux unités textuelles sans marqueur linguistique. Pour autant, la cohérence **peut** être étudiée par des marques linguistiques, autrement dit des indices de cohésion. Ces « ressources linguistiques pour créer du texte » (Péry-Woodley 2000 : 15) incluent des indices variés d'origine lexicale et grammaticale.

La description classique des indices de cohésion est *Cohesion in English* de Halliday et Hasan (1976) qui les examine de façon systématique pour les distinguer en cinq catégories différentes : la référence, l'ellipse, la substitution, la conjonction et la cohésion lexicale. Cette classification conforme la base de notre analyse, notamment, parce qu'elle est facile à opérationnaliser et parce que les distinctions entre les catégories sont claires. De plus, notre analyse englobe également d'autres indices linguistiques pouvant avoir des fonctions cohésives, tels que le positionnement initial des adverbiaux dans la phrase (cf. 3.4.4) et les amorces (cf. 4.2.4) qui peuvent précéder les séries linéaires. Pour une autre approche sur les indices de cohésion, voir par exemple le modèle de Winter (1982) développé également par Hoey (1983).

Suivant les termes de Halliday et Hasan (1976), les catégories les plus importantes pour notre thèse sont la référence, la conjonction et la cohésion lexicale. Ce sont surtout la conjonction (par ex. *premièrement... ensuite... mais... pour conclure*) et la référence (par ex. *le premier... le second... l'autre...*) qui peuvent introduire les items d'une série linéaire. La cohésion lexicale, c'est-à-dire les différentes formes de répétition, peut contribuer à l'identification des items mais ne peut pas les introduire de façon explicite (cf. 5.2.2). Il est à noter que, selon les termes de la linguistique systémique fonctionnelle (Halliday et Hasan 1976, 1989 ; Halliday 1985 ; Martin 1992), *la conjonction* se réfère au système des relations sémantico-logiques entre les phrases et les segments textuels. Elle est signalée par plusieurs marqueurs différents et ne doit donc pas être confondue avec les conjonctions de coordination ou de subordination, des groupes beaucoup plus restreints.

En définitive, si un texte peut être cohérent sans indices de cohésion, il est pertinent de se demander quelles sont les raisons de leur usage. Cette question a été posée, entre autres, par Mauranen (1993 : 161-163), qui note que ces marqueurs, bien qu'ils n'ajoutent rien au texte, le rendent plus clair et en facilitent ainsi la lecture. Cette fonction a aussi été signalée par Goutsos (1996 : 505) et Jackiewicz (2005 : 105), qui relèvent que le marquage devient nécessaire quand le découpage ne va pas de soi. Charolles (1995 : 125-126, 1997 : 2), pour sa part, considère que la fonction des marqueurs de cohésion est le guidage du destinataire dans le processus d'interprétation des séquences comme une unité. Les tests empiriques ont également prouvé que les indices de cohésion accélèrent la lecture des segments textuels (Bestgen et Vonk 1995 ; Sanders et Noordman 2000).

Charolles (1995 : 126, 129-130) signale que les marques de cohésion s'avèrent souvent problématiques du fait de leur sous-détermination et de leur polyfonctionnalité ; les instructions qu'elles fournissent sont insuffisantes pour aboutir seules à une interprétation souhaitée. L'expression anaphorique prototype *il*, par exemple, nécessite encore une interprétation adéquate par le destinataire afin que le texte soit cohérent. Ainsi, Charolles (*id.* : 130) note que ces marqueurs sont uniquement des instructions interprétatives n'offrant aucune garantie : « l'occurrence d'un connecteur et/ou d'une anaphore et/ou d'une quelconque autre marque de cohésion n'est en effet ni une condition nécessaire ni une condition suffisante pour que, par exemple, une suite de deux énoncés paraisse former une séquence cohérente ». Cette insuffisance est également notée par Halliday et Hasan (1976 : 5, 12), qui soulignent la nature relationnelle de la cohésion en constatant que ce n'est pas la présence d'un marqueur qui crée la cohérence mais la relation sous-jacente. Une occurrence isolée du mot *John*, par exemple, ne crée pas de cohésion lexicale, alors que c'est le cas pour la répétition de ce mot et le lien entre le mot original et la forme répétée.

3.4 INDICES DE COHESION

Les indices de cohésion guident le lecteur dans l'interprétation du texte et signalent (ou sont) des relations sémantiques entre les phrases ou segments textuels plus vastes. Dans la section précédente, nous avons déjà mentionné que les trois types d'indice de cohésion les plus importants à nos yeux, selon Halliday et Hasan (1976), sont la conjonction, la référence et la cohésion lexicale. Les indices linguistiques pouvant adopter ces fonctions sont fréquemment étudiés selon maintes perspectives. Dans cette section, nous les examinerons plus en détail.

3.4.1 CONJONCTION, CONNECTEURS ET ORGANISATEURS TEXTUELS

Les particules définies comme conjonctions par Halliday et Hasan (1976) sont examinées dans des études sous des dénominations variées : connecteurs, organisateurs textuels, marqueurs discursifs... Dans cette thèse, nous optons pour le terme *connecteur* comme dénomination globale de tous ces marqueurs. Lundquist (1980 : 47 ; cf. aussi Charolles 1997 : 2-3) en fournit la définition suivante :

[Les connecteurs] sont des particules de liaison, telles que « ainsi », « là », « donc », « pour cette raison », qui jouent le rôle de « passeurs » d'une phrase à une autre. Les connecteurs établissent une cohérence entre des phrases plutôt qu'entre des éléments de phrases, ils indiquent des relations entre des raisonnements, entre des thèses exprimées dans le texte.

Lundquist (1980 : 47)

Adam (1990 : 141), pour sa part, examine les connecteurs surtout dans une perspective textuelle et inter(trans)phrastique en soulignant les aspects pragmatiques et les fonctions énonciatives de ces marqueurs. Pour l'ensemble des connecteurs, il propose une

classification selon les fonctions des marqueurs. Ainsi, les organisateurs énumératifs accompagnent l'énumération (*ensuite, troisièmement*, cf. Turco et Coltier 1988 ; Schneuwly *et al.* 1989), les connecteurs de reformulation contribuent au marquage de la reformulation ou de la structuration du discours (*bref, alors* examinés dans Adam et Revaz 1989) et les connecteurs argumentatifs orientent le discours d'une façon argumentative et proposent de traiter un contenu propositionnel comme argument, conclusion ou contre argument. Ces derniers, affectés d'une fonction de marquage fort de prise en charge énonciative, incluent des connecteurs tels que *mais, car, et puisque* (entre autres Ducrot *et al.* 1980 ; Anscombe et Ducrot 1997).

Dans son analyse, Adam (1990) se concentre sur l'observation des connecteurs à un niveau local et pragmatique du discours. Une autre perspective, adoptée entre autres par Halliday et Hasan (1976, Halliday 1985), Lundquist (1980), Schneuwly *et al.* (1989) et Martin (1992) est de nature sémantique et propose d'étudier les connecteurs à partir du lien sémantique qu'ils signalent. Par exemple, Halliday & Hasan (1976 : 227) définissent que les connecteurs servent de « specification of the way in which what is to follow is systematically connected to what has gone before » (« spécification de la manière selon laquelle ce qui suit est systématiquement connecté à ce qui précède »). Cette approche sémantique est également celle que nous adoptons. En effet, elle s'ajuste le mieux à notre but d'observation du rôle des connecteurs dans la structuration du texte comme une unité sémantique.

Les connecteurs sont souvent définis comme des particules dont la polysémie complique l'analyse et la classification. Adam (1990 : 142), par exemple, note que les fonctions des connecteurs dépendent des types de séquences textuelles dont ils font partie. Il indique qu'« un connecteur comme ALORS, par exemple, n'a ni la même fonction ni la même valeur dans une séquence narrative et dans une séquence argumentative ». Cette polysémie s'observe également dans nombre d'études sur les marqueurs individuels, traitant de leurs fonctionnements divers (Pacelli Pekba 2003 ; Mosegaard Hansen 2005 ; Giraud 2008 ; Bras et Le Draoulec 2009 ; Vergez-Couret 2009, etc.).

La polysémie des connecteurs s'avère également épineuse pour les travaux tendant à classer les connecteurs ou les relations de cohérence selon leur nature sémantique. Le travail de Martin (1992 : 170-175, 179 et passim) vise à une description du système de la conjonction et une classification des relations logico sémantiques présentes dans un cadre systémique fonctionnel. Sa classification prend comme point de départ quatre types de relations de base : additives, comparatives, temporelles et conséquentielles. Admettant que toutes les classifications ont leurs atouts et faiblesses, Martin (*ibid.*) résout le problème en plaçant les mêmes connecteurs dans plusieurs systèmes si leurs contenus sémantiques l'exigent. Dans l'annotation du Penn Discourse Treebank (Prasad *et al.* 2008), pour les relations indiquées par un connecteur explicite, les annotateurs marquent le sens dans lequel le connecteur est utilisé.

Au lieu de mener une analyse minutieuse des contenus sémantiques des marqueurs, Mann & Thompson (1988, 2001, 2006) et les travaux sur la SDRT⁹ adoptent une solution alternative à l'étude des relations de cohérence. Ces travaux examinent seulement des relations sous-jacentes des formes de surface. La RST (Mann et Thompson 1988 : 256-257) distingue les relations « subject matter » (« sujet ») des relations présentationnelles. Les premières, cherchant à faire identifier la relation en question par le lecteur, comprennent des types tels que *élaboration*, *contraste*, *évaluation* et *condition*. Les secondes, tendant à renforcer l'envie d'agir du lecteur, intègrent des types tels que *motivation*, *concession*, et *évidence*. L'action souhaitée peut donc être, par exemple, le désir d'agir ou l'acceptation d'un fait. La SDRT part également des relations sémantiques entre les unités et non des indices linguistiques. Dans ce modèle, on distingue les relations rhétoriques fondées sur les fonctions rhétoriques des segments reliés (telles que *contraste*, *narration*, *raison*) et les relations de cohérence qui fonctionnent à partir des éventualités présentées dans les unités (telles que *cause*).

Quels sont les types de mots susceptibles de fonctionner comme connecteurs ? En français, plusieurs types d'adverbes, de conjonctions et de locutions conjonctives peuvent fonctionner comme connecteurs. Schneuwly & al. (1989) proposent des critères de surface spécifiques que les connecteurs, ou les organisateurs textuels, doivent remplir (pour une autre définition et critères, cf. Rossari *et al.* 2001). Ces marqueurs doivent ainsi : (1) joindre des structures propositionnelles en précisant leur insertion dans le cotexte ou dans le contexte de production, (2) être placés hors de la syntaxe de la phrase et (3) ne pas être soumis aux phénomènes d'accord. Schneuwly *et al.* (1989 : 43) concluent également que deux types d'organisateur textuels peuvent être distingués : ceux qui le sont toujours (conjonctions de coordination et de subordination) et ceux qui le sont lorsqu'ils sont placés hors de la structure syntaxique de la phrase (*à sept heures, plus tard...*)¹⁰.

Pour l'anglais, Halliday (1985 : 303) constate que les relations conjonctives peuvent être exprimées entre autres par des groupes adverbiaux ou prépositionnels et par des conjonctions, en général dans la position initiale de la phrase. De plus, l'expression doit pouvoir prendre cette fonction seule sans être accompagnée par ex. par un référent (cf. par ex. Halliday & Hasan 1976 : 226-233). Ces expressions incluent des marqueurs comme *accordingly*, *then*, *anyway* et *as a result*.

Les définitions des connecteurs de Schneuwly *et al.* (*ibid.*) et de Halliday et Hasan sont toutes deux exhaustives et comprennent un large éventail d'expressions. Nous suivons la définition de Schneuwly *et al.* et considérons comme connecteurs tous les marqueurs qui, selon leur schéma, le sont toujours. Cette définition nous convient bien parce qu'elle offre des critères spécifiques et facilement applicables pour l'analyse des connecteurs. De plus,

⁹ *Segmented Discourse Representation Theory*, la théorie des représentations discursives segmentées (Asher 1993 ; Asher & Lascarides 2003), est un modèle d'organisation du discours qui se base sur la sémantique formelle.

¹⁰ Ce groupe inclut également les cadratifs que nous discutons en section 3.4.4.

cette définition présente un aspect important : l'exclusion des connecteurs des anaphoriques et autres expressions, susceptibles de fonctionner comme sujets de la phrase. Un autre atout du travail de Schneuwly *et al.* tient à la vision hiérarchique du texte et de son organisation. Comparé à la conception générale des connecteurs comme dénotant la connexion entre deux segments, Schneuwly & al. (1989) opposent les connecteurs reliant des structures propositionnelles et les *organisateur textuels* structurant le texte de façon hiérarchique à différents niveaux, prenant ainsi en compte plusieurs dimensions du texte. Selon eux, la fonction des organisateurs textuels est en fait double : le *liage* assure la connexion des unités et l'*empaquetage* l'organisation hiérarchique des unités. Le texte résultant peut être perçu comme une structure hiérarchique composée d'unités et de sous unités reliées.

3.4.2 REFERENCE ET ANAPHORE

Il y a relation d'anaphore entre deux unités A et B quand l'interprétation de B dépend crucialement de l'existence de A, au point qu'on peut dire que l'unité B n'est interprétable que dans la mesure où elle reprend – entièrement ou partiellement – A. Cette relation existe quand B est un pronom dont la référence virtuelle n'est établie que par l'interprétation d'un N'' que le pronom répète. Elle existe également quand B est un N'' dont le caractère défini – c'est-à-dire le caractère identifié du référent – dépend exclusivement de l'occurrence, dans le contexte, d'un certain N'' [...].

Milner (1982 : 18)

Cette définition de l'anaphore de Milner en présente la conception traditionnelle : l'anaphore est perçue comme une relation asymétrique entre deux unités linguistiques, la deuxième nécessitant la récurrence à son antécédent pour être interprétée (cf. aussi Ducrot et Todorov 1972 : 358-359 ; Corblin 1985 : 126, 1995 : 31-34 ; Charolles 2002 : 245). Typiquement, l'anaphore est illustrée par le pronom *il*, ainsi dans l'Exemple 4 ci-dessous.

Exemple 4

Le chien dort. **Il** est grand.

Ce pronom remplit clairement les critères présentés ci-dessus. En effet, il ne peut pas être interprété par lui-même mais a besoin d'un antécédent pour être résolu. De plus, le pronom *il* est coréférentiel avec son antécédent, autrement dit il se réfère au même objet que *le chien*. Les corrélatifs anaphoriques ordinaux *le premier – le second* et *l'un – l'autre* - centre d'intérêt de notre thèse - partagent ces mêmes propriétés ; le lecteur ne peut pas savoir ce qu'ils désignent sans faire appel à leur point de référence placé à proximité (cf. Schnedecker 2000, 2006).

Sur un plan plus général, nous pouvons comprendre l'anaphore comme une sous-classe des expressions référentielles. Outre les anaphoriques dans une relation asymétrique avec leur antécédent, ce groupe englobe les cataphoriques et les expressions référentielles autonomes. En ce qui concerne les cataphoriques, elles sont autrement similaires aux

anaphoriques mais l'ordre de l'expression référentielle et de la source est y inversé, car la première précède la seconde : *il – le chien*. Quant aux expressions référentielles autonomes, elles se réfèrent au même objet que leur antécédent mais n'en dépendent pas pour être interprétées. Contrairement aux anaphoriques, ces expressions sont donc sémantiquement autonomes (cf. Corblin 1995 : 177).

Les expressions référentielles peuvent également être examinées selon le lieu de leur objet de référence. Halliday et Hasan (1976) distinguent les références *exophoriques*, se référant à la situation d'énonciation, et les références *endophoriques*, renvoyant au texte. Les exophoriques, aussi dénommées déictiques, sont typiquement des pronoms de première ou deuxième personne (*je, vous*) ou des constructions se référant à la situation (*maintenant*). Pourtant, d'autres expressions, comme le pronom personnel *il*, présentent les deux usages, déictique et endophorique. Par exemple, le *il* dans *Il est grand* de l'exemple ci-dessus pourrait, utilisé dans une phrase indépendante, faire référence à une personne dans la situation de communication (cf. aussi Reboul 1994 : 349-352 ; Corblin 1995 : 13-15.).

Dans les cas prototypiques des expressions anaphoriques, tel que le pronom *il*, la détermination dans le texte de l'antécédent ne pose pas de problèmes. La présence de cet antécédent et sa nature coréférentielle sont également souvent considérées comme les propriétés déterminantes de l'anaphore, comme l'extrait de Milner (1982) au début de cette section le montre. Dans des cas moins prototypiques, ces deux propriétés ne sont pourtant pas aussi évidentes¹¹. Considérons les exemples suivants, empruntés à Apothéloz (1995 : 23, 28, citant Karttunen 1969) et à Corblin (1995 : 39) :

Exemple 5

Un jeune homme soupçonné d'avoir détourné une ligne téléphonique a été interpellé il y a quelques jours par la police à Paris. Il avait « utilisé » la ligne de ses voisins à destination des Etats-Unis pour un montant d'environ 50000 F. **Le bavard** a été déferé devant le parquet. (Apothéloz 1995 : 23)

Exemple 6

L'homme qui a donné son salaire à son épouse a été plus sage que l'homme qui l'a donné à sa maîtresse (Apothéloz 1995 : 28, citant Karttunen 1969)

Exemple 7

Un professeur agrégé m'a parlé. **Ce professeur** agrégé est un ignorant. (Corblin 1995 : 39)

Exemple 8

Un professeur agrégé est venu. **Le professeur** agrégé est un ignorant. (Corblin 1995 : 39)

Tous ces exemples peuvent donc être considérés comme des anaphores. Or, dans le cadre d'une analyse plus en détail, nous notons que l'interprétation des anaphoriques ne dépend

¹¹ Sur la complexité des référents, voir le travail classique de Karttunen (1969).

pas, dans tous les cas, uniquement de l'antécédent. De plus, dans l'Exemple 5 et dans l'Exemple 6, l'identification même de l'antécédent peut s'avérer problématique bien qu'ils ressemblent autrement aux cas d'anaphore plus typiques. Il importe donc de réfléchir sur le statut de l'antécédent.

Dans l'Exemple 5, l'interprétation du *bavard* résulte du segment textuel entier et des connaissances du monde réel du lecteur (50 000 F est une somme importante et un appel aux États-Unis coûte cher). Ainsi, « l'antécédent » *un jeune homme* ne définit que partiellement le mot repris. De même l'antécédent de l'Exemple 6 est problématique car, *stricto sensu*, l'anaphorique *l'* n'a pas d'antécédent dans le texte. Bien que le mot *salaire* soit mentionné dans le texte, il ne se réfère pas à la même somme d'argent que celle désignée par le *l'*. Les points de référence de ces deux mots sont différents. Pour distinguer ce type d'antécédent de ceux qui le sont *stricto sensu*, Apothéloz (1995 : 28) fait la différence entre les anaphoriques *coréférentielles*, qui se réfèrent au même objet que leur antécédent, et les anaphoriques *co-signifiantes*, qui réfèrent aux objets du même type que leur antécédent.

Contrairement aux deux premiers exemples, la dépendance sémantique des reprises illustrées dans l'Exemple 7 et dans l'Exemple 8 l' fait l'objet de nombreux débats (cf. par exemple Milner 1982 ; Kleiber 1990 ; Corblin 1995 : 49-80 ; Apothéloz 1995 : 55-71). Corblin (1995 : 40), tout en examinant les différences entre ces deux constructions, décide que c'est « la capacité à faire reconnaître un terme comme antécédent » qui rend possible l'usage anaphorique de ces deux types de syntagmes (cf. aussi Kleiber 1990). Milner (1982 : 25), en revanche, considère que les constructions démonstratif+nom sont compréhensibles même sans référence à un antécédent. En effet, l'énoncé dans lequel la construction apparaît contient déjà tout le nécessaire pour son interprétation¹². A l'inverse, dans les constructions défini+nom, l'identification de l'antécédent se ferait grâce aux conditions décrites dans le groupe nominal dont il fait partie.

Le rôle de l'antécédent dans l'interprétation de l'anaphorique n'est donc pas évident dans tous les cas. Par conséquent, l'anaphore mérite une définition plus détaillée qui corresponde mieux à la réalité. Cornish (1999 : 42) propose une définition opérationnelle, selon laquelle c'est l'anaphorique (soit l'élément repris) et « its immediate utterance and predicational context » (« son énonciation immédiate et son contexte prédicationnel ») qui définissent l'interprétation et non seulement l'antécédent. Au lieu d'être un processus purement textuel, l'interprétation de l'anaphorique serait ainsi soutenue par le contexte textuel et par les connaissances du contexte du discours du lecteur.

Afin de différencier les antécédents « de surface » et les segments textuels qui en réalité contribuent à l'interprétation de l'anaphorique, Cornish (1999 : 41-43) propose de distinguer pour eux deux fonctions. La première fonction serait celle de *déclencheur*

¹² Encore, cf. Tesnière (1959) pour une opinion opposée. Voir aussi la discussion sur la saturation sémantique et référentielle dans Moeschler et Reboul (1994 : 360-372).

d'antécédent (antecedent trigger), soit le segment traditionnellement défini comme l'antécédent. Or, pour Cornish (*ibid.*), il s'agit uniquement d'un segment purement formel, insuffisant à lui seul pour interpréter l'anaphorique. Ainsi, *un jeune homme* dans l'Exemple 5 serait un déclencheur d'antécédent ; à lui seul, il ne permet pas d'interpréter le lien entre *un jeune homme* et *le bavard*. La deuxième fonction de l'antécédent, celle de *l'anaphore*, est réservée, en revanche, à l'entité textuelle permettant l'interprétation de l'anaphorique dans son contexte textuel et discursif. Cette fonction rend également possible l'interprétation du *bavard* dans l'Exemple 5. Cornish (*id.* : 42) conclut que cette distinction facilite la compréhension du lien entre l'anaphorique et l'antécédent comme une relation dynamique. Ce dynamisme autorise les changements dans cette relation et l'interprétation des anaphoriques, même dans les cas où elle ne constitue pas une reprise identique de l'antécédent mais doit être interprétée dans le contexte.

La perte d'importance de la distinction traditionnelle entre la référence anaphorique et la référence exophorique, ou déictique, est une conséquence notable de l'importance attribuée au contexte textuel et discursif dans le processus d'interprétation de l'anaphorique. Cornish (1999 : 69) constate que cette distinction est sans importance. En effet, dans les deux cas, c'est le contexte textuel et discursif qui définit l'interprétation des référents et non le positionnement de l'objet de référence dans le texte précédent ou dans le monde réel. De plus, Cornish (*id.* : 52-53) indique que ces usages ne sont pas toujours facilement identifiables et qu'il s'agit davantage d'un continuum entre les usages déictique et anaphorique que de groupes distincts¹³. Cette opinion est soutenue par Corblin (1995 : 17) qui note qu'« il est d'ailleurs facile de montrer que loin de s'exclure, les deux usages sont dans beaucoup de situations indistinguables ».

En définitive, à la définition de l'anaphore de Milner discutée au début de cette section peuvent donc s'ajouter maintes spécifications. Grâce à elles, l'anaphore se présente comme un phénomène plus complexe et varié. Cette variation autorise également la distinction d'un nombre de types d'anaphores, facilitant ainsi leur analyse. Nous illustrons, à la suite, les cas les plus fréquents.

Une première distinction, classique, s'établit entre les anaphores *fidèles* et *infidèles* discutée, entre autres, dans Apothéloz (1995 : 36-38), Riegel *et al.* (1994 : 614). Les anaphores fidèles sont des reprises d'un SN indéfini par un SN défini ou démonstratif, répétant le nom du syntagme : *un chien – ce / le chien*. Une anaphore infidèle, en revanche, ne répète pas le nom de l'antécédent mais présente souvent son synonyme ou hypéronymes : *un chien – cet / l'animal*. Cas particulier des anaphoriques infidèles, *l'anaphore conceptuelle* est définie par Moirand (1975) comme une reprise qui se distingue formellement de sa source mais en présente « l'objet de pensée », un SN lié à un démonstratif ou à l'expression *un tel / une telle*. Souvent, il s'agit d'une nominalisation d'un verbe, telle que *décider – décision* ou *mettre en liberté – libération*, qui sert « à

¹³ La distinction entre les expressions se référant au texte et celles se référant au monde extérieur est également centrale dans le métadiscours, cf. 2.2.2.

paraphraser un énoncé qui précède et permet un enchaînement dans l'expression du scripteur ainsi que dans le décodage, la compréhension du lecteur » (Moirand 1975 : 65)¹⁴.

Le deuxième type d'anaphore, *l'anaphore associative*, est plutôt abstrait et ne reprend pas directement un antécédent. Selon la définition traditionnelle de Blanche-Benveniste et Chervel (1966 : 31-32), cette anaphore relie des mots ou syntagmes dans une relation partie-tout au sein de laquelle le premier objet contient le second, c'est-à-dire l'anaphorique. De plus, Apothéloz (1995 : 40) précise que cette anaphore entend son antécédent comme étant déjà connu, bien qu'il n'ait pas été indiqué auparavant. Deux exemples typiques et généralement acceptés de l'anaphore associative sont *un village – l'église* (Kleiber 1992) et *un stylo – la plume* (Fradin 1984). Ces cas peuvent également être considérés comme des anaphores infidèles.

Pourtant, d'autres exemples plus complexes des anaphores associatives ont engendré une discussion ainsi que deux définitions différentes de ces reprises. Kleiber *et al.* (1994 : 8-9), notamment, distinguent une conception étendue et une conception étroite de cette relation. Selon la conception étendue, toutes les anaphoriques, c'est-à-dire les expressions identifiées par ce qui est exposé avant dans le texte, non coréférentielles sont des anaphores associatives. Cette définition est adoptée entre autres par Berrendonner (1994) et Apothéloz (1994, 1995). La conception étroite, en revanche, est beaucoup plus restrictive et n'accepte pas comme anaphoriques associatives les anaphores infidèles incluant un démonstratif et un pronom indéfini qui, selon la conception alternative, peuvent appartenir à cette catégorie. Ainsi, dans la pratique, Kleiber *et al.* (1994 : 8) expliquent que, selon la définition étendue, la paire *un village – cette église*, par exemple, est une anaphore associative, tandis que la conception étroite accepte uniquement la paire *un village – l'église*. Kleiber *et al.* (1994 : 8-9) précisent également que la définition étroite est plus répandue et peut ainsi devenir normative.

Dernier type d'anaphore examiné, *l'anaphore résomptive* que certains classifient également dans le même groupe que les anaphores conceptuelles (ainsi Charaudeau et Maingueneau 2002 : 49, cf. aussi Auricchio *et al.* 1995). Étudiant ces expressions dans le texte narratif du XV^e siècle, Guillot (2006 : 298) les définit comme une reprise condensée et résumée du contexte précédant sous la forme d'un tout unifié. La forme de cette anaphorique est une nominalisation précédée par un déterminant, tel que *le* ou *ce*. Un exemple illustratif en est la clôture des séries linéaires : *Le premier chapitre... Le second chapitre... Ces sections...* Comme le constate Guillot (2006 : 301), ces expressions signalent les frontières des séquences et se révèlent ainsi un outil utile dans la structuration du texte. C'est également ce fonctionnement qui nous a conduit à garder celles-ci et les

¹⁴ Comme nous l'avons déjà remarqué, le statut anaphorique des anaphores fidèles, infidèles et conceptuelles peut être mis en question car les travaux divergent sur l'indépendance ou la dépendance sémantique de ces reprises. Moirand (1975 : 62), afin de les analyser comme anaphoriques, adopte une approche vaste et, suivant Maillard (1974), les définit comme un fragment énonciatif qui se réfère à un énoncé antérieur dans le texte.

anaphores associatives en tant que groupes séparés. Les propriétés distinctives sont principalement la fonction synthétisante des anaphores résomptives, permettant la reprise de plusieurs éléments sous un seul et, comparée à la conception étroite des anaphores associatives, la possibilité d'être introduite par *ce* ou même par *un tel*.

En définitive, l'anaphore est donc un phénomène textuel complexe qui relie des mots et des segments textuels et contribue ainsi à la cohésion et à la cohérence. Outre le fait de relier des paires de mots, les anaphoriques et autres expressions référentielles peuvent également se combiner pour former des *chaînes de référence*, c'est-à-dire « des suites d'expressions d'un texte entre lesquelles l'interprétation construit une identité référentielle » (Corblin 1985 : 123, voir aussi Charolles 1988 ; Moeschler et Reboul 1994 : 461 ; Suomela-Salmi 1997). Ces constructions se composent d'au moins trois expressions coréférentielles (Schneidecker 1997 : 9-10) : *Un chien... Il... Ce chien...* Souvent, comme le mentionne Corblin (1995 : 169, 174-175 et aussi Suomela-Salmi 1997), se combinent dans ces chaînes des références hétérogènes, comme les pronoms personnels, les démonstratifs et les syntagmes nominaux définis. Ainsi, une partie des références d'une chaîne peut correspondre à des anaphoriques tandis que d'autres ne sont que des coréférentiels. *Les chaînes anaphoriques*, en revanche, sont composées uniquement d'expressions anaphoriques.

3.4.3 COHESION LEXICALE

La cohésion lexicale (étudiée et développée également par Hoey 1991, cf. aussi Tanskanen 2006) est la dernière catégorie d'indices de cohésion définie par Halliday et Hasan (1976) que nous examinons. Dans sa forme la plus évidente, la cohésion lexicale s'établit entre deux items lexicaux répétés dans le texte, tels que *chapitre... chapitre...* Pourtant, des relations lexicales de ressemblance et d'opposition plus complexes peuvent également avoir lieu, telles que *chapitre... thèse... et thèse... travail*. Comme le notent Halliday et Hasan (1976 : 285), la cohésion lexicale peut relier « any pair of lexical items that stand to each other in some recognizable lexicosemantic relation » (« une paire quelconque d'items lexicaux liés entre eux par une relation lexico sémantique reconnaissable ») (cf. aussi Halliday & Matthiessen 2004 : 570-571).

Patterns of Lexis in Text par Hoey (1991) continue le travail de Halliday et Hasan (1976) sur la cohésion lexicale et l'examine comme une propriété organisatrice des textes. Dans une perspective textuelle, ce travail traite essentiellement des liens de répétition s'établissant entre les items lexicaux d'un texte. Grâce à ces relations, un texte devient un réseau de mots à partir duquel l'organisation du texte peut être étudiée.

Hoey (1991 : 7-8) indique que la cohésion lexicale diffère des autres indices de cohésion en ce qu'elle est d'abord une relation lexicale et, seulement ensuite, une relation textuelle. Autrement dit, c'est uniquement la relation lexicale identifiée dans le texte par le lecteur qui permet la création de la relation textuelle contribuant, pour sa part, à la cohérence.

Cette relation textuelle est pourtant très importante. Notant qu'il s'agit d'un moyen dominant de l'établissement du texte, Hoey (*id.* : 10) indique que « the study of the greater part of cohesion is the study of lexis, and the study of cohesion in text is to a considerable degree the study of patterns of lexis in text » (« l'étude de la majeure partie de la cohésion est celle du lexique et l'étude de la cohésion dans le texte est, à un degré considérable, celle du lexique dans le texte »). Hoey (*id.* : 9-10) justifie cet argument en constatant que, dans une analyse de Halliday & Hasan (1976), la répétition couvrait en tout 48 % des relations cohésives.

Dans certains cas, la cohésion lexicale peut se confondre avec l'anaphore ou les expressions référentielles en générale. La différence la plus notable entre ces deux catégories est que la cohésion lexicale ne nécessite pas de relation coréférentielle (ou co-significative) entre les éléments liés. De plus, selon Halliday et Hasan (1976 : 284) et Halliday (1985 : 311), la cohésion lexicale est surtout une relation entre deux formes équivalentes qui cohèrent aussi quand elles ne sont pas coréférentielles. L'anaphore, en revanche, relie plutôt des contenus sémantiques, comme le démontre, par exemple, l'existence des anaphores associative et conceptuelle, dans lesquelles les formes reprises peuvent différer de leur source. Certes, Halliday & Hasan (1976 : 278-279, 282 ; Halliday 1985 : 310-311) constatent également qu'il n'est pas toujours aisé de distinguer les expressions référentielles de la cohésion lexicale et que, par ailleurs, les deux types d'indices sont souvent utilisés ensemble. Par exemple, nous pouvons interpréter la paire *un chapitre... le chapitre* de plusieurs façons, notamment comme une répétition du mot *chapitre*, comme une construction combinée de répétition et d'une expression référentielle (*un...le*) ou comme une expression anaphorique.

Outre la répétition d'un même mot, la cohésion lexicale peut, par exemple selon Halliday et Hasan (1976 : 278-284), concerner la répétition d'un (quasi-)synonyme ou d'un hypéronyme. Ainsi, la seconde occurrence de la paire peut être identique (*chien – chien*), inclusive (*chien – animal*) ou exclusive (*chien – chat*) par rapport à la première. Hoey (1991 : 52-68) propose également une typologie plus détaillée des différents types de répétition et en distingue quatre types. *La répétition simple* relie deux items lexicaux qui diffèrent uniquement au niveau d'un paradigme grammatical fermé : *chien – chiens*. *La répétition lexicale complexe*, en revanche, relie aussi deux items ayant le même morphème lexical mais différant au niveau de la forme ou de la fonction grammaticale : *médecin – médecine*. *La paraphrase simple* se produit quand un item lexical remplace un autre sans en modifier le contenu sémantique, ainsi la paire *livre – volume*. Il peut donc s'agir, par exemple, des synonymes. Enfin, le dernier type de répétition, *la paraphrase complexe*, est un type très vaste que Hoey (*ibid.*) analyse dans trois cas. D'abord, la paraphrase complexe peut s'établir entre les antonymes, tels que *chaud – froid*. Un deuxième cas de paraphrase complexe relie les mots B et C quand A est une répétition complexe de B (*travail – travaillons*) et B une paraphrase simple de C (*travail – emploi*). Dans ce cas, B & C (*travaillons – emploi*) seraient liés par un lien de paraphrase complexe. Il s'agit donc avant tout du produit de deux autres liens qui en créent un troisième. Enfin, le dernier type de

paraphrase complexe est similaire au deuxième et peut survenir même si B est absent du texte, sous condition qu'il puisse s'agir d'une paraphrase simple de A dans son contexte. Ainsi, nous pouvons relier *travaillons* et *emploi* même lorsque *travail* est absent.

3.4.4 ENCADREMENT DU DISCOURS

Les marques de cohésion sont souvent considérées comme une classe fermée, composée des cinq types d'indices déterminés par Halliday et Hasan (1976). Pourtant, Charolles (1997, 2005) propose d'en ajouter un nouveau à ces types d'indice, celui d'*encadrement du discours*. Cette fonction peut être attribuée aux expressions détachées à l'initial de la phrase et marquer « que plusieurs unités doivent être traitées de la même manière relativement à un critère (plus ou moins) spécifié par ces expressions » (Charolles 1997 : 3). Ces expressions peuvent étendre leur zone d'influence au-delà de la phrase dont elles font syntaxiquement partie et ainsi contribuer à l'organisation du texte et à la cohérence.

Le fonctionnement des adverbiaux en position initiale dans l'organisation du texte et du discours est également examiné selon d'autres perspectives, entre autres par Virtanen (1992, 2004), Chafe (1984) et Bestgen et Vonk (2000) (voir aussi l'introduction à cette thématique par Charolles et Péry-Woodley 2005). À la suite, nous nous concentrerons cependant sur l'approche de Charolles. Le travail de Virtanen (1992, 2004) est discuté en section 4.1.

Selon Charolles (1997 : 3, 2005 : 4), les adverbiaux détachés à l'initial de la phrase peuvent donc *cadrer* des segments textuels sous leur influence sémantique, c'est-à-dire spécifier le critère selon lequel les éléments textuels suivants se regroupent. Comme cette influence sémantique peut également excéder la phrase dont l'expression fait syntaxiquement partie, le cadre assure une fonction organisatrice dans le texte.

Charolles (2005 : 4) constate que, le plus typiquement, la fonction cadrative est assurée par des adverbiaux spatiaux et temporels tels que *En France*. De plus, les adverbiaux plus abstraits, tels que *en français*, sont aussi capables de fonctionner comme cadres. Charolles (*ibid.*), en revanche, poursuit en relevant que, malgré leur usage commun dans la position initiale de la phrase, les adverbiaux ayant une valeur évaluative, comme *malheureusement*, ou qualitative, tels que *par hasard*, ne peuvent pas prendre cette fonction.

La différence fondamentale entre les relations cohésives et les cadres est la direction de référence ; tandis que les premières établissent des relations avec des éléments antérieurs dans le texte, le critère spécifié par l'expression cadrative crée des liens avec les éléments suivants dans le texte (Charolles 1997 : 3). Plus spécifiquement, les cadres ouvrent un *univers de discours* que Charolles (1997 : 5) suivant R. Martin (1983 : 37) définit comme « l'ensemble des circonstances [...] dans laquelle la proposition peut être dite vraie ». Une expression cadrative peut soit fermer le cadre précédent, soit rester sous la portée sémantique de ce cadre et en devenir une subordonnée. Le choix dépend du contenu

sémantique du cadre. Le cadre *À Paris...*, par exemple, deviendrait subordonné au cadre *En France...* tandis que le cadre *En Angleterre* le fermerait et en ouvrirait un autre (Charolles 1997 : 7-14). Comme dans la lecture de tout texte, des connaissances encyclopédiques et de contexte sont donc également nécessaires dans l'interprétation des cadres.

Le fonctionnement des cadres dans l'organisation du texte est bien illustré par l'Exemple 9 et l'Exemple 10 de Charolles (1997 : 18), qui opposent deux petits segments textuels.

Exemple 9

Le dimanche, Paul va au café. Il joue au tiercé. (Charolles 1997 : 18)

Exemple 10

Paul va au café le dimanche. Il joue au tiercé. Charolles 1997 : 18)

Selon Charolles (1997 : 18-20), la différence entre les exemples réside dans le fait que, dans le premier, la position initiale de l'adverbial temporel permet d'ouvrir un univers de discours tandis que, dans le second, l'adverbial est placé à la fin de la phrase et n'a pas ce pouvoir. Charolles (*ibid.*) explique que les deux exemples permettent deux interprétations. Premièrement, il est possible de comprendre que Paul va au café le dimanche et y joue au tiercé **ce jour-là**. Deuxièmement, nous pouvons comprendre que Paul va au café le dimanche et y joue au tiercé **le dimanche** mais **aussi les autres jours**. La première interprétation prend donc la seconde phrase sous la portée sémantique de la première mais ce n'est pas cas dans la seconde interprétation. Poursuivant, Charolles (*ibid.*) souligne que les deux exemples favorisent des interprétations différentes. Dans l'Exemple 9, grâce à la portée sémantique de l'adverbial détaché, la tendance serait de favoriser la première interprétation, selon laquelle Paul jouerait au tiercé le dimanche. Dans l'Exemple 10, en revanche, les deux interprétations seraient selon Charolles (*ibid.*) autant probables l'une que l'autre. Enfin, Charolles (*id.* : 19) clarifie le fait que la préférence de la première interprétation dans l'Exemple 9 s'explique par l'univers de discours ouvert par le cadre temporel *Le dimanche*, dans lequel la phrase suivante peut être intégrée. C'est donc grâce à cette capacité de cadrage que ces expressions organisent le discours, fonctionnent comme indices de cohésion et, finalement, contribuent à la cohérence.

Les expressions cadratives les plus importantes pour notre étude sont celles pouvant introduire un item d'une série linéaire. Typiquement, ces expressions sont surtout des adverbiaux spatiaux combinés à un numéral, qui, suivant la théorie de l'encadrement, ouvrent un cadre spatial : *dans la première partie, dans la section trois*, etc. Les MIL (cf. 4.2.1) constituent un deuxième groupe d'expressions cadratives pouvant introduire un item. Selon Charolles (2005 : 4), ils sont utilisés pour leur capacité organisatrice et, qui plus est, sont souvent présents dans une série linéaire. Poursuivant, Charolles (1997 : 27) relève pourtant qu'à la différence des expressions ouvrant un univers de discours, ces expressions ne portent pas sur la vérité ou la fausseté des propositions. Par conséquent, au lieu d'initier des univers de discours, ces expressions ouvrent des *espaces de discours*. Pareillement, les connecteurs conclusifs tels que *en somme* et *bref*, également capables

d'introduire les clôtures des séries linéaires, peuvent adopter cette même fonction (cf. 4.2.4).

En définitive, suivant Charolles (1997, 2005), différents types d'expressions peuvent adopter une fonction cadrative et ainsi contribuer à la cohérence et à l'organisation textuelle. Or, cette théorie a également été critiquée. Dans son article sur les adverbiaux dans des textes argumentatifs en anglais, Crompton (2006) affirme que la capacité organisatrice des adverbiaux est seulement question de « la continuité par défaut » (cf. section 4.1) et que le critère d'organisation introduit par l'adverbial est valide jusqu'à ce qu'un nouveau critère soit proposé. Ainsi, selon ses résultats, il n'y a pas de différence dans les portées sémantiques des adverbiaux en position initiale et finale de la phrase. En concluant ses résultats sur les fonctions de la position initiale dans l'organisation du discours, Ho Dac (2007 : 283) constate également que ce n'est pas l'expression cadrative elle-même qui peut étendre sa portée sémantique au-delà de sa phrase d'accueil. En effet, cette fonction est la conséquence d'autres phénomènes discursifs, tels que le pouvoir structurant alloué aux index d'une structure énumérative, la continuité topicale et le changement de paragraphe ou de section. Poursuivant, Ho Dac (*ibid.*) souligne que « dans les autres situations discursives, i.e. en situation d'isolement, un adverbial spatial ou temporel ne porte pas réellement [de champ sémantique] ». Selon Ho Dac (2007), ce ne serait donc pas l'expression cadrative mais le contexte dans lequel elle est utilisée qui serait déterminant dans l'établissement d'un univers (ou espace) de discours.

Enfin, dans notre travail, une partie des expressions cadratives pourrait aussi fonctionner comme des cadres dans le sens restreint ci-dessus décrit par Ho Dac (2007) : utilisées dans des énumérations prototypiques avec items parallèles et facilement identifiables. Ainsi, elles peuvent supposer une fonction cadrative grâce à leur placement dans une série linéaire. Suivant Ho Dac, ce ne serait pas le cas d'autres expressions, appartenant aux séries déséquilibrées dans lesquelles le parallélisme entre les items n'est pas évident (cf. 4.2.4). Cette distinction nécessiterait pourtant encore des analyses pour être confirmée.

4. ORGANISATION TEXTUELLE

Dans la section précédente, nous avons défini le texte comme une unité cohérente. Selon notre définition de la cohérence, cette propriété a pour conséquence qu'un texte est aussi forcément organisé. S'il ne l'était pas, il ne serait non plus cohérent et ne remplirait pas la définition du texte. Ces deux propriétés vont donc ensemble.

Outre la cohérence, le besoin de l'organisation trouve une raison fondamentale dans la contradiction entre le texte et ce qu'il représente ; le texte est forcément de nature linéaire, tandis que les représentations métalinguistiques qu'il décrit ne le sont pas (cf. notre discussion sur le texte et le discours dans 3.2 et Levelt 1981). Par conséquent, comme le notent Ho Dac (2007 : 65) et Ho Dac et Péry-Woodley (2009, cf. aussi Goutsos 1996 : 504 et Adam & Revaz 1989 : 66 pour les MIL), l'auteur doit attribuer au texte une certaine structure pour bien faire passer le message. Phénomène plutôt complexe, l'organisation textuelle peut pourtant se dérouler à des niveaux différents selon l'approche adoptée. Exemple simple et classique, le modèle IMRAD de représentation de l'organisation des articles de recherche divise l'article en sections thématiques : *Introduction, Méthode, Résultats, Discussion* (cf. *inter alia* Swales 1990). Un autre exemple d'un niveau d'organisation textuelle est les séries linéaires au centre de notre étude.

Comme le constate Péry-Woodley dans l'extrait suivant, un aspect central de l'organisation textuelle est *la segmentation* :

Quel que soit le niveau d'organisation envisagé, structurer revient toujours à segmenter, c'est-à-dire à la fois diviser et rassembler.[...] C'est le cas des IU (introduceurs d'univers) [...] qui fournissent un critère pour l'interprétation des propositions dans leur portée et qui déterminent ainsi un segment.

Péry-Woodley (2000 : 136)

De même que l'organisation textuelle, la segmentation ne se limite pas à un seul niveau du texte. L'extrait de Péry-Woodley (*ibid.*) illustre un niveau de segmentation à l'aide des expressions cadratives mais la segmentation peut également se dérouler à d'autres niveaux, notamment thématiques et typographiques.

Dans cette section, nous nous concentrerons sur l'organisation textuelle à partir de la segmentation du texte ainsi que sur la façon dont les séries linéaires effectuent cette tâche. Nous commencerons par examiner l'organisation textuelle comme le résultat d'une fonction de segmentation et étudierons un modèle qui a comme objet la segmentation du texte en séquences se suivant linéairement dans le texte. Ce modèle, celui de *la séquentialité du discours* présenté par Goutsos (1996), part de la contradiction entre la linéarité du texte et la non linéarité de ce qu'il représente, se référant souvent au terme

*discours*¹⁵. Centré sur des textes expositifs, le modèle adopte une approche holistique de segmentation qui ne dépend pas d'un seul niveau mais tient compte de tous les niveaux d'organisation du texte et du discours. Dans un second temps, nous aborderons l'organisation et la segmentation textuelle plus spécifiquement dans la perspective de la présente thèse. Dans cette section, nous explorerons les travaux antérieurs sur les séries linéaires et présenterons notre propre définition de ces structures et de leurs fonctions.

4.1 SEQUENTIALITE DU DISCOURS (GOUTSOS 1996)

Le modèle de la séquentialité du discours¹⁶ proposé par Goutsos (1996) cherche à examiner l'organisation du discours *via* sa segmentation en séquences se suivant linéairement. Sur un plan global, ce modèle bénéficie des travaux de Virtanen (1992, 2004) et Enkvist (1975, 1987) sur les stratégies textuelles et l'examen du processus de production du texte dans une perspective textuelle et discursive. Le modèle de Goutsos (1996) peut également être placé dans le modèle texte-typologique présenté dans Virtanen (1992 : 41-55), appliqué à l'origine à l'étude des fonctions discursives des adverbiaux dans une position initiale de la phrase (Virtanen 1992, 2004, voir section 3.4.4).

Comme motivation pour son modèle, Goutsos (*ibid.*) note le besoin d'une théorie holistique pour examiner la segmentation du discours, une théorie qui ne se limite pas à un seul niveau d'organisation. Goutsos (1996 : 501-502, 527) critique les travaux dans ce domaine, trop centrés sur la cohérence sémantique qui lui semble insuffisante. Goutsos (*ibid.*) constate également que, bien que quelques indices linguistiques individuels de la séquentialité (comme ceux étudiés par Halliday et Hasan 1976) soient beaucoup étudiés, ce n'est pas le cas de leur usage séquentiel dans le discours, soit la façon dont ils sont utilisés ensemble.

Goutsos (*id.* : 505) souligne que son modèle prend en compte uniquement la segmentation linéaire du texte et reste neutre en regard de son contenu propositionnel. Ainsi, contrairement à la RST, par exemple, qui segmente le texte en unités reliées par des relations sémantiques telles que la *motivation*, la théorie de la séquentialité ne pose pas de critères préalables d'organisation. Les séquences peuvent être organisées par rapport à tous types de critères, notamment sémantiques, thématiques ou typographiques.

¹⁵ Voir notre discussion dans 3.2 sur le discours comme représentation mentale construite de façon hiérarchique et le texte comme une entité physique présentant le discours linéairement.

¹⁶ Comme nous l'avons discuté dans la section 3.2, l'usage des concepts de *texte* et de *discours* peut différer dans des travaux très similaires. À cet égard, Goutsos dit examiner le discours, alors que l'objet de notre travail est le texte. En discutant du travail de Goutsos, nous nous référons cependant à sa terminologie et utilisons le terme discours.

4.1.1 SCHEMA DE GOUTSOS : STRATEGIES, TECHNIQUES ET ZONES

Le modèle de Goutsos (1996 : 504) procède à la division du texte en séquences essentiellement selon deux principes. En premier lieu, suivant le principe de *la continuité par défaut de la lecture* d'un texte, le lecteur présume que les phrases qui se suivent sont liées aux phrases précédentes. Ensuite, si ce n'est pas le cas, c'est à l'auteur d'indiquer la discontinuité. Comme le remarque Goutsos (*ibid.*), « the writer is faced with the task to manage the interaction through discourse in sequential terms and to segment discourse into chunks and indicate their boundaries, i.e. the discontinuity between one and another » (« le rédacteur fait face à la tâche de gérer les interactions *via* le discours en termes séquentiels et de segmenter le discours en fragments tout en indiquant leurs limites, soit la discontinuité entre eux »). Poursuivant, il énonce (*id.* : 504-507) que le résultat, c'est-à-dire le texte segmenté, peut être perçu comme un ensemble de séquences formées à partir de ces deux principes. Plus spécifiquement, il s'agit d'une alternance de zones de continuation construites par des phrases reliées et de zones de transition délimitant les séquences formées.

Pour schématiser la segmentation du texte en séquences, Goutsos (1996 : 507) propose un modèle à trois niveaux décrivant les aspects du processus de segmentation, présenté ci-dessous en Figure 1. Parmi ces niveaux, les *stratégies* sont le composant le plus abstrait et elles fonctionnent au niveau cognitif de l'auteur. Ces stratégies sont réalisées dans le texte par des *techniques linguistiques* différentes, également signalées dans la surface du texte par des mots et des indices typographiques. Ces techniques opèrent au deuxième niveau du modèle. *Les zones de continuation et de transition*, enfin, sont le résultat textuel des stratégies et correspondent à des séquences du texte. Les zones sont donc le produit final du processus de segmentation.

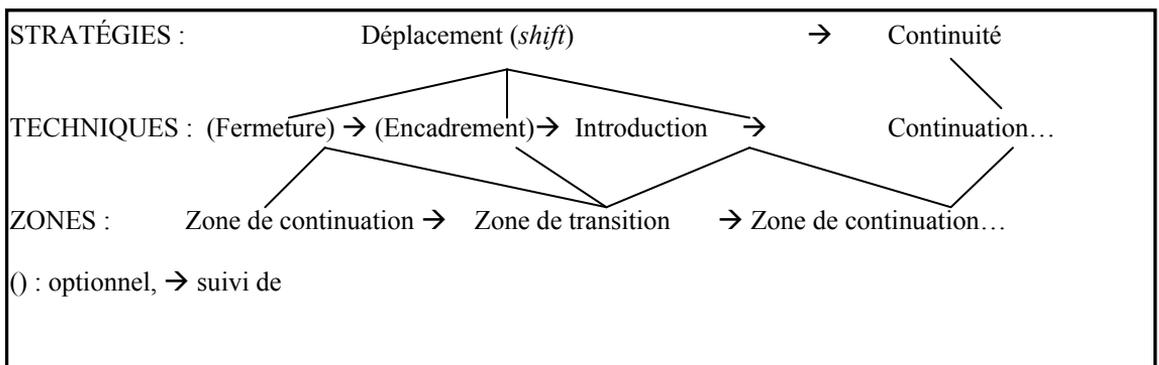


Figure 1 Modèle de Goutsos (1996 : 507).

Les stratégies textuelles sont le composant crucial de la segmentation. En effet, c'est à leur niveau, soit le niveau cognitif de l'auteur, que la création des séquences est déterminée. Les stratégies textuelles relient également le modèle de Goutsos aux travaux de Virtanen (1992, 2004) et de Enkvist (1987) sur les stratégies textuelles. Selon ces derniers, les

stratégies peuvent être définies comme « the aggregate of all decisions concerning textualization that the text producer makes in view of a communication goal » (« l'agrégat de toutes les décisions concernant la textualisation que le producteur du texte réalise dans un but de communication ») (Virtanen 1992 : 51, et discussion 41-55 ; cf. aussi Enkvist 1975 : 18-22, 1987). Il s'agit donc des décisions de l'auteur concernant la mise en texte du discours en fonction de son but communicatif. Autrement dit, dans la pratique, l'auteur prend ces décisions afin d'organiser son texte du mieux possible pour atteindre le but communicatif choisi.

De plus, les stratégies textuelles peuvent également correspondre à des stratégies plus spécifiques, également présentes à la surface du texte. Suivant Virtanen (1992 : 84-128), nous les dénommons *text strategic continuities (TSC)* (Continuités stratégiques du texte). Les TSC peuvent être définies comme « a thematic or topical uniform text-structuring orientation chosen to attain [...] a maximally profitable text organization » (« une orientation de structuration du texte uniforme, thématique ou topique, afin d'atteindre [...] une organisation textuelle profitable au maximum ») (*id.* :85) et être signalées dans le texte par des indices linguistiques. Par exemple, une TSC temporelle peut se manifester par des adverbiaux temporels placés à l'initiale de la phrase : *In January... In March... In May...* (exemple de Virtanen 1992 : 134-136). Ces structures créent de la cohésion et de la cohérence tout en segmentant le texte en fonction des critères spécifiés par les adverbiaux.

Dans le modèle de Goutsos (1996), les deux stratégies textuelles principales sont la discontinuité, nommée déplacement, et la continuité. Goutsos (*id.* : 507-508) note que, la continuité étant le défaut qui n'a pas nécessairement à être explicitement signalé, c'est la stratégie de déplacement qui joue le rôle central dans la création des séquences. Cette stratégie a pour fonction d'indiquer la discontinuité ou, plus précisément, la transition d'une zone de continuation à une autre. Les techniques linguistiques pour réaliser cette stratégie sont la fermeture, l'encadrement et l'introduction dont seule l'introduction est obligatoire. Ces techniques fonctionnent également comme signaux linguistiques des zones de continuation et de transition.

4.1.2 TECHNIQUES DE DEPLACEMENT ET DE CONTINUATION

Dans cette section, nous examinerons les différentes techniques linguistiques qui servent les stratégies de déplacement et de continuité. Ces techniques étant indiquées par des marqueurs linguistiques à la surface du texte, leur examen illustre bien la façon dont les séquences sont formées et signalées dans la pratique. Nous commencerons par les techniques servant la stratégie de déplacement pour aborder ensuite la technique de continuation.

4.1.2.1 Techniques de déplacement

Le modèle de Goutsos (1996) comprend trois techniques de déplacement : *l'introduction*, *l'encadrement* et *la fermeture* (cf. la Figure 1). *L'encadrement* signale simultanément la fermeture d'une zone de continuation et l'ouverture d'une zone de transition. Il indique donc qu'une séquence se termine et qu'une autre s'ouvre bientôt.

Selon Goutsos (*id.* : 508-512), un changement de paragraphe constitue un signal fort et fréquent de l'encadrement. Nous relevons, en outre, d'autres signaux comme les expressions métadiscursives du type «*Now I am opening a new domain* » à la fin d'un paragraphe, les connecteurs *now* et *then* et les reprises formées, par exemple, d'un démonstratif et d'un nom, telles que *ces aspects*. Pour notre travail, la *prédiction* est, enfin, un signe d'encadrement important. Elle correspond aux amorces dans les séries linéaires, illustrée par *three* dans l'Exemple 11, ci-dessous, emprunté à Goutsos (*id.* : 511). Goutsos (*ibid.*) note que la prédiction indique l'ouverture d'une zone de transition qui couvre la phrase dont la prédiction fait partie.

Exemple 11 Les en gras sont de nous.

For railway fans, privatisation should, in principle, offer **three advantages**.

First, it should encourage [...]

Second, private operators [...]. **Third**, [...].

Les amorces sont optionnelles dans les séries linéaires ; si elles sont absentes, le début de la série est introduit par le marqueur énumératif *first*. De même, les encadrements sont optionnels dans le modèle de Goutsos (1996). En leur absence, la technique d'encadrement n'est simplement pas utilisée. En ce cas, le début de la séquence est signalé par la technique *d'introduction* et par le connecteur *first* qui indique, sur le niveau textuel, l'ouverture d'une nouvelle zone de continuation. Goutsos (1996 : 511) précise également que le premier connecteur fonctionne comme indice pour les deux autres connecteurs, introduisant les autres items de la série.

Outre les marqueurs indiquant le premier item d'une série linéaire, Goutsos (1996 : 512-513) note comme indices possibles de la technique d'introduction et du début d'une nouvelle zone de continuation les *structures presentationnelles*, telles que les propositions avec un sujet indéfini (*Un chien... ou Un exemple typique...*). De plus, dans les *chaînes de référence*, l'introduction d'une nouvelle zone de continuation peut être signalée par la répétition d'un nom au lieu d'un pronom : *Saddam... He... He... Allowing Saddam...*

L'introduction diffère des autres techniques de déplacement en ce qu'elle est obligatoire. Une nouvelle zone de continuation ne peut pas être établie sans elle. La fermeture d'une zone de continuation (précédente), en revanche, est possible soit par l'encadrement, soit automatiquement, sans zone de transition. Or, Goutsos (*id.* : 512) note qu'il est très rare qu'une introduction ne soit pas précédée par l'encadrement, ce qui paraît logique étant donnée la diversité du marquage de l'encadrement dans le texte. Les changements de

paragraphe, par exemple, sont dans la plupart des textes très fréquents et signalent aussi, comme nous l'avons dit, des zones de transition.

La *fermeture* est la dernière technique pour réaliser la stratégie textuelle de déplacement. De même que l'encadrement, elle assure l'ouverture d'une zone de transition et fonctionne ainsi, comme son nom l'indique, comme un signe de fermeture. Pourtant, contrairement à l'encadrement, la fermeture est incapable de fermer seule la zone de continuation ; elle ne peut que l'anticiper. La zone de continuation, ouverte plus tôt, et la zone de transition, ouverte par la fermeture, restent par conséquent ouvertes simultanément jusqu'à ce que l'encadrement ferme la zone de continuation ou l'introduction ouvre une nouvelle zone de continuation et la zone de continuation précédente se ferme automatiquement. Comme l'encadrement, la fermeture peut aussi être signalée par le changement de paragraphe, les reprises nominales et les marqueurs discursifs, tels que *therefore* et *and*. L'Exemple 12 ci-dessous illustre le fonctionnement de la conjonction *and* comme marqueur de fermeture. Cet exemple se compose de trois phrases, dont les deux premières étiquetées 6.7 font partie d'une zone de continuation déjà ouverte et la seconde, la phrase 7.1, en ouvre une nouvelle.

Exemple 12 Les italiques sont de Goutsos.

- (6.7) We do not yet have such theory, and we may still be a long way from having one, but we do already know many of the properties that it must have.
- (6.7) *And we shall see, in later chapters, that we already know a fair amount about the predictions a quantum theory of gravity must make.*
- (7.1) Now, if you believe that the universe is arbitrary...

(Goutsos 1996 : 514-515)

Goutsos (*id.* : 514-515) note que, dans cet extrait, *and* revêt une triple fonction. La première est effectivement la fermeture, c'est-à-dire l'ouverture d'une zone de transition ; on sait que *and* marque le dernier élément d'une succession. Ensuite, l'encadrement, soit la fermeture de la zone de continuation dont les deux phrases 6.7 font partie, est également ainsi anticipé. Cette fonction d'encadrement est logique car la zone de continuation ne peut pas se poursuivre après la fin du dernier élément de la liste. Enfin, la troisième fonction de *and* décrit par Goutsos (*id.* : 514-515) est l'indication de la continuation : en marquant l'addition de la première phrase 6.7 à la seconde, elle indique également son inclusion dans la zone de continuation déjà active.

En définitive, comme le fonctionnement de *and* dans l'Exemple 12 ci-dessus le montre, un marqueur peut signaler plusieurs techniques. Aussi Goutsos (*id.* : 512) note que ces indices sont polyfonctionnels. Cette relation complexe entre les techniques et leurs indices linguistiques est également motivée par leur redondance partielle. La fermeture et l'encadrement, notamment, peuvent tous deux ouvrir une zone de transition et être signalés par un changement de paragraphe.

4.1.2.2 Technique de continuation

Contrairement à la stratégie de déplacement qui correspond à trois techniques, la stratégie de continuité est servie par une seule, la technique de continuation. Cette technique diffère des techniques de déplacement sous certains aspects importants, décrivant également son fonctionnement de manière satisfaisante.

Tout d'abord, comme nous l'indiquions, le fait de ne pas toujours être explicitement signalé est caractéristique de la technique de continuation. En lisant le texte, le lecteur présume que les phrases sont liées aux phrases précédentes et les interprète en tant que partie de la zone de continuation active, en l'absence d'indices contraires exprimant la transition. De plus, la technique de continuation est également obligatoire. Comme le rappelle Goutsos (1996 : 515), il est improbable que deux introductions se suivent dans le texte sans zone de continuité entre elles. Goutsos (*ibid.*) poursuit enfin que la continuation ne saurait fonctionner seule : pour qu'elle puisse être appliquée et ajouter un nouvel élément à la zone de continuation, celle-ci doit déjà être ouverte par l'introduction.

Nous avons déjà mentionné que la technique de continuation peut être signalée par la conjonction *and*, indiquant qu'un élément textuel, tel qu'une phrase, doit être lié à l'élément précédent. Goutsos (*id.* : 516-517) liste également d'autres types d'indices de continuation, notamment les connecteurs additifs comme *again* et *too* et les pronoms personnels tels que *il*. Ces pronoms sont surtout utilisés dans les chaînes de référence *Un chien... Il... Il...* (cf. ci-dessus), par exemple. Se référant à l'élément antérieur dans le texte, ces expressions anaphoriques attachent l'élément dont elles font partie au précédent et créent ainsi un élément homogène, plus vaste, formant une zone de continuation. Ce genre d'homogénéité à l'intérieur d'un élément peut être créée également par une continuité temporelle, établie par l'usage répétitif des verbes au même temps verbal : *Un chien dormait... Il rêvait...*

En définitive, de prime abord, les indices linguistiques de la technique de continuation diffèrent de ceux indiquant le déplacement. Bien que les connecteurs soient aussi utilisés pour indiquer le déplacement, il ne s'agit pas du même type de marqueurs. Goutsos (*id.* : 516-517) constate une différence fondamentale entre ces deux groupes de marqueurs : tous les indices de continuation fonctionnent à un niveau plutôt local du texte, entre les éléments adjacents. Ainsi, le pronom *il* se réfère à un élément précédent dans le texte. Pareillement, l'ellipse fonctionne à ce niveau local. Les indices servant le déplacement, en revanche, peuvent relier ou couvrir des segments plus vastes. C'est le cas des reprises nominales du type *ces chiens...* qui, au lieu de se référer seulement à la phrase précédente, peuvent aussi reprendre des passages plus longs. Goutsos (*ibid.*) poursuit en indiquant que ce type de cohésion, établie entre les éléments plus distants, peut être le signe d'une zone de transition. Suivant cette logique, des marqueurs différents seraient donc utilisés pour signaler la cohésion locale et globale.

4.1.3 CORRESPONDANCES ET CONFLITS ENTRE LE MODELE DE GOUTSOS ET LES TRAVAUX SUR LA COHERENCE

Grâce à sa perspective holistique de l'organisation textuelle qui ne se restreint ni à un seul niveau d'analyse, ni à un seul type de marqueurs, la théorie de la séquentialité de Goutsos (1996) fournit une méthode efficace et opérationnelle d'examen des séries linéaires. Pourtant, comparée aux études reposant sur la cohérence, examinées dans la section 3.1, et à notre perspective des séries linéaires (cf. la suite), la théorie de Goutsos comporte quelques aspects qui méritent une remarque.

Une première différence entre le modèle de Goutsos (1996) et les études sur l'organisation textuelle reposant sur la cohérence concerne le marquage explicite et le non marquage. Pour les travaux fondés sur la cohérence (cf. *inter alia* Mann et Thompson 1988 ; Charolles 1995, 1997, 2006), l'organisation textuelle ne repose pas sur les indices linguistiques. Un texte peut ainsi être effectivement cohérent et organisé sans marqueurs explicites. Les relations de cohérence dans la RST, notamment, peuvent se produire sans indices de surface. Dans le modèle de Goutsos (1996), en revanche, la segmentation du texte en séquences naît uniquement des indices linguistiques : « the identification of signal is both necessary and adequate for the establishment of sequential relations » (« l'identification du signal est à la fois nécessaire et adéquate pour l'établissement de relations séquentielles ») (Goutsos 1996 : 518).

Les rôles opposés des indices linguistiques dans les deux approches s'expliquent toutefois par les différentes façons de comprendre le marquage explicite. Comme nous le disions, Goutsos analyse tout type de marqueurs, indices typographiques, thématiques et pronoms indéfinis inclus. Les travaux sur la cohérence et la cohésion, en revanche, sont bien plus restrictifs à l'égard de la notion de marquage explicite. En outre, ils abordent souvent un seul type de marqueur, tel que les connecteurs (Schneuwly *et al.* 1989 ; Martin 1992), les expressions référentielles (Apothéloz 1995 ; Cornish 1999) ou la cohésion lexicale (Hoey 1991 ; Tanskanen 2006). Cette différence de perspective explique le fond du conflit : ce qui, dans les travaux sur la cohérence et la cohésion, peut être considéré comme non marqué ne l'est pas nécessairement pour Goutsos (1996). Or, il est clair que les relations de cohérence sont possibles sans aucun indice de surface, même selon termes de Goutsos. L'existence de ces relations dans le texte¹⁷ ne peut pourtant pas être confirmée sans une analyse de corpus. Cette question est étudiée lors de l'examen des items « non marqués ».

Un deuxième aspect nécessitant une discussion sur le modèle de Goutsos (1996) concerne le niveau d'analyse et, surtout, l'importance attribuée à l'aspect cognitif. Nous avons déjà souligné, en introduction de la section 2, que nous examinons les séries linéaires dans une perspective textuelle. Le modèle de Goutsos (1996), en revanche, adopte une perspective holistique composée des niveaux cognitif, linguistique et textuel. Pour éviter les

¹⁷ Des relations de cohérence sans aucun marqueur linguistique entre deux phrases sont faciles à créer. Leur fréquence dans la pratique, dans le texte, n'est pourtant pas évidente.

confrontations possibles, nous devons préciser que nous examinons les séries linéaires selon une perspective essentiellement textuelle. Sans ignorer le niveau cognitif du modèle de Goutsos, nous traitons avant tout les aspects linguistiques et textuels des séries linéaires. À nos yeux, les stratégies textuelles, qui correspondent au niveau cognitif de l'auteur, sont surtout un composant supplémentaire susceptible d'être pris en compte en tant que nécessaire mais sans constituer l'objet principal de notre étude.

Enfin, le modèle de Goutsos (1996) comporte également des atouts et tient compte des aspects de l'organisation textuelle absents de la plupart des travaux relatifs à la cohérence et la cohésion. Outre la perspective holistique, déjà mentionnée à plusieurs reprises, la distinction claire entre les marqueurs globaux et locaux dans le texte correspond à un tel aspect facilitant l'analyse des séries linéaires dans le texte. Comme nous l'avons constaté dans la section précédente, Goutsos (1996 : 516-517) distingue les marqueurs locaux, signalant des relations entre des phrases adjacentes, et les indices globaux, capables de reprendre des éléments plus distants. L'un de nos objectifs dans cette thèse est l'examen de la façon dont la longueur du segment à organiser influence la façon dont la série linéaire est introduite. Cette distinction nous est particulièrement utile. Dans la plupart des travaux sur la cohérence et la cohésion, cette distinction est pourtant totalement absente ou vite dépassée : Adam et Revaz (1989 : 73), par exemple, se contentent de noter l'existence de cette distinction. Ensuite, un autre atout du modèle de Goutsos central pour la présente étude concerne sa neutralité envers le contenu propositionnel des éléments textuels organisés, d'une part, et le critère d'organisation appliqué, d'autre part. Comme nous le montrons dans la section suivante, cette caractéristique s'harmonise particulièrement bien avec le fonctionnement des séries linéaires.

En définitive, les différences potentielles entre le modèle de Goutsos et les travaux sur la cohérence et la cohésion sont donc facilement explicables. Leur combinaison ne nous cause pas de problèmes. De plus, le modèle de Goutsos (1996) tient compte des aspects qui ne sont pas examinés en détail dans les autres travaux. Cependant, le modèle de Goutsos ne suffit pas à lui seul pour l'examen des séries linéaires. Le fonctionnement des types de marqueurs dans les techniques différentes, surtout, y reste flou¹⁸. Un état de fait, au moins partiellement, imputable à leur nature multifonctionnelle, comme le montre l'Exemple 12 sur *and*. Toutefois, les travaux sur la cohérence et la cohésion offrent une vision plus détaillée des types de marqueurs et même des marqueurs individuels. Pour ces raisons, nous estimons que la combinaison des deux approches est la solution la plus avantageuse : elle fournit un cadre théorique couvrant et expliquant au mieux tous les phénomènes et caractéristiques liés aux séries linéaires.

¹⁸ Les travaux de Virtanen (1992, 2004) sur les adverbiaux placés à l'initiale de la phrase et capables d'ouvrir une zone de continuation constituent une exception à cette tendance.

4.2 ORGANISATION TEXTUELLE PAR DES SERIES LINEAIRES

Dans cette section, nous nous efforcerons d'aborder l'organisation textuelle plus spécifiquement selon la perspective de la présente étude en la traitant à l'aide des séries linéaires. Nous commencerons par présenter des travaux antérieurs relatifs aux séries linéaires et aux structures similaires à celles-ci. Ces études incluent également quelques travaux traitant essentiellement des marqueurs présents dans les séries. Nous réfléchirons ensuite sur la façon dont les séries linéaires devraient être déterminées afin que la définition comprenne toutes les structures possibles et non seulement les plus prototypiques. En ce faisant, nous discuterons également du rôle des indices de surface dans la définition des séries linéaires et en avancerons quelques atouts et problèmes. Pour finir, nous examinerons les composants additionnels de ces structures, les amorces et les clôtures, qui peuvent également être décisifs dans l'identification et le fonctionnement des séries linéaires.

4.2.1 SERIES LINEAIRES ET LEURS MARQUEURS DANS DES TRAVAUX ANTERIEURS

Des travaux sur les séries linéaires, les premiers français, surtout ceux de Turco et Coltier (1988) et d'Adam et Revaz (1989) traitent essentiellement les marqueurs individuels. Plus tard, les travaux de Jackiewicz (2004, 2005), Jackiewicz et Minel (2003), du projet Annodis (Ho Dac *et al.* 2009) et d'autres se fixent comme objet d'étude les combinaisons de ces marqueurs et le mode global de participation des séries linéaires à l'organisation du discours. Par surcroît, dès les années 1970, entre autres F. Danes (1974) examine des structures très similaires aux séries linéaires à partir de la progression thématique.

L'article *Des agents doubles de l'organisation textuelle, les marqueurs d'intégration linéaire* de **Turco et Coltier (1988)** constitue un travail classique sur les marqueurs d'énumération en fondant beaucoup d'autres (dont le nôtre). L'article examine les marqueurs qui « accompagnent l'énumération sans fournir de précision autre que le fait que le segment discursif qu'ils introduisent est à intégrer de façon linéaire dans la série » (Turco et Coltier 1988 : 57) et observe les fonctions des ces marqueurs dans le texte. Ajoutant aux MIL des marqueurs comme *d'abord, ensuite, aussi, enfin, d'une part – d'autre part*, etc, les auteurs les définissent comme des termes polysémiques et polyfonctionnels qui « perdent leur vocation sémantique primitive pour désigner la succession des constituants discursifs » (Turco et Coltier 1988 : 58). Ainsi, les MIL peuvent être classés selon leur origine sémantique en marqueurs de numération (*premièrement, deuxièmement,...*)¹⁹, de lieu (*d'une part – d'autre part,...*) et de temps (*dans un premier temps*). De plus, dans les combinaisons de plusieurs marqueurs, les MIL reçoivent trois fonctions possibles : indiquer le début de la série (par exemple *premièrement, primo*), servir de relais (*secundo, deuxièmement*) et indiquer la clôture de la série (*enfin, de l'autre*). La longueur minimale d'une liste étant considérée comme deux

¹⁹ Pour les marqueurs en *-ment*, voir aussi Guimier (1996).

éléments, les marqueurs de relais peuvent également introduire le dernier élément. Dans ce cas, c'est le contexte qui permet d'identifier la fin de la série.

En cas d'exemple des MIL, Turco et Coltier (1988 : 62 et passim) analysent plus en détail le fonctionnement du couple de marqueurs *d'une part – d'autre part*. Constatant que la longueur des séries dont le premier élément est introduit par *d'une part* ou *d'un côté* n'est pas limitée à deux items, ils examinent les façons dont ces marqueurs peuvent entrer en combinaison avec d'autres MIL et des connecteurs. Entre autres, sont analysées les combinaisons du type *si d'une part – si d'autre part*, *d'une part – mais d'autre part* et *d'une part – d'autre part enfin*. Les séries *non homogènes*, qui ne consistent pas en deux éléments ou dans lesquelles les deux marqueurs du couple ne sont pas tous deux utilisés, présentent un aspect intéressant de l'analyse. Comme introducteurs possibles du troisième élément, les auteurs indiquent la répétition de *d'autre part* et l'usage d'un marqueur de relais, tel que *en troisième lieu*. Le couple *d'une part – et surtout* est fourni comme exemple d'une série acceptable dans laquelle les deux marqueurs du couple ne sont pas tous deux utilisés.

Enfin, Turco et Coltier (1988) réfléchissent également sur les fonctions de ces marqueurs dans le texte. Comme pour d'autres travaux sur ce même sujet (cf. *infra*), ils (*id.* : 63-73) classent les fonctions d'empaquetage et de guidage (balisage) *id.* comme des fonctions fondamentales des MIL : cruciaux surtout pour le récepteur, ces marqueurs le guident dans l'interprétation et l'aident dans le traitement d'informations du texte, imposées au format linéaire du texte. Pourtant, leur emploi présente également un risque : selon les auteurs, (*id.* : 73), une erreur dans leur manipulation peut attaquer la crédibilité du scripteur.

Un deuxième article sur les MIL paru dans les années 1980 est publié par Adam et Revaz (1989). Cet article traite principalement des marqueurs d'énumération et de reformulation dans le texte descriptif. Constatant qu'une énumération pure n'inclut pas d'ordre entre les éléments, il s'agit, selon les auteurs (*id.* : 66), de « développer linéairement un ensemble de propositions dont l'organisation n'est à l'origine ni causale ni chronologique ». Ce sont les organisateurs textuels, ou les MIL, qui ordonnent une énumération et assistent le lecteur dans l'interprétation du texte. Pourtant, les auteurs (*id.* : 71) notent également que tous les items d'une énumération ne sont pas nécessairement introduits par un marqueur alors que le dernier l'est le plus souvent. Cette observation est intéressante, surtout pour notre travail.

De même que Turco et Coltier (1988), Adam et Revaz (1989) indiquent aussi que les marqueurs peuvent avoir leur origine dans des principes d'organisation différents, en particulier les systèmes temporel ou spatial. Ils incluent même dans leur analyse des organisateurs spatiaux tels que *au nord*, *au sud* et temporels tels que *la veille*, *le lendemain*. De plus, les auteurs étendent la liste initiale des organisateurs possibles pour inclure également des organisateurs additifs (*aussi*, *encore*, *sans oublier*). La liste des autres MIL s'agrandit également en incluant, outre les marqueurs déjà notés par Turco &

Coltier (1988), des constructions telles que *deuxième exemple, un autre, le second* et *après*. Bien que ces extensions permettent une analyse plus globale des indices de structuration du texte, elles comportent également le risque de la compliquer du fait des fonctions variées que ces constructions peuvent adopter en syntaxe et dans le texte. Le marqueur *le premier*, par exemple, peut servir de sujet de la phrase. Comme organisateur textuel, il n'est ainsi pas nécessairement équivalent à l'adverbial *premièrement*.

Les marqueurs linguistiques d'énumération sont également au centre du travail de **Jackiewicz (2004, 2005)** et **Jackiewicz et Minel (2003)** sur les séries linéaires, c'est-à-dire des structures composées d'items au moins partiellement introduites par des MIL. Ayant comme objectif final l'application des résultats en repérage automatique des séries linéaires dans le cadre de l'extraction d'information, ces articles décrivent comment les séries sont construites et introduites dans le texte. Le travail est effectué dans le cadre de l'hypothèse d'encadrement du discours de Charolles (1997) à partir d'observations dans un corpus d'articles du *Monde diplomatique* et de *Frantext*.

L'objectif computationnel de ces travaux est clairement perceptible dans la visée modélisatrice ; les séries linéaires sont examinées *via* la création des paradigmes d'introducteurs d'items et des types de séries afin d'établir une description complète des séries linéaires et de leurs marqueurs possibles. Pour les introducteurs du deuxième item de la série, par exemple, Jackiewicz constate (2005 : 102) avoir calculé 49 possibilités différentes.

Dans notre cas, les résultats les plus importants du travail de Jackiewicz (2005) et Jackiewicz et Minel (2003) concernent le marquage des séries linéaires. Le marquage incomplet des items et les séries hétérogènes avec des types de marqueurs différents sont, dans ces articles, considérés comme des caractéristiques des séries linéaires : Jackiewicz (2005 : 109) conclut « qu'il n'était pas possible d'exhiber un seul type de marque comme indice déclencheur pour l'identification des séries », bien que, le plus souvent, les items soient introduits par des MIL. De plus, Jackiewicz (*id.* : 98) note que le marquage des items se fait plus manifeste avec l'allongement de la série : plus elle est longue, plus le marquage est explicite et régulier.

La modélisation des séries linéaires a conduit Jackiewicz (2005 : 104-107) à mener une réflexion sur leur nature cognitive . Cette réflexion illustre bien sa conception de ces structures et décrit d'une façon plus abstraite le processus de sériation. Suivant Grize (1974), Jackiewicz (*id.* : 104) indique trois propriétés de base des items d'une série. Tout d'abord, le fait d'être disjoint indique surtout une certaine discontinuité dans le discours établie par la série. Les segments sont introduits par des marques discursives qui servent à « séparer les éléments du discours les uns par rapport aux autres » et « du contexte plus large » (*id.* : 105). Le marquage est nécessaire surtout lorsque le découpage ne va pas de soi.

La deuxième propriété définie par Jackiewicz (2005 : 104-106) concerne l'équivalence entre les items d'une série : les items sont considérés comme « étant de même nature ou de même importance » (*id.* : 105). Le trait commun, soit le principe d'énumération (PE), peut être indiqué par une amorce (cf. section 4.2.4), sans que ce soit une obligation. Enfin, la dernière propriété discutée par Jackiewicz (2005 : 107) insiste sur l'ordre entre les items de la série. Contrairement à ce qu'énoncent Adam et Revaz (1989), Jackiewicz (2005 : 107) constate que « étant donné que tous les éléments d'une série sont équivalents, la seule façon de les distinguer est de les ordonner ». Poursuivant, elle ajoute (*ibid.*) que le rang exact de l'item est surtout précisé dans les séries longues, comptant quatre items ou plus.

Un autre travail notable sur le marquage des séries linéaires est celui de **Hempel et Degand (2008)** qui se base sur les études de Jackiewicz et Minel (2003) et de Jackiewicz (2005), et a comme objet de décrire l'usage pratique des séries linéaires dans le texte. Quoique l'article traite l'anglais, il nous est très central grâce à son analyse relativement détaillée de la façon dont le marquage des séries est effectué dans la pratique, et grâce à ses résultats quantitatifs que nous pouvons comparer aux nôtres. De plus, un des genres discursifs analysés dans l'article est le discours de recherche.

Le point de départ de l'analyse de Hempel et Degand (2008) est les séries prototypiques, définies à partir d'un examen des modèles théoriques et des occurrences pratiques des séries linéaires. Ainsi, selon les auteurs (*id.* : 681-682), les séries prototypiques devraient :

- être précédées d'une phrase introductrice avec un classificateur et un quantificateur, du type *He gives three main reasons for his doubts* (cf. section 4.2.4)
- être explicitement signalées par des marqueurs de sériation (notamment les correspondants anglais des MIL)
- être homogènes, c'est-à-dire se composer d'items signalés par le même type de marqueurs

Les auteurs (*id.* : 686) précisent que la distinction des types de marqueurs se base sur leur origine sémantique et concerne les marqueurs temporels (*at first, then, finally*), les marqueurs spatiaux (*on one hand, on the other*) et les marqueurs numériques (*firstly, secondly*).

Les résultats de Hempel et Degand (*id.* : 688-689, 691) attestent la longueur supérieure des séries numériques, qui souvent couvrent plusieurs phrases tandis que les deux autres groupes de structure ne fonctionnent qu'à un niveau intraphrastique. Majoritairement, toutes les séries se composent de deux items. De surcroît, l'analyse présentée souligne la nature non prototypique des séries repérées. Seules les structures signalées par des marqueurs numériques sont souvent précédées d'une phrase introductrice. De plus, seules les séries spatiales se composent majoritairement d'items explicitement signalés. Dans les séries numériques et temporelles, en revanche, ce ne sont que les derniers items qui sont le plus souvent introduits par un marqueur. Ainsi, les auteurs (*id.* : 691) concluent que l'usage

des séries linéaires ne semble pas correspondre à leurs descriptions dans des manuels didactiques ou dans des modèles théoriques.

Les travaux de Jackiewicz (2004, 2005), de Jackiewicz et Minel (2003) et de Hempel et Degand (2008), fournissent donc une description générale des séries linéaires et de leur marquage. Une autre étude sur leur marquage linguistique, celle de **Porhiel (2007)**, examine un cas particulier des séries, celui des **séries à deux temps**. Ces séries sont composées de deux séries enchaînées dont la deuxième dépend de la première et est typiquement introduite par des marqueurs thématiques : *D'abord A... Ensuite B... Enfin C... En ce qui concerne A... Quant à B... Pour C...* Dans ce type de série, le premier temps constitue une zone de référence à laquelle le deuxième temps se réfère. Ainsi, tandis que le premier temps pourrait être indépendant et présente de l'information nouvelle, le deuxième temps est non autonome et ses items sont déjà présentés dans le premier temps. Notre analyse tient également compte de ces séries (cf. 6.1).

Les études sur les séries linéaires et leurs marqueurs, que nous venons de présenter, traitent essentiellement du marquage linguistique des items. Or, le marquage des séries linéaires, ou des structures énumératives, est également possible par des moyens visuels et typographiques, tels que les puces, numéros et tirets. Une approche qui tient compte de l'ensemble des moyens linguistiques et typographiques servant de marqueurs d'énumération est présenté par le **MAT (Modèle d'architecture textuelle)** (Virbel 1989, Pascual 1991, Luc 2001, Luc et Virbel 2001 et Luc *et al.* 2002, voir aussi Bouraoui et Vigouroux 2003). Ce modèle part de l'hypothèse qu'il existe une relation d'équivalence entre des formulations comportant des propriétés visuelles particulières et des formulations entièrement discursives (De même, suivant Greenbaum (1969) et Guimier (1996 : 126), Schnedercker (2001 : 261) note que les adverbes ordinaux tels que *premièrement* « se laissent facilement remplacer [...] par des signes de ponctuation comme les tirets, et aussi Péry-Woodley (2000 : 142) constate qu'il est impossible de traiter séparément les marqueurs discursifs et visuels, qu'ils sont « fonctionnellement équivalents »²⁰. Ainsi, Péry-Woodley liste (*id.* : 141-142) comme traits possibles d'énumération 1) les indices dispositionnels, soit les espaces horizontaux et verticaux entre les items, 2) les indices ponctuationnels (souvent les deux-points), 3) les indices lexico-syntaxiques, notamment le parallélisme syntaxique entre les items et 4) les indices lexicaux ou typographiques, tels que la numérotation des items ou l'usage des MIL comme marqueurs (voir aussi le travail de Suomela-Salmi 1993). Outre les items des structures énumératives, les marquages ponctuationnel et dispositionnel concernent également l'amorce de la série : les deux-points en fin d'amorce et le retour à la ligne et le blanc vertical entre l'amorce et le premier item peuvent selon Péry-Woodley (2000 : 141) constituer des indices d'amorce significatifs. Cette perspective des moyens de « marquage » des séries linéaires permet inévitablement une étude plus englobante des formes possibles de l'ensemble des structures énumératives.

²⁰ Aussi le travail de Honeste et Froissart (2003) qui se concentre sur les listes composées d'items courts inférieurs à la phrase prend en compte les marqueurs typographiques

Les études sur les structures énumératives dans le cadre du MAT se focalisent surtout sur le parallélisme (ou non) fonctionnel et syntaxique des items. En classifiant les structures énumératives différentes, Luc (2001) définit trois catégories qui, chacune, représentent une caractéristique distinctive des énumérations parallèles et non parallèles. Grâce à son analyse détaillée et à ses critères de classification bien définis, ce travail nous est particulièrement utile dans l'examen des séries linéaires atypiques. Tout d'abord, la première catégorie définie par Luc (2001) concerne l'indépendance et l'équivalence entre les items et inclut trois types d'énumérations : les énumérations *paradigmatiques*, dont tous les items sont fonctionnellement équivalents, les énumérations *syntagmatiques*, dont tous ne le sont pas et les énumérations *hybrides*, incluant des items des deux types d'énumération.

Selon Luc (2001), l'équivalence entre les items peut être syntaxique mais aussi rhétorique, signifiant que les items sont « sémantico pragmatiquement » similaires. Ce dernier cas évoque la remarque de Jackiewicz sur « la même importance » des items. Dans l'Exemple 13 ci-dessous, emprunté à Luc (2001 : 3), les items entretiennent une relation syntaxique, le premier item fonctionnant comme sujet et le deuxième comme complément de la phrase. L'énumération est donc syntagmatique parce que les fonctions des items diffèrent. Le second item, surtout, ne pourrait pas fonctionner de façon autonome sans le premier.

Exemple 13

Le Lindy Hop est :

- une danse swing,
- dont la naissance remonte aux années 20.

La deuxième catégorie définie par Luc (2001) distingue les énumérations fonctionnant à un même niveau structural du texte de celles incluant des items à différents niveaux, correspondant en l'espèce à la longueur des items. Dans les énumérations *hétérogènes*, l'un des items peut être, par exemple, intraparagaphique tandis que les autres couvrent chacun un paragraphe et fonctionnent ainsi à un niveau différent du premier. Dans les énumérations *homogènes*, en revanche, les items fonctionnent tous au même niveau hiérarchique du texte et, de la sorte, sont approximativement de la même longueur. Enfin, la dernière catégorie de Luc (2001) distingue les séries *isolées* et les séries *liées* dont l'amorce ou la clôture incluent déjà une autre série plus courte. Ainsi, les énumérations à deux temps définies par Porhiel (2007), que nous venons de présenter, seraient selon cette classification liées.

A l'aide de ces trois catégories, Luc (2001) définit les énumérations parallèles comme *paradigmatiques*, soit composées d'items équivalents, *homogènes*, donc de même longueur, et *isolées*, la structure n'incluant pas d'autres énumérations. Les énumérations non parallèles, en revanche, possèdent au moins l'une des caractéristiques suivantes, qui sont associées au non parallélisme et rendent, chacune, les structures non parallèles : les énumérations *syntagmatiques* ou *hybrides* sont, au moins partiellement, constituées d'items

déséquilibrés, les énumérations *hétérogènes* incluent des items des longueurs différentes et les énumérations *liées* sont composées d'énumérations inférieures.

Pour la fréquence des types d'énumération présentés par Luc (2001), Péry-Woodley conclut (2000 : 140) que, selon une étude relativement modeste portant sur 50 énumérations issues de textes variés, les énumérations paradigmatiques s'élèvent à 33, les syntagmatiques à 9 et les énumérations hybrides à 8 occurrences. Ainsi, de même qu'avec les marqueurs linguistiques, l'hétérogénéité et le non parallélisme des items semble fréquent dans les séries linéaires, même en tenant compte du marquage visuel dans l'analyse.

Enfin, nous présentons dans ce chapitre un dernier groupe de travaux sur les séries linéaires, le projet multidisciplinaire **Annodis**²¹, issu de la collaboration des équipes CLLE-ERSS (Université de Toulouse-le-Mirail), Greyc (Université de Caen) et de l'IRIT (Université Paul Sabatier). Le but du projet est de construire un corpus de référence annoté pour l'étude de l'organisation du discours en français. Les textes du corpus, notamment des articles de recherche, sont expositifs. Le projet comporte deux niveaux d'analyse. Le niveau *micro* part de l'analyse des unités minimales du discours pour construire des structures plus complexes *via* les relations de discours. Le cadre théorique de cette perspective repose sur la SDRT, la théorie des représentations discursives segmentées (Asher 1993, Asher & Lascarides 2003). Le niveau *macro*, en revanche, traite le développement d'un modèle pour l'annotation de la séquentialité du discours (Goutsos 1996) et des structures discursives de haut niveau. Dans la pratique, ces structures correspondent à des séries linéaires de types différents.

Dans l'article sur l'annotation du niveau macro du projet Annodis, Ho Dac *et al.* (2009 : 900) relèvent que les structures énumératives sont choisies pour représenter l'organisation séquentielle du discours. Les auteurs (*id.* : 895) indiquent également que, selon le modèle de Goutsos (1996), le modèle d'annotation construit correspond à des zones de continuation et de transition, facilitant l'étude des techniques linguistiques pour indiquer les stratégies de continuité et de discontinuité. De plus, ils (*id.* : 897) comprennent les structures énumératives comme une stratégie textuelle de base. Utilisée pour adapter le discours au format linéaire du texte, elles l'organisent hiérarchiquement en éléments énumérés, correspondant notamment aux propriétés, arguments, sous-topiques, etc.

De même que le MAT présenté ci-dessus, Annodis prend également en compte les deux types de marqueurs visuels et lexicaux (Ho Dac & al.*id.* : 897). Toutefois, contrairement à notre conception et à celle présentée dans Jackiewicz (2005), le projet Annodis n'insiste pas sur l'ordre entre les items de la série. Ainsi, comme Adam et Revaz (1989), Ho Dac *et al.* (2009 : 900) comptent comme marqueurs d'énumération des indices temporels et spatiaux autres que les MIL. Ces expressions sont surtout des cadratives (Charolles 1997, cf. 3.4.4), telles que *En France...* ou *Depuis septembre 2004...* Elles indexent plusieurs

²¹ <http://w3.erss.univ-tlse2.fr:8080/index.jsp?perso=annodis&subURL=>, Ho Dac *et al.* (2009).

segments, s'enchaînent et, par conséquent, forment une série linéaire. Les structures énumératives analysées dans l'Annodis sont donc extrêmement variées.

En définitive, outre les travaux sur les marqueurs individuels et l'organisation du texte par des énumérations, aussi ceux sur l'organisation thématique du texte étudient les structures très similaires aux séries linéaires. Le modèle de la progression thématique présenté par **Danes (1974)**, surtout, est très proche de nos séries linéaires, bien qu'il ne tienne pas compte des marqueurs linguistiques utilisés pour signaler les items, ni de l'ordre entre ceux-ci.

Sur un plan général, le travail de Danes (1974) examine la connexité du texte à travers de la progression thématique. Danes (1974) présente trois modèles de la progression thématique dans lesquels le thème, soit l'information déjà connue, et le rhème, soit la nouvelle information, sont organisés différemment²². Le premier modèle, la *progression à thème linéaire*, est le type de progression thématique le plus simple ; dans celui-ci, le rhème de la première phrase devient le thème de la phrase suivante, dont le rhème devient le thème de la troisième phrase (*id.* : 118). Dans l'Exemple 14 emprunté à Danes (*ibid.*), le rhème *Sir Alexander Fleming* de la première phrase est le thème de la seconde phrase, *He*.

Exemple 14 Nous ajoutons les numéros 1 et 2.

(1) The first of antibiotics was discovered by **Sir Alexander Fleming** in 1928. (2) **He** was busy at the time investigating a certain species of germ [...]

Le deuxième modèle de la progression thématique présenté par Danes (*id.* : 118-119) est celui à *thème constant*. En l'espèce, le même thème apparaît dans une série de phrases. Ce modèle peut, notamment, se réaliser dans les chaînes de référence déjà présentées en discutant des expressions anaphoriques : *Le chien... Il... Il...* Les séries linéaires peuvent aussi s'organiser de la sorte : *D'abord, le chien.... Ensuite, il... Enfin, il...*

Enfin, le troisième type de la progression thématique de Danes correspond le plus typiquement aux séries linéaires. Ce modèle de *thème dérivé* (*id.* : 119-120) se compose d'un hyperthème dont les sous thèmes sont discutés dans les parties inférieures. Ce modèle est illustré ci-dessous dans la Figure 2.

²² Pour la définition du thème et du rhème, voir Danes (1974 : 112-113 et in passim). Le même modèle est également présenté dans Combettes (1983 : 90-101). De plus, le « General Particular Pattern » de Hoey (1983 : 135-167) est similaire.

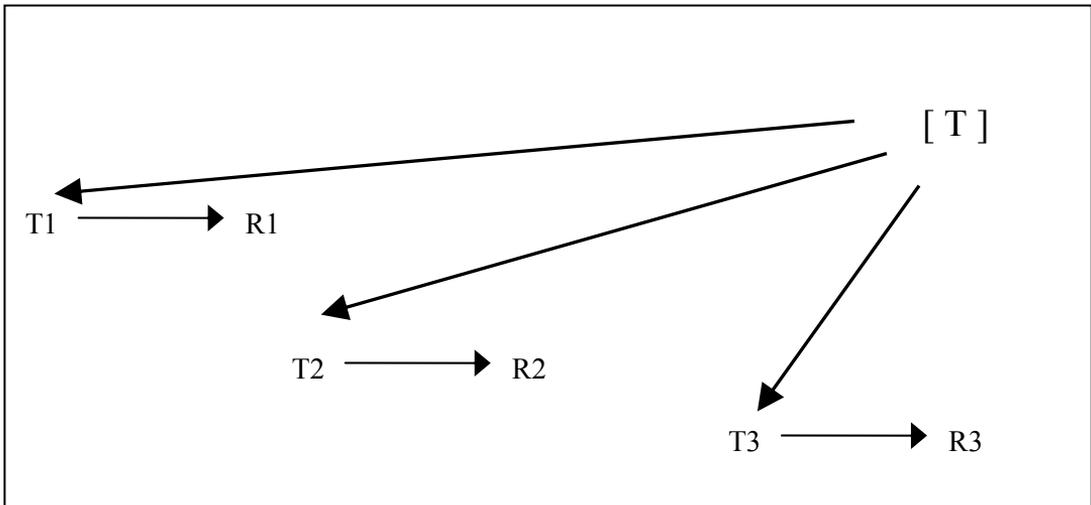


Figure 2 Modèle de thème dérivé (Danes 1974 : 119).

Dans le modèle de thème dérivé, l'hyperthème peut être explicitement annoncé, sans que ce soit une obligation. Dans l'Exemple 15 ci-dessous, extrait de notre corpus, il n'est pas exprimé par une amorce. La série discute des chercheurs et des travaux sur l'idée absolutiste aux XV^{ème} et XVI^{ème} siècles, les items les énumérant chacun dans un ordre temporel. Ainsi, chaque item présente un sous thème de l'hyperthème implicite. Selon nos critères d'annotation (cf. section 5.2.2), seul le dernier item est explicitement signalé par un marqueur d'addition, ou d'ordre, dans ce cas par le MIL *enfin*.

Exemple 15

- 1 Dans sa thèse sur *L'évolution et l'influence de l'idée absolutiste en France, de 1498 à 1559* (1955), **Jacques Poujol**, spécialiste de Jean Ferrault et de Claude de Seyssel (1450-1520), retient également le nom de Jean Montaigne, en qui il reconnaît un des interprètes de ce qu'il nomme « la monarchie style Louis XII ». Mêlant dans son commentaire les additions de Nicolas Bohier au texte original, comme tous les auteurs à sa suite, il juge que la thèse de la supériorité du roi assisté du Grand Conseil sur les parlements était utilisée pour affirmer le pouvoir absolu du roi.
- 2 En 1975, un historien italien, **Enzo Sciacca**, reprend la première partie de la problématique développée par J. Poujol. Leurs conclusions, issues des réflexions précédentes, se rejoignent sur un certain nombre de points. Pour E. Sciacca, l'attitude absolutiste se partage en deux courants ; [...]
- 3 L'historien américain **Julian H. Franklin**, fin connaisseur de l'œuvre de Jean Bodin (1530-1596), insiste au contraire sur la « *persistance [au XVI^e siècle] du constitutionnalisme médiéval* ». La monarchie des débuts de la Renaissance présente, selon lui, un double aspect : [...]
- 4 **Jacques Krynen** *enfin* a mis davantage encore l'accent sur l'intérêt du Tractatus celeberrimus pour la compréhension de l'idéologie parlementaire, à propos de la querelle du Parlement-Sénat. Nous partageons avec lui l'idée selon laquelle l'ouvrage prend place au cœur d'un des plus brûlants débats entre Moyen Âge et Renaissance. [...]

Arabeyre

4.2.2 POUR UNE DEFINITION DES SERIE LINEAIRES : QUELLES FONCTIONS ET CARACTERISTIQUES PRENDRE EN COMPTE ?

Dans la section précédente, nous avons observé des travaux antérieurs sur les séries linéaires et leurs marqueurs. Comme nous l'avons montré, les séries et leurs fonctions dans le texte et le discours peuvent être appréhendées de maintes façons. À la suite, nous réfléchissons sur ces idées et travaux présentés afin de créer une définition des séries linéaires et de leur fonction qui corresponde aux besoins de notre étude. Cette définition doit également inclure une décision sur l'exigence de l'ordre entre les items de la série.

Un aspect souvent présent dans les travaux sur les séries linéaires est leur côté cognitif : de même que la segmentation du discours en général (Goutsos 1996 : 504 ; Ho Dac 2007 : 65), les séries linéaires sont considérées comme nécessaires pour répondre au problème soulevé par le contraste entre les représentations mentales de l'émetteur, d'une part, et le format obligatoirement linéaire du texte reçu par le lecteur, d'autre part. Ho Dac *et al.* (2009 : 897), par exemple, notent que « enumerative structures correspond to a textual strategy used to adapt discourse to the linear format of text » (« les structures énumératives correspondent à une stratégie textuelle utilisée pour adapter le discours au format linéaire du texte »). De même, Turco et Coltier (1988 : 71) indiquent que les MIL permettent au lecteur de prendre en compte les contraintes issues de cette linéarité forcée du texte tout en fournissant des instructions d'interprétation des informations qui, contrairement au texte, ne sont pas de nature linéaire.

L'une des fonctions des MIL serait donc de guider le lecteur dans l'interprétation du texte, forcé au format linéaire. Ce rôle de guidage est remarqué également par Schnedecker (2001 : 279, 281) qui, allant plus loin, note également que l'usage de ces marqueurs « requiert de la part du rédacteur une capacité de planification textuelle relativement élaborée » et permet au lecteur de schématiser la structure globale du texte. Cette fonction de guidage n'est pourtant pas nécessairement différente de celle de tous les indices de cohésion (cf. 3.3) et ne définit pas la façon spécifique dont ces marqueurs se distinguent des autres indices de cohésion. Comment les MIL guident-ils le lecteur ? Quel est le type d'organisation signalé par les séries linéaires ?

Adam & Revaz (1989 : 66) proposent une réponse aux questions sur les fonctions des séries linéaires. Ils constatent que l'énumération indique une organisation linéaire de propositions qui n'est ni causale ni chronologique. L'énumération serait donc un mode d'organisation de base, dominant en l'absence d'indication d'autres modes. De plus, divers travaux soulignent l'équivalence entre les items des séries linéaires. Déjà, Adam et Revaz remarquent (1989 : 73) que les MIL, en particulier *d'abord*, *puis*, *ensuite* et *enfin*, permettent de segmenter le texte en morceaux placés à un même niveau hiérarchique du texte. Par ailleurs, selon Turco et Coltier (1988 : 63), ils signalent que les éléments « sont à traiter dans le même mouvement ». Cette propriété est encore plus accentuée dans le travail de Jackiewicz (2005 : 104-107) qui la contemple comme l'une des trois principales des séries linéaires. Elle note ainsi que les items « sont considérés comme étant de même

nature ou de même importance » (*id.* : 105). L'équivalence entre les items est aussi présente dans le constat de Ho Dac *et al.* (2009 : 897) sur la possible correspondance des items avec des entités sémantiques différentes, telles que propriétés ou arguments. C'est également un aspect très important des énumérations parallèles définies par Luc (2001)²³.

La fréquence du parallélisme ou de l'équivalence entre les items dans les définitions des séries linéaires est également notée par Péry-Woodley (2000 : 138) et Porhiel (2007 : 107-109, cf. aussi Luc 2001). Poursuivant pourtant, tous deux ajoutent que cette conception s'avère problématique et restrictive en raison de la variation dans les structures énumératives et de la fréquence des séries non parallèles, également au centre du travail de Luc (2001). Les résultats de Hempel et Degand (2008) ainsi que notre analyse pilote sur un corpus restreint (Laippala 2008) soutiennent ces remarques en montrant que le non marquage des items ainsi que l'usage des marqueurs différents dans une même série sont fréquents. Cette diversité formelle n'empêche pourtant pas l'identification et l'examen des séries linéaires. Comme l'indique Péry-Woodley (2000 : 139), les énumérations sont fonctionnellement acceptables et leur statut énumératif n'est pas questionné malgré les « déviances » de marquage et de structure.

La définition des séries linéaires à partir de l'équivalence de leurs items semble donc insuffisante pour notre étude. Nous avons, en effet, pour objectif de couvrir également les cas moins typiques dans lesquels les items de la série ne sont pas nécessairement parallèles ni introduits d'une façon homogène. Une façon alternative de définir les séries linéaires pourrait être leur examen à partir d'un modèle d'organisation du texte, tel que la RST (cf. 3.4.1). Cette approche est pourtant critiquée par Goutsos (*id.* : 527) qui, admettant que les relations entre les items pourraient, dans le cadre de la RST, être analysées comme *séquence* ou *liste*, constate que « *it is doubtful whether essentially presentational relationships like enumeration can be couched in ideational terms* » (« il est douteux que des relations essentiellement présentationnelles, comme les énumérations, puissent être formulées en termes idéationnels »).

Goutsos (*ibid.*) estime donc que les relations sémantiques (ou idéationnelles) définies dans la RST ne saisissent pas l'organisation établie par les séries linéaires, qui sont surtout de nature textuelle et segmentent le texte à partir de critères textuels et organisationnels. Une perspective partagée par Ho Dac *et al.* (2009 : 897-898) qui définissent les structures énumératives comme un mode de structuration du discours de base, qui n'est pas lié à une seule dimension d'organisation. Afin de décrire convenablement l'organisation établie par les séries linéaires, Goutsos (*id.* : 511-512) propose de les définir à partir de son modèle comme une séquence dans laquelle le premier item introduit une zone de continuation et indexe aussi les items suivants. Ce modèle reste également neutre envers l'équivalence ou le déséquilibre des items et prend en compte le marquage des items. En effet, les introductions des zones de continuation peuvent être signalées par des marques de sériation

²³ Pour des exemples des structures parallèles, voir Suomela-Salmi (1997 : 99-100).

telles que *premièrement* ou *un autre*. De plus, l'un de ses atouts est son appréhension globale du marquage de l'organisation.

Bien que Goutsos (1996) et Ho Dac *et al.* (2009) définissent les séries linéaires comme un mode d'organisation de base, segmentant seulement le texte en séquences, d'autres modes d'organisation plus précis peuvent s'appliquer à ces structures. Ho Dac *et al.* (2009 : 897-898), par exemple, énoncent que, le principe d'énumération dans une seule série étant variable, les items co-énumérés peuvent correspondre à des événements, sous topiques, arguments, etc. La possibilité de se combiner avec d'autres modes d'organisation est aussi indiquée par Jackiewicz (2005 : 107). Elle note que sur « les opérations de caractère purement organisationnel, peuvent se greffer des relations argumentatives [...] ou [...] des rapports sémantiques d'opposition, de concession, de hiérarchisation ou conséquence entre les items ». Ces rapports sémantiques peuvent aussi être indiqués en surface du texte par des connecteurs, tels que *mais* et *parce que* qui peuvent se combiner avec des MIL (cf. aussi Turco et Coltier 1988). Grâce à cette capacité de combinaison, les séries linéaires deviennent une structure multifonctionnelle susceptible d'adopter, outre sa fonction de base, des modes d'organisation variés dans le texte.

En définitive, l'approche présentée dans Goutsos (1996) et Ho Dac *et al.* (2009) s'ajuste bien à la variété des fonctionnements des séries linéaires. Elles peuvent donc être considérées comme un mode d'organisation de base, une stratégie textuelle nécessaire pour adapter le discours au format linéaire du texte. La segmentation du texte en items d'une série linéaire étant purement organisationnelle, d'autres critères d'organisation, tels qu'une dimension thématique ou sémantique, peuvent s'y ajouter. Le mode d'organisation de la série linéaire peut être explicitement dénoté par des indices fonctionnant comme guide pour le lecteur dans l'interprétation de cette organisation.

Un critère d'organisation sémantique qui peut s'ajouter aux séries linéaires est l'ordre entre les items. Dans cette étude, nous considérons cette caractéristique comme obligatoire pour ces structures, et justifions cette décision dans la suite. Nous avons déjà montré que les travaux sur l'énumération conçoivent l'ordre entre les items de diverses manières. Dans le projet Annodis, l'ordre entre les items n'est pas une propriété nécessaire d'une énumération (Ho Dac *et al.* 2009). Adam et Revaz (1989 : 66) constatent également qu'une énumération pure n'a pas d'ordre et que c'est à cette fin que les marqueurs d'énumération sont utilisés. Bilger (2003 : 29-30, 33), en revanche, énonce le contraire en disant que les éléments énumérés sous forme de liste, l'un après l'autre, sont par défaut interprétés comme ordonnés et que la linéarité du texte établit un certain ordre entre les items. Jackiewicz (2005 : 107), enfin, compte l'ordre entre les items comme une caractéristique essentielle d'une série linéaire.

Dans notre étude, nous suivons le modèle de Jackiewicz (2005) et prenons un certain ordre entre les items comme une propriété essentielle des séries linéaires. Notre décision est motivée fondamentalement par notre désir de restreindre le champ de l'étude afin d'inclure

uniquement les séries linéaires dont les items sont, au moins partiellement, ordonnés. Pour la même raison, nous excluons les séries introduites par les marqueurs visuels, tels que les tirets, et privilégions les marqueurs linguistiques. Cette exigence d'ordre est aussi légitimée par l'annotation qui doit, dans une étude quantitative et comparative, opter pour une méthode fiable et systématique. Sans aucune garantie de véracité de l'hypothèse, nous avons présumé que les critères de surface fiables sont plus faciles à établir pour les séries, au moins partiellement, ordonnées, qu'elles comprennent des items équivalents ou non. L'établissement des critères de surface fiables pour les séries non ordonnées et déséquilibrées, dont les items sont diversifiés, nous semble, surtout, difficile à établir du fait de la variation des indices de sériation possibles dans ces structures. Les expressions cadratives, par exemple, auraient sûrement posé des problèmes comme marqueurs d'item dans les séries déséquilibrées incluant également d'autres marqueurs. Si un item d'une structure est introduit par le cadratif *En France* et les autres non, devrait-elle être analysée comme une série ? Comment alors définir la nature d'une série linéaire ? Les indices linguistiques d'expression de l'ordre, en revanche, forment un groupe plus aisément discernable et restreint. Leur présence ou absence permet également de déterminer si les items de la série sont (au moins partiellement) ordonnés ou pas. Ainsi, le choix de l'ordre entre les items comme propriété essentielle dans les séries linéaires de cette étude remplit donc l'objectif de restriction du champ d'étude, d'une part, et de garantie d'établissement de critères d'annotation fiables et opérationnels, d'autre part. Par surcroît, il inclut également dans l'analyse les structures déséquilibrées.

4.2.3 TRAITS DE SURFACE COMME SIGNES DE SERIE LINEAIRE – ATOUTS ET PROBLEMES

Dans la section précédente, nous venons d'indiquer que nous établissons des critères de surface pour l'annotation des séries linéaires. Cette option peut sembler conflictuelle avec notre choix méthodologique d'une fonction et non une forme définie à l'avance (cf. 2.1) comme point de départ de l'analyse. Nous clarifierons donc, à la suite, la motivation de cette décision.

Selon Porhiel (2007 : 107), les séries linéaires sont facilement identifiables pour le lecteur. Ho Dac *et al.* (2009 : 900) indiquent aussi que ces structures sont intuitivement familières aux annotateurs et ainsi bien adaptées à cette tâche. Or, dans les sections précédentes, nous avons également discuté de la grande variation des séries linéaires en termes de structures, de fonctions et de marquages. En outre, une analyse quantitative et comparative fiable exige de définir les séries linéaires afin de pouvoir justifier systématiquement pour chaque occurrence les raisons de son inclusion ou de son exclusion de l'analyse. Afin que ces raisons ne reposent pas sur l'intuition de l'annotateur, elles devraient à notre avis se fonder sur des indices linguistiques. Une définition du fonctionnement des séries linéaires, comme nous l'avons formée dans la section précédente, ne suffit pas. C'est à cet effet que nous considérons l'établissement des critères de surface nécessaire.

Bien évidemment, les critères d'annotation, au lieu de traits de surface, pourraient également être sémantiques ; il existe certainement des séries linéaires d'un type uniquement identifiables à partir de critères de sens. Or, dans notre cas, une telle méthode ne serait pas fiable non plus. Étant donné que, à quelques exceptions près, nous sommes la seule annotatrice du corpus, une annotation fondée sur des critères sémantiques risquerait de reposer uniquement sur notre propre intuition et notre appréhension du texte. Pour éviter une telle situation, les traits de surface nous semblent, en l'occurrence, la seule option opérationnelle à même de garantir l'objectivité de l'annotation. En outre, nous ne sommes pas seule à nous fonder sur des critères de surface. Ho Dac *et al.* (2009 : 901-902) notent également que « enumerative structures cannot be defined (and interpreted) independently of the cues that signal them » (« les structures énumératives ne peuvent être définies (et interprétées) indépendamment des indices les signalant »).

Comme toutes les méthodes, l'application des critères de surface n'est pourtant pas sans désavantages. Surtout, ces critères risquent, d'une part, d'exclure de l'analyse les structures dont la fonction est similaire à celles qui sont incluses et, d'autre part, d'y inclure des structures dont la fonction primaire n'est pas celle des séries linéaires. Nous avons cependant préféré cette méthode pour les raisons déjà exposées. Quoique le corpus résultant risque d'être fonctionnellement hétérogène, il rassemble du moins systématiquement toutes les structures du corpus remplissant les critères établis et permet ainsi leur étude quantitative et systématique.

Enfin, malgré la nécessaire formulation de quelques critères formels pour les séries linéaires, nous souhaitons garder autant que possible la fonction comme point de départ de notre analyse. C'est envisageable essentiellement du fait que, bien que nous précisions un **type** d'expression devant être inclus par une série linéaire, nous n'avons pas de liste pré-établie d'expressions possibles. Les seuls critères de surface indiquent que l'expression doit introduire un item de la série et exprimer l'ordre dans un sens très vaste, progression, addition, etc. inclus. Ainsi, bien que le type d'expression soit décidé à l'avance, l'analyse tient compte de toutes les expressions qui remplissent cette fonction dans le contexte. De plus, l'application des critères de surface ne signifie pas que nous nous concentrons uniquement sur les séries linéaires identifiables par la cohésion ou les structures bien formulées. D'abord, les items non marqués ne sont pas nécessairement introduits par des items linguistiques et peuvent ainsi faire partie de la cohérence. L'identification de ces items et le rôle de la cohérence et de la cohésion dans ce processus seront examinés lors de l'analyse. En ce qui concerne les structures mal formulées et possiblement incohérentes, notre analyse englobe toutes les séries linéaires que nous avons pu identifier et ainsi inclure dans l'analyse. Bien évidemment, les structures les plus incohérentes que nous n'avons pas pu bien interpréter sont ainsi exclues, ce qui nous paraît justifié notre but n'étant pas d'examiner l'organisation textuelle de la perspective de bonne ou de mauvaise formation mais les séries linéaires qui doivent satisfaire certains critères de base même quand moins bien formulées.

4.2.4 AUTRES COMPOSANTS D'UNE SERIE LINEAIRE : AMORCE ET CLOTURE, EQUIVALENCE ENTRE LES ITEMS

Une série linéaire est obligatoirement composée d'au moins deux items co-énumérés qui se suivent dans le texte. Pourtant, les items proprement dits peuvent également être précédés d'une amorce et suivis d'une clôture. Ces composants contribuent au marquage d'une série linéaire, surtout lorsque les items ne sont pas explicitement introduits. Le parallélisme, ou l'équivalence, souvent associé aux items des séries linéaires, peut avoir ce même effet. Dans cette section, nous définissons ces trois concepts.

Selon les termes de Goutsos (1996), les amorces sont des encadrements qui ferment la zone de continuation précédente et ouvrent une zone de transition, anticipant la nouvelle zone de continuation. Elles sont pourtant optionnelles. En effet, en leur absence, l'introduction et l'ouverture de la zone se fait par le premier item de la série. Dans d'autres travaux, outre l'annonce du début de la série linéaire, les amorces sont également censées exprimer le principe fédérateur des items, c'est-à-dire *le principe d'énumération* (dorénavant PE, Ho Dac *et al.* 2009 : 896-898 ; voir aussi Jackiewicz 2005 : 106 et *Méthod of Development* dans Fries 1983). Ce principe est en général un hypéronyme, un concept plus général qui rattache les items et qui peut être répété dans les items, sans être une obligation. De plus, Ho Dac *et al.* (2009 : 896-898) indiquent que ce principe peut également changer au fil de la structure.

Selon Jackiewicz (2005 : 98, 106 ; voir aussi Hempel & Degand 2008 et Schnedecker 2001 : 273-275), le principe d'énumération, le *classificateur*, peut être accompagné par un *quantificateur*, un numéro cardinal ou un adjectif indéfini. Dans l'Exemple 16 ci-dessous, que nous lui empruntons (*ibid.*), c'est le SN *ensembles de choix* qui fonctionne comme classificateur et le numéro *six* qui sert de quantificateur.

Exemple 16

[...] L'Europe va devoir faire **six ensembles de choix des plus importants.**

Premièrement, ce sera, au mois de décembre à Copenhague, l'élargissement à vingt-cinq pays.

Deuxièmement, dans les six premiers mois de 2003, la présentation par la Convention, dirigée par Valéry Giscard d'Estaing, d'une nouvelle constitution.

Troisièmement, (...). **Quatrièmement**, (...). **Cinquièmement**, (...). **Enfin**, (...).

Le Monde, 30 juillet 2002, p.12

Jackiewicz (2005) et Hempel et Degand (2008) considèrent les amorces comme un composant complémentaire des séries linéaires. Une autre approche est proposée par Tadros (1994). Elle examine les prédictions selon une perspective interactive du discours, les considérant comme des actes de discours par lequel l'auteur s'engage à structurer son texte d'une façon spécifique. Selon Tadros (1994 : 70), ne pas réaliser les attentes du lecteur et manquer à cet engagement mettraient la crédibilité de l'auteur en question. Dans ce modèle, la prédiction est composée d'une paire, l'élément prédictif et l'élément prédit. L'élément prédit peut également se composer d'un seul item, car le modèle s'applique

pareillement aux structures autres que les séries linéaires. Tadros (1994 : 70) définit en tout six catégories de prédiction : *enumeration*, *advance labelling*, *question*, *reporting*, *recapitulation* et *hypotheticality*. Pour les amorces des séries linéaires, nous appliquons les trois premières catégories qui sont discutées plus tard en section 5.2.2.

En définitive, si l'amorce est, dans le signalement d'une série linéaire, définie comme un indice complémentaire, la clôture l'est encore plus. Étant donné qu'elle n'apparaît qu'à la fin, Sinclair (2004 : 88) constate que l'amorce l'emporte sur la clôture d'une façon naturelle. De même, Porhiel (2007 : 111) note qu'elle est moins développée et que ses éléments saillants sont difficiles à déterminer. Malgré ces difficultés constatées, un accord dans la définition de la clôture semble être atteint, au moins dans les travaux de Jackiewicz (2005) et de Ho Dac *et al.* (2009). Ils définissent ainsi la clôture comme un composant qui clôt la série et porte sur l'ensemble des items. Les deux travaux notent comme formes potentielles de ces expressions les références, surtout les anaphores résomptives, et les marqueurs de reformulation : *l'ensemble...*, *les trois segments...*, *ces deux...*, *en bref...*, *autrement dit...*, etc. Dans le modèle de Goutsos (1996) aussi, ces mêmes marqueurs sont analysés comme des signes de fermeture, une technique optionnelle anticipant la fermeture de la zone de continuation ouverte et correspondant aux clôtures des séries linéaires.

L'amorce et la clôture sont des composants linguistiques et textuels qui peuvent s'attacher à une série linéaire. Le parallélisme, ou l'équivalence, entre les items, diffère de ceux-ci en n'étant qu'un trait associé aux items de ces structures. Or, de même que les amorces et les clôtures, le parallélisme est, à nos yeux, un trait optionnel des séries. Il peut contribuer à la formation de la série et à son identification dans le texte, mais également être absent sans nécessairement nuire à la cohérence du segment. Cependant, il s'agit d'une notion importante dans l'examen des séries linéaires qui nous est utile surtout dans l'étude du marquage et du non marquage. Pour terminer, nous rappelons encore la façon dont l'équivalence ou le parallélisme des items est définie dans les travaux antérieurs et réfléchissons sur ce que nous entendons à cet égard.

Dans les premiers travaux sur les MIL, l'équivalence entre les items est donc plutôt implicite. Comme nous l'avons exposé, Adam et Revaz (1989 : 73) remarquent que les MIL segmentent le texte en morceaux placés à un même niveau hiérarchique du texte. En outre, selon Turco et Coltier (1988 : 63), ces marqueurs signalent que les éléments « sont à traiter dans le même mouvement ». Pour Jackiewicz (2005 : 104-106), en revanche, l'équivalence entre les items est plus centrale, signifiant que les items sont « de même nature ou de même importance ». « L'homogénéité » des items est fondamentale également dans le travail d'Honeste et Froissart (2003). De plus, comme nous venons de l'exprimer, l'équivalence ou le parallélisme des items peut être examiné également selon *le principe d'énumération* des items annoncé dans l'amorce et possiblement répété dans les items. Ces travaux ne détaillent pourtant pas comment l'équivalence peut être créée ou rompue.

Les études de Hempel et Degand (2008) et de Luc (2001), en revanche, proposent des analyses plus détaillées du déséquilibre entre les items. Dans le travail de Hempel et Degand (2008), l'équivalence est associée avec les séries prototypiques, qui doivent se composer d'items explicitement signalés par des marqueurs d'origine sémantique similaire et être précédées d'une phrase introductrice, autrement dit d'une amorce²⁴. Luc (2001), pour autant, distingue trois catégories de caractéristiques associées aux énumérations parallèles et non parallèles : l'équivalence fonctionnelle des items, leur longueur similaire, ou non, et la présence, ou non, d'une énumération inférieure dans l'amorce ou la clôture de l'énumération. Dans son analyse, la première catégorie, c'est-à-dire l'équivalence fonctionnelle des items ressemble à la notion du *principe d'énumération* dont nous venons de discuter.

Selon le travail de Hempel et Degand (2008) et la typologie de Luc, nous pouvons former une définition claire des séries linéaires parallèles et non parallèles. À notre sens, les séries parallèles sont donc surtout composées d'items de longueurs invariables et, selon un critère syntaxique ou « rhétorique », similaires. L'interprétation de cette similarité se fonde, comme la lecture, sur la cohérence et l'interprétation du lecteur, des aspects centraux dans celui-ci étant le principe d'énumération de la série ainsi que l'introduction explicite de tous les items par des marqueurs similaires. Si les longueurs des items varient d'une façon importante ou si les items et leur marqueurs sont par ailleurs très différents, nous considérons la série comme non parallèle. Pourtant, contrairement à Luc (2001) et à Hempel et Degand (2008), nous ne considérons pas que l'absence d'une phrase introductrice ou la présence d'une série inférieure dans l'amorce ou dans la clôture d'une série supérieure puisse rendre toute la structure non parallèle. Bien que ces séries soient différentes des autres, leurs items peuvent tout de même être explicites et de même longueur, aussi bien que ceux des séries isolées ou des séries précédées d'une amorce. De ce fait, il nous semble exagéré de les considérer non parallèles comme un ensemble.

4.3 BILAN : ORGANISATION TEXTUELLE

Dans les deux dernières sections, nous avons abordé de nombreux aspects liés aux principes d'organisation du texte, aux séries linéaires et aux façons dont ces structures ont été définies. De plus, nous avons également réfléchi sur la façon dont les séries linéaires devraient être définies dans cette étude. Avant de passer à la partie analytique de ce travail, nous proposerons encore une conclusion sur quelques notions principales discutées et reformulons notre pensée. Ensuite, nous réfléchirons sur les raisons sous-jacentes de notre décision de combinaison du modèle d'organisation du discours de Goutsos (1996) et des études traditionnelles sur la cohérence et la cohésion dans notre approche.

²⁴ Voir aussi notre analyse pilote du marquage des séries linéaires (Laippala 2008) qui a démontré la fréquence des items non marqués et de l'usage des marqueurs différents dans une seule série.

4.3.1 NOTRE CONCEPTION DE CERTAINES NOTIONS CENTRALES DANS L'ORGANISATION TEXTUELLE

Commençons par les notions sous-jacentes à la formation du texte : la cohérence, la cohésion et les termes texte et discours. Bien que souvent invisibles dans l'analyse pratique, ces concepts sont centraux dans l'organisation textuelle et, par conséquent, dans les fonctions et fonctionnements des séries linéaires.

La **cohérence** est donc le principe selon lequel les phrases se suivant linéairement sont interprétées comme reliées et forment un tout sémantique. Nous avons abordé ce sujet selon deux perspectives : la perspective cognitive et la perspective textuelle. Pour la première, soutenue entre autres par Charolles (1983, 1995, 1997, 2006), la cohérence est surtout un processus cognitif qui guide le lecteur dans l'interprétation des phrases. Autrement dit, il ne s'agirait pas d'une propriété du texte mais d'un principe cognitif d'interprétation. De nature pragmatique et sémantique, ce processus est, toujours selon Charolles, soutenu par les connaissances textuelles, linguistiques et extralinguistiques du lecteur. Il est ainsi, d'une certaine façon, subjectif.

La perspective textuelle de la cohérence, adoptée entre autres par Halliday et Hasan (1976), Halliday (1985) et Mann et Thompson (1988), souligne la nature sémantique du texte et le rôle des relations de cohérence, c'est-à-dire des liens sémantiques entre les phrases et segments textuels. Un texte cohérent est compris comme un tout organisé dans lequel chaque élément revêt un rôle discernable.

Dans notre travail, nous appréhendons la cohérence selon ces deux perspectives. Ainsi, nous la considérons d'une part comme un principe cognitif de l'interprétation du texte et, d'autre part, comme une propriété essentielle du texte, un résultat des liens sémantiques reliant les phrases et les segments textuels. Or, nous soulignons que l'interprétation du lecteur doit toujours reposer sur les indices linguistiques du texte même. Quoique la cohérence soit un principe cognitif influencé par les capacités personnelles du lecteur et le contexte d'interprétation, elle doit également être ancrée dans le texte : le texte et sa cohérence sont deux aspects d'un même objet. Le lecteur ne peut donc pas faire ce qu'il veut avec le texte.

La perspective textuelle de la cohérence implique également la définition du **texte**. Nous le considérons comme une unité sémantique, construite par le lecteur selon les indices linguistiques du texte. Or, le texte présente également d'autres aspects qui méritent une remarque, du fait de leur rôle important pour notre travail. Par conséquent, suivant Halliday (1985) et Halliday et Hasan (1976), le texte est également pour nous une unité fonctionnelle avec une tâche à remplir : il a toujours une fonction.

Le texte est souvent confondu avec la notion de **discours**. Comme nous l'avons expliqué, les deux termes peuvent être utilisés dans des contextes très similaires. Certains chercheurs ont même proposé de ne plus les dissocier. D'autres, encore, continuent de les distinguer.

Dans ce travail, nous considérons la distinction entre le texte et le discours comme similaire à celle entre le processus et le produit. Ainsi, le discours est pour nous un processus multidimensionnel et complexe, englobant également des dimensions extralinguistiques. Le texte, en revanche, est la trace textuelle de ce processus, son produit, qui garde cependant la même complexité. Le texte ne peut donc plus être analysé sans considérer *inter alia* son contexte de production et son but communicatif.

L'objet de notre étude est le texte. Bien que les différences entre les deux termes ne soient pas toujours très importantes, nous considérons que cette notion décrit le mieux ce que nous étudions : la trace écrite du discours, dont il est le produit. Quoique les références au processus d'écriture soient parfois nécessaires dans notre analyse, nous pensons qu'en tant qu'objet de recherche le texte forme une entité concrète et (en principe) facilement accessible. À nos yeux, en revanche, le discours demeure une reproduction qui est, comme l'a indiqué Cornish (1999 : 35), d'une certaine façon probabiliste. La reconstruction exacte du discours, d'un processus englobant un nombre de dimensions, nous semble ardue en l'absence d'accès direct à la situation de production.

Un texte peut former une entité cohérente sans indices linguistiques reliant les segments les uns aux autres. Toutefois, la formation du texte peut également être examinée à partir de la **cohésion**. Suivant Péry-Woodley (2000 : 15), nous la considérons comme des « ressources linguistiques pour créer du texte » qui guident le lecteur dans l'interprétation des phrases. La cohésion ne garantit pourtant pas la cohérence : si les indices sont mal utilisés ou mal interprétés, les phrases ne peuvent pas être interprétées comme un tout et le texte être considéré comme cohérent.

Les indices de cohésion essentiels pour notre travail sont ceux susceptibles d'introduire les items des séries linéaires : **connecteurs** et **expressions anaphoriques**. Les connecteurs ont fait objet de divers travaux selon diverses perspectives. Nous les considérons comme des signes des relations sémantiques entre les phrases qui, malgré leur polysémie attestée, guident les lecteurs en démontrant la façon dont les phrases sont combinées. Suivant Schneuwly *et al.* (1989)²⁵, nous proposons les critères suivants :

- Ils doivent joindre des structures propositionnelles en précisant leur insertion dans le cotexte ou dans le contexte de production.
- Ils doivent être placés hors de la syntaxe de la phrase.
- Ils ne peuvent être soumis aux phénomènes d'accord.

Souvent, les connecteurs sont des adverbes ou des locutions adverbiales. Les **expressions cadratives**, étudiées surtout par Charolles (1997), forment un groupe similaire²⁶.

²⁵ Comme nous l'avons indiqué en section 3.4.1, Schneuwly *et al.* (1989) utilisent l'expression *organisateur textuel*. Par simplicité, nous nous référons uniquement aux connecteurs.

²⁶ Pour cette raison, lors de l'analyse, nous examinons ces deux groupes de marqueurs ensemble dans quelques cas.

Contrairement aux organisateurs textuels, ces marqueurs peuvent pourtant être soumis aux phénomènes d'accord (*dans le premier exemple vs dans la première section*).

Comme exposé en section 3.4.2, **l'anaphore** et **la référence** forment un cas complexe. La définition typique d'une relation asymétrique entre deux objets, dont le second renvoie au premier pour son interprétation, ne recouvre pas les anaphores plus abstraites, ainsi les anaphores associatives et résomptives. De plus, la relation coréférentielle entre l'antécédent et l'anaphorique ainsi que la dépendance sémantique de l'anaphorique envers sa source ont suscité nombre de discussions sur les propriétés typiques ou nécessaires des anaphores.

Dans les séries linéaires, les types d'anaphores possibles sont plutôt restreints. Pour les marqueurs d'items, les expressions anaphoriques possibles sont les paires *le premier – le second* et *l'un – l'autre* et les dérivatifs (cf. Schnedeker 2000, 2006). Ces expressions sont effectivement sémantiquement dépendantes de leur antécédent et coréférentielles avec lui. Par conséquent, leur statut d'anaphorique n'est pas questionné. De plus, d'autres expressions référentielles des séries linéaires peuvent apparaître dans les clôtures. Il s'agit souvent de reprises synthétisantes ou reformulatives des items de la série, ainsi le SN pluriel démonstratif : *Le premier exemple... Ensuite, l'exemple... Le dernier exemple... Ces exemples...* La clôture peut également être un SN pluriel défini voire un SN défini singulier. Ces expressions ne sont pourtant pas des anaphoriques typiques. En effet, elles ne sont pas coréférentielles avec leurs antécédents et, de plus, leur dépendance sémantique peut être questionnée (voir notre discussion dans la section 3.4.2 sur les SN définis et démonstratifs). Malgré ces propriétés atypiques, ces constructions sont souvent analysées comme des anaphores résomptives, et Auricchio et al (1995 : 28), par exemple, constatent que ces structures ne peuvent pas être interprétées sans récurrence à leur antécédent. Pour cela, nous analysons ce type de clôtures comme des anaphores résomptives.

4.3.2 NOTRE CONCEPTION DES SERIES LINEAIRES

Dans les sections précédentes, nous avons discuté des séries linéaires et des façons dont elles ont été définies dans des travaux antérieurs. Nous avons également réfléchi sur la façon dont elles devraient être définies dans notre étude.

Afin d'accéder à une définition des séries linéaires qui recouvre toutes les structures possibles et soit, simultanément, aisément applicable dans la pratique, nous avons décidé de les définir à partir des indices de surface. Autrement dit, pour être analysées, les séries doivent inclure un certain nombre d'indices de surface signalant addition, la progression ou l'ordre. Bien que cette méthode nous amène certainement à inclure dans le corpus des structures douteuses tout en excluant d'autres structures correspondant davantage à des séries linéaires typiques, nous la considérons comme la seule option fiable et opérationnelle dans notre cas. En effet, le corpus est de taille relativement importante et l'annotation est, pour la plupart, le fruit de notre travail. Les traits de surface garantissent

également une analyse systématique des structures étudiées, un élément primordial dans une étude quantitative. Afin de recentrer l'étude sur les fonctions des séries linéaires dans l'organisation du texte et non sur les marqueurs et indices de cohésion individuels, les critères posés ne spécifient pourtant pas des traits de surface particuliers, tels que marqueurs ou types de marqueurs, mais indiquent seulement une fonction nécessaire pour une partie de composants des séries analysées. Comme nous le détaillons dans la section suivante, cette fonction est soit un *ordre relatif*, c'est-à-dire une *addition*, signalé entre autres par *de plus*, *et également*, soit un *ordre absolu*, signalé par exemple par *d'abord*, *le second*, et *le troisième exemple* (cf. Jackiewicz 2005 : 107). De plus, il est à noter que seule une partie des items d'une série doivent être introduits par ces marques de surface ; d'autres peuvent être non marqués, permettant ainsi l'examen de l'alternance du marquage explicite et du non marquage.

Les traits de surface fournissent donc un cadre exact et opérationnel pour l'annotation et l'analyse des séries linéaires. Cependant, la définition de la variété de leurs fonctionnements dans l'organisation du texte reste difficile. Comme mentionné, une définition à partir des liens sémantiques ou du parallélisme entre les items serait très restrictive. Ce défaut est également noté par Goutsos (1996 : 511-512, 527), qui propose de les définir selon son modèle de séquentialité comme une zone de continuation dans laquelle le premier item de la série indexe également les items suivants. Poursuivant, un autre atout de son modèle est qu'il, de même que les énumérations, opère sur le plan organisationnel du texte et ne dépend pas de dimensions sémantiques ou thématiques. De plus, il adopte une perspective très large du marquage de l'organisation, incluant tout type d'indices linguistiques, textuels et typographiques. Pour ces raisons, son modèle s'ajuste bien à la réalité de nos séries linéaires sur lesquelles des critères d'organisation sémantiques et thématiques peuvent se greffer tout en restant à même de fonctionner comme une structure de base, segmentant simplement le texte en items. Grâce à cette adaptabilité, le modèle de Goutsos est également appliqué dans cette étude.

L'approche de Goutsos (1996) attache pourtant très peu d'intérêt aux façons dont les zones de continuation et de transition sont réalisées en surface textuelle. D'une part, cela s'explique par lapolysémie des indices possibles. Comme le montre l'Exemple 12 des trois fonctions de la conjonction *and*, l'analyse détaillée des marqueurs risquerait d'être très complexe et de s'avérer peu utile pour des fonctions très nombreuses et dépendantes du contexte. D'autre part, comparé aux travaux sur la cohérence et cohésion, le modèle de Goutsos est relativement peu examiné. Peut-être un examen des indices textuels des zones de continuation et de transition serait-il justifié mais, pour quelque raison, il n'a pas été réalisé. Le travail de Virtanen (1992, 2004) sur les adverbiaux dans la position initiale de la phrase constitue, toutefois, une exception à cette tendance.

Bien que l'imprécision dans la définition des indices de surface du modèle de Goutsos soit explicable, elle complique son application dans notre étude. En effet, les diverses façons dont les séries linéaires sont introduites dans le texte et les raisons textuelles expliquant

l'usage des marqueurs différents sont centrales à nos yeux. De ce fait, nous fondons également notre analyse sur les études relatives à la cohérence et la cohésion. Comme exposé entre autres dans la section 4.2.1, ces études incluent des travaux très détaillés sur les types d'indices de cohésion différents ainsi que sur des marqueurs individuels. Elles fournissent donc un outil exact à l'analyse des types de marqueurs utilisés dans les séries linéaires. En outre, comme nous exigeons que les items des séries linéaires soient au moins partiellement ordonnés, soit également organisés selon un critère sémantique, ces travaux adoptant une perspective sémantique de l'organisation correspondent bien au cadre de notre étude. Par conséquent, les deux approches se complètent.

L'imprécision du modèle de Goutsos (1996) en regard de la relation entre les séries linéaires et les zones de continuation constitue un autre inconvénient. Une zone de continuation correspond-elle à une série linéaire ou à un item ? Comme nous l'avons mentionné, le marqueur énumératif *first* ouvre, selon Goutsos (1996 : 511-512), une zone de continuation. Par conséquent, nous pourrions penser que les autres marqueurs du même type fonctionnent d'une manière similaire et ouvrent eux aussi leurs propres zones. Dans un autre exemple, Goutsos (*id.* : 515-516) note pourtant que le connecteur additif *too* signale la technique de continuation. Ainsi, les types de marqueurs différents fonctionneraient de façons différentes. Ou l'ouverture d'une nouvelle zone dépend-elle toujours du contexte et non du marquage ?

Sans une analyse compréhensive de ces structures, aucune réponse ne peut être offerte à la question de correspondance réelle des zones de continuation et des séries linéaires ou des items. Or, une définition s'impose. Nous avons donc décidé de les analyser comme des structures récursives dans lesquelles chaque item appartient à une même zone de continuation tout en formant simultanément sa propre zone de continuation. Suivant Goutsos (1996), nous considérons également que ces structures sont de nature organisationnelle. Pourtant, contrairement aux énumérations composées d'items non ordonnés, qui peuvent opérer sur un plan purement organisationnel, les séries linéaires de notre étude, comportant des items au moins partiellement ordonnés, doivent également inclure un aspect sémantique, celui de l'ordre. En section 4.2.2, nous avons justifié cette décision par la restriction de l'objet d'étude et par les aspects pratiques liés au marquage de l'ordre, qui facilitent l'annotation et la fiabilisent. Par surcroît, cet aspect sémantique motive aussi le fait que nous nous reposons sur les travaux relatifs à la cohérence et la cohésion, également fondés sur une conception sémantique de l'organisation du texte.

5. CORPUS ET METHODE

5.1 CORPUS

De même que le texte n'est pas une suite quelconque de phrases, un corpus n'est pas un ensemble quelconque de textes. Au contraire, le corpus et son contenu englobent de nombreux aspects cruciaux pour la réussite l'étude. Dans les deux sections qui suivent, nous discutons de la collecte du corpus et de la façon dont nous l'avons effectuée.

5.1.1 CORPUS ET CORPUS. QUE SONT-ILS ? COMMENT SONT-ILS ?

Sinclair (1996) définit le corpus de la façon suivante (cf. aussi Habert *et al.* 1997 : 11) :

A corpus is 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 the language.

(Un corpus est une collection d'éléments de langage sélectionnés et ordonnés selon des critères linguistiques explicites afin d'être utilisés comme échantillons de langage.)

Sinclair (1996)

Cette définition, que nous faisons nôtre, inclut deux aspects importants d'un corpus : d'abord, l'application de critères explicites de sélection des données et, ensuite, l'emploi du résultat comme un échantillon de la langue. Condamines (2003 : 31-32) ajoute, en outre, à ces caractéristiques le fait que le corpus doit être composé de textes entiers et non d'extraits (cf. ci-dessous) et que des critères extralinguistiques peuvent aussi être utilisés pour la sélection du matériel.

L'emploi du corpus comme un échantillon de la langue qu'il représente a une conséquence importante : il doit être représentatif de cette langue. C'est dans cette perspective que les critères linguistiques et extralinguistiques utilisés pour la collecte sont cruciaux et doivent être bien définis. Autrement, il n'est pas possible de garantir la représentativité du corpus résultant et donc sa capacité à répondre aux questions de recherche posées. Comme l'illustre l'exemple de Biber *et al.* (1998 : 246), « a corpus composed primarily of news reportage would not allow a general investigation of variation of English » (« Un corpus composé principalement de reportages d'actualités ne saurait fonder une recherche générale sur les variations de l'anglais. »). De plus, la représentativité affecte la possibilité de généraliser les résultats de l'étude. *Inter alia*, Condamines (2005 : 19-21) et Biber *et al.* (1998 : 246-250) notent que, d'une part, si le corpus est mal construit, les résultats ne peuvent être généralisés au-delà du corpus et, d'autre part, même pour un corpus bien construit, les réponses fournies sont valables uniquement pour le genre discursif du corpus et ne sauraient être généralisées au-delà.

Outre la représentativité, un corpus se voit souvent attribué certaines caractéristiques, quatre selon Sinclair (1996). La première, *la quantité*, est évidente. Un corpus destiné à une étude quantitative doit intégrer un nombre important de mots et d'occurrences examinés. Sinclair (1996) l'exprime très simplement : « *The default value of Quantity is large* » (« La valeur de Quantité par défaut est grande. »). Objet de discussions fréquentes (entre autres Habert *et al.* 1997 : 145, Biber *et al.* 1998 : 248-249), la question du nombre suffisant de mots d'un corpus est pourtant dépourvue de réponse simple : tout dépend, notamment, de l'objet de l'étude et de sa fréquence dans la langue, ce qui peut varier d'une façon très importante (cf. aussi Leech 1992 ; Barkema 1994 : 271 ; Péry-Woodley 2005 : 186-188).

Une deuxième propriété *par défaut* des corpus, selon Sinclair (1996), est *la qualité* qui doit, tout d'abord, être authentique. Cette caractéristique peut également inclure les autres critères appliqués à la collecte du corpus. De plus, Sinclair (*ibid.*) y associe les corpus spécialisés ne visant pas à décrire la langue ordinaire, tels que les corpus littéraires ou ceux des locuteurs non natifs. Cette propriété caractérise également notre corpus, composé d'articles de recherche. Enfin, les troisième et quatrième propriétés de Sinclair (*ibid.*), *la simplicité et la documentation*, concernent les aspects techniques. La simplicité signifie surtout que le texte proprement dit du corpus doit afficher un format simple et facilement séparable des annotations ainsi que des autres informations et remarques ajoutées. De cette remarque, nous déduisons aussi que le corpus est généralement au format électronique (cf. McEnery & Wilson 1996 : 223-24). La documentation, finalement, assure que les informations sur les documents sont séparées des documents eux-mêmes, que « full details about the constituents of a component are kept separately from the component itself » (« Les détails complets des constituants d'un composant sont maintenus séparés du composant lui-même. ») (Sinclair 1996).

Dans la pratique, ces quatre critères sont, ou non, remplis par un nombre de corpus qui existent surtout pour l'anglais. Le Brown corpus²⁷, composé de 1 014 312 mots, représentant « edited English prose printed in the United States during the calendar year 1961 » (« prose anglaise éditée, imprimée aux États-Unis durant l'année calendaire 1961 ») (site web du projet) constitue un corpus classique. Le British National Corpus (BNC)²⁸, en revanche, est un corpus très étendu, intégrant 100 million de mots dont la dernière version a été publiée en 2007. Le BNC inclut des parties écrites et orales, représentant « a wide cross-section of British English from the later part of the 20th century » (« Une ample section croisée de l'anglais britannique depuis la fin du XXème siècle »). Pour le français, le Frantext²⁹, composé de 500 œuvres de la littérature française couvrant en tout 146 693 289 caractères, propose un corpus similaire. Ces corpus offrent des ressources significatives pour plusieurs types d'études. Cependant, une lacune les rend difficilement applicables dans un nombre d'études : ils sont composés d'extraits de textes

²⁷ <http://icame.uib.no/brown/bcm.html>

²⁸ <http://www.natcorp.ox.ac.uk/corpus/index.xml>

²⁹ <http://www.cnrtl.fr/corpus/frantext/>, www.frantext.fr

et ne permettent pas l'examen des occurrences dans leur contexte textuel. Ce défaut, également relevé par Condamines (2003 : 31-32) et Péry-Woodley (2005 : 188), constitue surtout une pierre d'achoppement pour les études de l'analyse du discours et de l'organisation textuelle, requérant un contexte textuel vaste voire des textes entiers pour leurs analyses.

Comme discuté en section 2.1, le défi de l'analyse quantitative de l'organisation textuelle est qu'elle est difficilement repérable automatiquement dans le texte. Cette difficulté est souvent surmontée par l'annotation manuelle du phénomène étudié. Des corpus annotés de la sorte ont été développés, notamment selon le modèle RST de Mann et Thompson (1988). Le résultat, *The RST Discourse Treebank* (Carlson *et al.* 2003), est composé de plus de 176 000 mots d'articles du *Wall Street Journal*. Le *Penn Discourse Treebank* (Miltsakaki *et al.* 2004 ; Prasad *et al.* 2008) propose un autre corpus de ce type, construit sur le *Penn Treebank*, un corpus doté d'une annotation syntaxique (Marcus *et al.* 2004). Contrairement au corpus RST, le PDTB reste neutre par rapport aux différentes approches théoriques et part de quatre relations de cohérence de base, qui peuvent être explicitement ou implicitement introduites. Le PDTB comprend un million de mots du journal *Wall Street* et inclut 40 600 relations sémantiques annotées. Pour le français, il n'existe pas de corpus annotant les relations de cohérence, en excluant le résultat visé par le projet Annodis (Ho Dac *et al.* 2009). Les seuls corpus proposant une annotation discursive sont la *Collection de corpus du français contemporain*³⁰ et le corpus du projet Ananas³¹ qui, tous deux, sont annotés pour les relations anaphoriques.

5.1.2 CORPUS DE L'ETUDE : CRITERES DE COLLECTE ET TAILLE FINALE

Toute étude en corpus suppose une caractérisation de l'ensemble de texte étudiés, de manière à se donner les moyens d'évaluer la portée des descriptions ou des propositions avancées ; *à fortiori*, toute étude qui vise la comparaison de corpus à partir de l'hypothèse d'une différence spécifique en rapport avec le phénomène étudié doit pouvoir justifier le choix des sous-corpus comparés.

Péry-Woodley (2005 : 200)

Le corpus de notre étude se compose, comme vu en 2.2, d'articles de recherche en linguistique, en éducation et en histoire. Afin que le corpus respecte les critères de Péry-Woodley (ci-dessus) et ceux discutés dans la section précédente, nous avons lors de la collecte tenu compte de plusieurs aspects potentiellement problématiques. Il s'agit surtout des traits susceptibles d'affecter l'usage des séries linéaires et le marquage de l'organisation textuelle, en particulier donc la langue, la revue et le domaine spécifique de rédaction de l'article (cf. section 2.2). Dans la suite, nous discuterons de ces aspects et, enfin, nous aborderons brièvement les questions liées à la taille du corpus. Il importe de noter que la collecte implique des choix et des compromis : respecter simultanément tous

³⁰ http://catalog.elra.info/product_info.php?products_id=634&language=fr

³¹ <http://www.atilf.fr/ananas/>

les critères de collecte n'est pas toujours possible. Le corpus résultant peut être considéré comme un compromis, la meilleure solution possible dans les circonstances.

Tout d'abord, l'accessibilité de l'article au format électronique a constitué, pour nous, un critère d'ordre pratique absolu. Étant donné notre méthode et le fait d'être seule pour la collecte, la transcription manuelle des articles n'aurait pas été possible. De plus, idéalement, les articles inclus devraient tous être librement accessibles sur Internet. Dans la pratique, pourtant, le choix final des revues et des articles collectés est un compromis entre leur accessibilité et leur représentativité. Bien que les revues électroniques soient toujours plus communes, surtout *via* le portail *revues.org*, elles n'étaient pas encore très fréquentes à l'heure de la collecte, ce qui a limité le nombre de revues accessibles.

Abordons maintenant les questions sur la langue et la revue. Pour la langue, la réponse paraît simple : c'est le français. Or, afin d'assurer sa qualité, quelques aspects nécessitent une discussion, le plus important étant peut-être le niveau linguistique du ou des auteurs. En théorie, il serait facile d'exiger que la langue maternelle de l'auteur soit le français. Dans la pratique, pourtant, comme le mentionnent Fløttum *et al.* (2006 : 10), la nationalité ou la langue maternelle ne peuvent pas être déduites avec certitude du nom de l'auteur (voir aussi Bondi 2009a : 87). De surcroît, la problématique de la distinction entre un locuteur natif et non natif suscite de nombreuses discussions depuis quelques années (cf. Jenkins 2003 : 80-83 ; Lesznyák 2004 : 25-31 ; Mauranen 2009). Bien que ces études portent surtout sur l'anglais comme *lingua franca* ou langue internationale, la même difficulté de définition du locuteur natif et des compétences nécessaires à ce statut est présente pour le français. Pour ces raisons, l'inclusion ou l'exclusion d'un article sur le fondement du nom de l'auteur nous a semblé superficielle.

Bien que la question de la langue maternelle de l'auteur soit délicate, nous soulignons que l'objet de notre étude est la façon d'écrire français. Par conséquent, nous avons décidé d'inclure dans le corpus seulement les articles dont l'auteur, ou au moins l'un des auteurs, travaille dans une université française ou francophone. Sur le fondement de cette restriction, nous pouvons exclure les articles rédigés au sein d'une culture risquant de différer gravement de la culture française, ou francophone, tout en bénéficiant d'une interprétation plus aisée et départie de préjugés sur le nom de l'auteur.

L'usage du français dans le monde scientifique ne se limite pas à la France. Pour cette raison, le corpus inclut également divers articles écrits par des chercheurs suisses et belges ainsi que quelques revues canadiennes. La question se pose du caractère notable de la différence des revues canadiennes en regard des françaises, susceptible d'affaiblir la représentativité du corpus. Or, discutant de ce même sujet, Fløttum *et al.* (2006 : 10-11, 118) constatent que les textes écrits dans une même langue se ressemblent et que cette similarité semble dépasser les frontières nationales. Poursuivant, ils ajoutent que la norme pour tous les textes écrits en français semble être la norme française. Par conséquent,

aucune différence notable entre les revues francophones et françaises ne devrait être relevée.

Enfin, il faut également tenir compte du fait que le ou les auteurs ne sont pas nécessairement seuls à avoir influencé le format et le contenu de l'article ; la revue peut aussi éditer l'article selon ses propres normes. La définition de l'auteur réel du texte et la distinction entre les parties éditées et celles rédigées par les auteurs est pourtant impossible et, pour nous, aussi inutile. En effet, l'objet de notre intérêt est le produit final du processus d'édition, c'est-à-dire l'article. En outre, ce processus d'édition constitue plutôt un atout dans la collecte à nos yeux. De fait, en éditant les articles selon leurs normes, les revues les ajustent également mieux aux critères convenus de la communauté scientifique, assurant ainsi leur représentativité dans le domaine. De la sorte, ce processus rend, par exemple, la question de la langue maternelle de l'auteur secondaire.

Pour le choix des revues du corpus, nous avons appliqué plusieurs critères principaux. Afin de garantir le niveau scientifique du corpus, les revues choisies pratiquent toutes un processus d'évaluation. Les différents actes de colloque du corpus sont suivis d'un colloque avec un comité scientifique. Tous les articles inclus sont également publiés depuis l'an 2000.

Le sujet des articles choisis doit également être pris en compte, car les différents sous-domaines, tels que la linguistique appliquée et la linguistique théorique, peuvent présenter des différences. Pour la sélection, nous avons d'abord évité les articles dont le sujet risque de représenter un sous-domaine particulier, accroissant ainsi l'hétérogénéité du corpus, notamment les études linguistiques très techniques sur le traitement automatique des langues ou très appliquées sur le français en tant que langue étrangère. De plus, nous avons préféré les articles présentant leur propre recherche et ne se limitant pas uniquement à discuter celles des autres (cf. aussi Fløttum *et al.* 2006 : 8).

Afin d'éviter la concentration du corpus sur un échantillon très restreint et, possiblement, sur certaines thématiques spécifiques, le nombre de revues de chaque sous-corpus (linguistique, éducation et histoire) doit être suffisant. Par ailleurs, si les revues différentes sont trop nombreuses, le corpus risque de devenir dispersé. Le nombre de revues dans chaque sous-corpus dépend également de la taille finale du corpus. La quantité nécessaire est pourtant difficile à déterminer préalablement pour une étude qui, comme la nôtre, manque de précurseurs. En effet, le nombre de séries linéaires est très difficile à estimer avant l'annotation. Parmi les travaux antérieurs, le corpus KIAP (Fløttum *et al.* 2006) comporte 450 articles de recherche (2 250 868 mots) et le corpus de Hyland 1998a 28 articles (160 000 mots). Les corpus annotés *RST Discourse Treebank* (Carlson *et al.* 2003) et *Penn Discourse Treebank* (Miltsakaki *et al.* 2004, Prasad *et al.* 2008) intègrent 176 000 et un million de mots, respectivement. Ces chiffres sont pourtant difficiles à comparer à ceux de notre étude du fait des problématiques, méthodes et ressources différentes.

Le corpus final de notre étude se compose de 796 072 mots, soit 90 articles de recherche (30 en linguistique, 30 en éducation et 30 en histoire). Ce nombre d'articles nous a semblé suffisant pour une étude comparative tout en restant compatible avec leur annotation par nous-même. Un corpus plus important aurait nécessité un annotateur additionnel ou une prolongation importante du projet de thèse. Revues et tailles des sous-corpus sont présentées dans le Tableau 1 ci-dessous et plus en détail dans les annexes.

Discipline	Revue	Nb d'articles de la revue dans le corpus	Nb total de mots	Nb de mots / article (moyenne)
ling	Cahiers de grammaire	7	44 775	6 396
ling	Corpus	8	77 936	9 742
ling	French Language Studies	4	36 297	9 074
ling	Marges linguistiques	3	41 891	13 964
ling	Linguisticae investigationes	1	9 474	9 474
ling	Revue romaine	7	59 943	8 563
Sous total		30	270 316	9 011
édu	Actes du colloque « Innovations, Usages, Réseaux »	4	20 189	5 047
édu	MJE	7	66 207	9 458
édu	Recherches et éducations	5	26 340	5 268
édu	Spirale	7	33 823	4 832
édu	Revue des sciences de l'éducation	7	71 469	10 210
Sous total		30	218 028	7 268
his	Annales historiques de la Révolution française	5	57 200	11 440
his	Cahiers de recherches médiévales	8	79 937	9 992
his	Cahiers d'histoire	5	33 525	6 705
his	Revue d'histoire du IX siècle	6	62 613	10 436
his	Revue d'histoire de l'Amérique française	6	74 453	74 454
Sous total		30	307 728	10 258
Total		90	796 072	8 845

Tableau 1 Sous-corpus de l'étude.

Pour garantir la variation thématique et stylistique, chaque sous-corpus inclut 4 à 8 articles issus de 5 à 6 revues différentes. Bien que nous ayons évité les articles traitant un sujet très technique ou très appliqué, nous n'avons pu empêcher l'inclusion de plusieurs articles plutôt techniques ou appliqués dans le corpus. Dans l'ensemble, leur nombre reste

cependant minime. Nous pouvons même nous demander si la variation suscitée par ces articles dans le corpus dépasse la variation personnelle causée par le style unique de chaque auteur (cf. aussi Fløttum *et al.* 2006 : 8, Fløttum 2006). En outre, cette différence possible entre les articles plutôt techniques ou appliqués et ceux plus typiques peut être testée par des méthodes statistiques. La même procédure peut être appliquée pour confirmer la similitude ou la divergence entre les articles publiés dans des revues canadiennes et françaises.

5.2 METHODE

Comme nous l'avons mentionné dans 2.1, ce travail est une étude de corpus reposant surtout sur des méthodes quantitatives. Par ailleurs, nous utilisons également des analyses qualitatives afin d'interpréter les résultats et pour un examen plus détaillé des aspects difficilement repérables automatiquement. Comme l'affirment Biber et al (1998 : 5, 9-10), les études quantitatives doivent aussi dépasser les simples fréquences d'occurrences et interpréter les résultats par des méthodes qualitatives, complétant les approches quantitatives. À la suite, nous clarifions la méthode quantitative que nous avons appliquée dans l'analyse des séries linéaires. Nous commencerons par présenter la méthode pratique et poursuivrons par la définition du schéma d'annotation qui définit les critères que les séries linéaires annotées doivent remplir et les classes des marqueurs utilisées dans l'analyse de la façon dont les items, les amorces et les clôtures sont signalés.

5.2.1 DEMARCHE PRATIQUE

Dans la pratique, la démarche de notre méthode comporte deux étapes, la première comprenant l'annotation des articles et la deuxième les analyses poursuivies sur le fondement de ces annotations.

La préparation des articles de recherche, d'abord enregistrés au format .txt lisible par l'outil d'annotation, constitue le premier pas de l'annotation. L'étape préparatoire inclut également un prétraitement des marqueurs pouvant fonctionner comme indices de sériation, tels que les MIL et les numéros ainsi que les connecteurs et autres expressions fréquents dans les amorces et les clôtures. Afin de rendre ces expressions plus facilement visibles dans le texte, et faciliter ainsi l'annotation, nous faisons précéder automatiquement ces expressions de la figure XX en employant un script Unix.

Le marquage des possibles indices de sériation par XX a, selon notre expérience, effectivement facilité l'annotation. Il s'avère donc ainsi justifié. Une partie de l'étape de préparation a pourtant été menée de manière inappropriée. À l'origine, les articles adoptaient un format soit pdf, soit html. De ce fait, les articles en html auraient pu, pour le moins, être annotés également dans leur format d'origine. Nous aurions pu ainsi conserver automatiquement quelques informations typographiques, telles que les titres et les

changements de paragraphe. Comme cela n'a pas été le cas, ces informations ont par conséquent été perdues.

Après la transformation des articles au format .txt et l'ajout des XX aux endroits nécessaires, nous avons annoté les articles avec l'outil d'annotation Callisto³². Parmi les différents outils XML, Callisto nous a paru le plus approprié. En effet, il a été développé surtout pour l'annotation des phénomènes linguistiques et bénéficie d'un manuel clair et exhaustif. Callisto s'avère également très convivial et adaptable à des tâches différentes. Nous l'avons personnalisé pour l'annotation des séries linéaires avec une fiche de style DTD. La Figure 3 ci-dessous illustre la tâche d'annotation avec Callisto. Le texte à annoter apparaît au milieu et les séries linéaires annotées et numérotées en bas de l'écran.

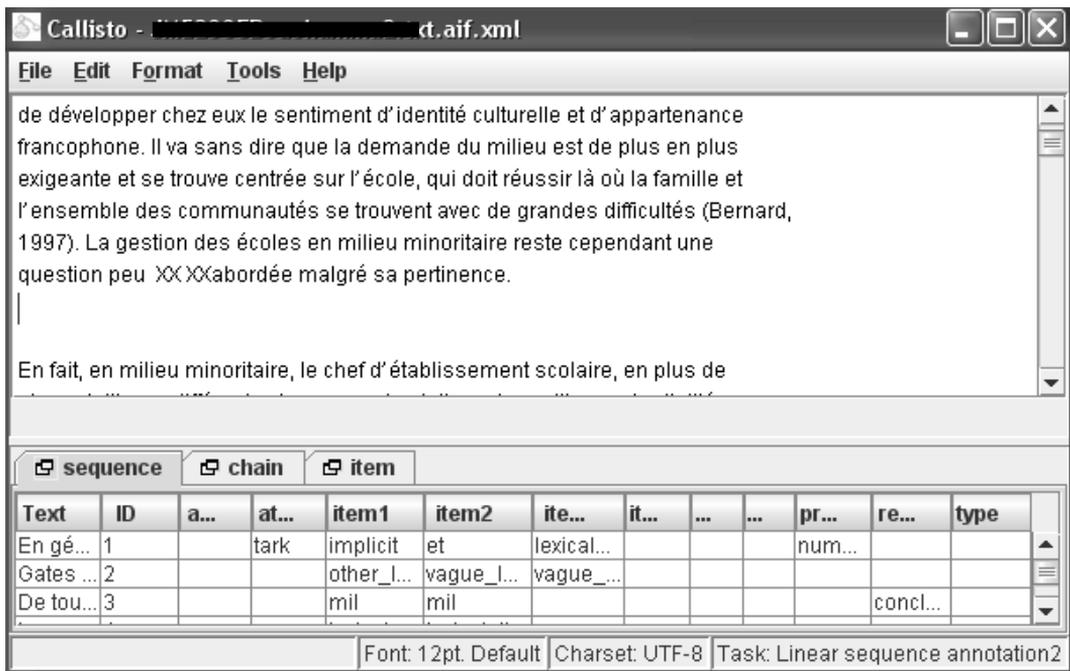


Figure 3 Le logiciel Callisto.

Nous avons annoté tous les articles et vérifié puis corrigé les annotations complétées, au moins à deux reprises. De plus, nous avons analysé les cas problématiques conjointement avec notre directeur de thèse. Comme nous n'avons pas eu la possibilité de mener à bien une double annotation par plusieurs personnes, pour mesurer ainsi l'accord inter-annotateur, nous avons jugé ces discussions nécessaires afin de fiabiliser l'analyse. Nous n'avons pas mesuré le temps consacré à l'annotation. Nous estimons, toutefois, que l'annotation d'un article avec sa double vérification et correction, selon notre procédure, a requis plus d'une journée de travail.

³² callisto.mitre.org

Les articles annotés sont sauvegardés au format XML. Les propriétés annotées dans ce format peuvent être analysées automatiquement d'une façon similaire aux mots individuels. Le schéma d'annotation est présenté en section 5.2.2. L'Exemple 17 illustre cependant une simplification du code d'une série linéaire annotée.

Exemple 17

```
<?xml version="1.0" encoding="UTF-8"?>
<annotation>
  <corpus>
    <chain>
      <sequence ID="4" arguable="*null*" >
        <item ID="0(4)" arguable="*null*" marking="vague" type="prediction">
          De XX nombreux départements actifs dans le pétitionnement répondent à ce modèle
          gardois.
          </item>
        <item ID="1(4)" arguable="*null*" " marker="*null*"
          marking="other_lexical_cohesion" type="item">
          Dans la Sarthe, 11 800 signatures sont récoltées dans 90 communes, dont seulement
          16 % au Mans.
          </item>
        <item ID="2(4)" arguable="*null*" marker="egalement"
          marking="vague_linear_connector" type="item">
          Dans le Bas-Rhin également, seuls 22 % des 12 252 signatures ont été recueillies à
          Strasbourg.
          </item>
        <item ID="R(4)" arguable="*null*" marking="anafora"
          type="restrospection">
          Dans ces départements, le pétitionnement se caractérise par la prépondérance de
          l'initiative villageoise et la dissémination rurale. Le pétitionnement permet alors
          d'apprécier l'organisation et la capacité d'action du parti républicain au niveau local.
          On entrevoit déjà certains des traits caractéristiques de l'insurrection de décembre
          1851. Face au contrôle policier croissant qui s'exerce sur les villes, l'activité
          montagnarde se replie dans les campagnes.
          </item>
        </sequence>
      </chain>
    </corpus>
  </annotation>
```

Le XML de l'Exemple 17 illustre l'annotation d'une série linéaire avec les informations nécessaires pour son analyse. Les informations introduites lors de l'annotation sont en gras. Les éléments annotés dans l'exemple sont donc la *chain*, qui peut inclure une ou plusieurs *sequences*, c'est-à-dire des séries, qui se décomposent en *items*. La *chain* est une structure spécifiée seulement si elle inclut deux ou plusieurs séries, soit si elle remplit certains critères spécifiques indiqués dans le schéma d'annotation ci-dessous. La *sequence* inclut deux *attributs* : le numéro de la série, c'est-à-dire l'*ID* et l'attribut *arguable*, auquel nous pouvons donner, par exemple, la valeur *questionnable* si la série semble douteuse. Les *items* de la *sequence* sont également numérotés : 0 pour l'amorce, des numéros pour les items proprement dits et R pour la clôture. De même que les *sequences*, les items

possèdent aussi l'attribut *arguable*. Les autres informations présentes sont *marker* offrant l'option de marquage manuel du marqueur utilisé, *marking* proposant la classe du marqueur de l'item et *type* spécifiant si l'item est une amorce, un item proprement dit ou une clôture. Le texte couvert par le composant suit les informations marquées. Il est à noter, d'une part, que ces attributs (*ID*, *arguable* et autres) ne sont pas les seuls que Callisto permet d'annoter et, d'autre part, que d'autres informations sont également inscrites dans la source XML. Les attributs présentés dans l'exemple sont pourtant ceux qui sont le plus souvent utilisés dans l'analyse. Les autres seront précisés lors de l'analyse, en tant que besoin.

Dans l'analyse des annotations, nous avons exploité principalement de courts programmes *python* rédigés par nous-même et de *R*, un environnement d'étude des données statistiques³³. La Figure 4 ci-dessous illustre une analyse *python* de l'annotation. Les lignes numérotées de 1 à 8 correspondent chacune à une série linéaire, ces premières chiffres présentant le numéro identifiant de la série que nous lui avons attribué lors de l'annotation (*sequence ID* dans l'exemple précédent). La deuxième colonne indique la longueur de la série calculée par le programme python et **null** que la série n'a pas de fonction spécifique attribuée lors de l'annotation³⁴.

Les colonnes suivantes traitent surtout du marquage des séries. *PNM* et *HOMO* sont des classes de séries définies par le programme, *PNM* présentant que la série inclut des items explicitement signalés et des items non marqués, *HOMO* que tous les items de la série sont signalés par le même type de marqueurs. Les colonnes suivantes, par contre, se basent sur l'annotation manuelle des classes de marqueurs utilisées dans l'introduction des items, des amorces et des clôtures. Ces classes sont définies dans le schéma d'annotation que nous présentons dans la section suivante. *0=...* et *R=...* décrivent des types d'amorce et de clôture présents dans la série, *C*, *L*, et *CL* que les marqueurs utilisés dans la série sont des connecteurs (*C*), des constructions numéro+nom (*L*) ou des deux (*CL*). *E*, *V*, et *EV*, encore, définissent l'ordre exprimé par les marqueurs d'item : *E* correspond à l'ordre exact, *V* à l'ordre vague ou l'addition et *EV* indique que les marqueurs d'item de la série font partie des deux groupes. *HISQC*, en revanche, définit que la série est issue d'un article du sous-corpus histoire et que la revue est parue au Canada. La dernière colonne, enfin, précise la position des items non marqués dans la série. Quand la série n'inclut pas d'item non marqué, cette colonne porte l'étiquette *EX*.

³³ <http://www.r-project.org/>

³⁴ Cf. section 6.1

```

mave1a@babelfish ~/data0410 $
mave1a@babelfish ~/data0410 $ cat trans/his/RevuedHistoiredAmerique
Fr57\4\Histon2.txt.aif.xml | ./xml2data.sh

1 147 *null* PNM 0=0 R=0 CL V HISQC 1
2 530 *null* PNM 0=0 R=ANA L V HISQC 1
3 51 *null* HOMO 0=0 R=0 C E HISQC EX
4 161 *null* PNM 0=Q R=0 C V HISQC 1
5 630 *null* PNM 0=0 R=0 CL EV HISQC 1
7 150 *null* PNM 0=VAG R=0 L V HISQC 12
8 153 *null* PNM 0=0 R=0 C V HISQC 1
mave1a@babelfish ~/data0410 $ █

```

Figure 4 Exemple de l'analyse *python* du XML annoté

Une fois analysées à l'aide de *python*, les annotations XML sont faciles à examiner d'une manière quantitative à l'aide du logiciel R. De plus, quand nécessaire, les annotations peuvent être soumises à des nouvelles analyses *python* pour compléter les résultats. Ainsi, une fois accomplies, les annotations manuelles permettent une analyse relativement aisée et flexible du corpus.

Enfin, pour terminer, nous présentons encore la démarche pratique de notre méthode à l'aide de la Figure 5 ci-dessous.

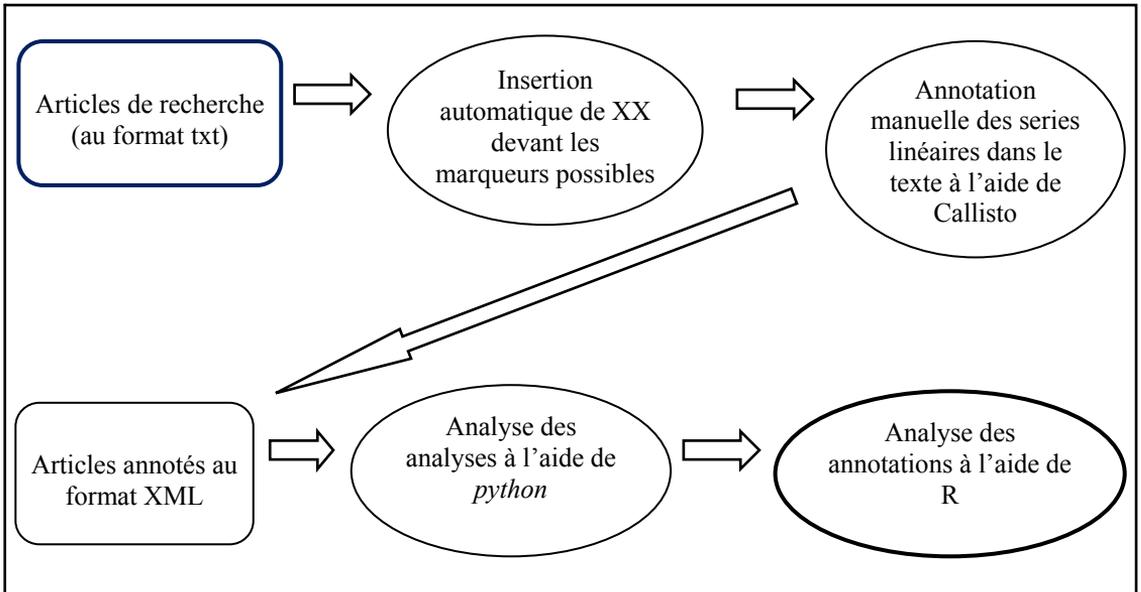


Figure 5 Démarche pratique de notre méthode

Comme l'illustre la Figure 5, le point de départ de notre méthode est les textes de notre corpus, soit les articles de recherche choisis et sauvegardés au format txt. Après

l'annotation manuelle des séries linéaires, le résultat, c'est-à-dire les articles au format XML, sont analysés automatiquement par des programmes python et par le logiciel R.

5.2.2 SCHEMA D'ANNOTATION

Nous avons élaboré le schéma d'annotation pendant les analyses préparatoires. La version finale s'est stabilisée pendant les premières annotations. Après quelques changements, nous avons parcouru toutes les annotations afin de les ajuster toutes selon le même schéma. Ce schéma se compose de deux parties : les définitions des marqueurs individuels, d'une part, et les critères que les séries annotées doivent remplir, de l'autre. Comme nous l'avons expliqué dans 4.2.3, afin de fiabiliser l'analyse, l'annotation doit se fonder sur des indices linguistiques.

La section commencera par discuter divers critères de classification que nous appliquons à l'annotation des marqueurs individuels, des amorces et des clôtures. Ensuite, nous présenterons cette classification et les critères de surface établis pour les séries linéaires analysées.

5.2.2.1 Quelques critères de classification des marqueurs d'items individuels

L'annotation des marqueurs individuels tend à généraliser des propriétés de ces indices afin de pouvoir les classer en groupes. Dans notre cas, l'analyse des marqueurs des composants des séries linéaires, deux propriétés générales nous ont semblé essentielles : l'ordre, ou le nombre, d'items exprimés par le marqueur et le type de marqueur employé.

La question de l'ordre exprimé est pertinente³⁵. En effet, selon notre définition (cf. 4.2.2 et 4.3.2), les items des séries linéaires doivent être au moins partiellement ordonnés. Les amorces et les clôtures exprimant le nombre d'items soulignent également la structure de la série et la définissent plus en détail. Au sens large, nous pouvons distinguer deux degrés d'ordre que les marqueurs d'item, d'amorce et de clôture peuvent exprimer : l'ordre absolu, ou exact, indique la place exacte de l'item dans la série (*d'abord, finalement, le troisième*) ou le nombre exact d'items (*cette série comprend deux items, Ces deux items*), tandis que l'ordre relatif, ou additif, note seulement l'addition (*de plus, un autre + NOM*) ou l'organisation du texte à venir dans une série (*cette série comprend quelques items*). Une clôture indiquant l'ordre relatif peut être une expression référentielle du type *ces items*.

Les différents types de marqueur, en revanche, sont nécessaires à la distinction du fait des fonctions différentes qu'ils revêtent dans l'organisation du texte (cf. 3.4). Les expressions anaphoriques, par exemple, se réfèrent toujours au texte antérieur tandis que les adverbiaux détachés à l'initiale de la phrase peuvent adopter une fonction cadrative et ainsi étendre leur zone d'influence au-delà de la phrase d'origine. Les MIL, en revanche, peuvent relier

³⁵ Voir aussi la discussion sur ce concept par Schnedecker (2001: 262-266).

et indexer des segments mais, contrairement aux adverbiaux, ils ne peuvent ouvrir d'univers de discours.

Pour les items, nous distinguons également quelques marqueurs qui n'expriment pas directement l'ordre. Les connecteurs adversatifs, tels que *mais* et *par contre*, sont analysés comme des marqueurs d'items explicites puisqu'ils peuvent aussi adopter une fonction additive. Les items introduits par des indices qui ne signalent aucun ordre ou addition sont, en revanche, analysés comme non marqués. De fait, ces indices ne peuvent pas introduire seuls les items d'une série dont l'ordre est une propriété essentielle. Ainsi, les connecteurs thématiques, tels qu'*en ce qui concerne*, ne comptent pas comme marqueurs d'items explicites. Or, ces indices peuvent tout de même contribuer à l'identification des items et à l'organisation de la série. C'est pourquoi nous les considérons comme des *indices complémentaires de sériation*, pouvant être présents dans les items que nous considérons pourtant non marqués. Outre les connecteurs, les titres et la répétition de mêmes items lexicaux dans plusieurs items constituent d'autres indices complémentaires pouvant contribuer à l'identification de l'item mais sans signaler l'ordre ou l'addition.

Dans 4.1, nous avons discuté du modèle de Goutsos (1996) qui comprend le marquage explicite et le non marquage d'une façon différente de celle des travaux sur la cohérence. Notre définition des items non marqués ou introduits implicitement (ci-dessus) s'approche cependant de ces derniers. En effet, pour Goutsos (1996), les items « marqués » par la répétition des items lexicaux ou les connecteurs ne signalant pas l'addition ou l'ordre peuvent bien être introduits explicitement dans le texte. La différence réside pour autant dans ce qui doit être signalé : pour Goutsos, les indices nécessaires sont seulement ceux à même d'exprimer les techniques d'introduction et de continuation, tandis que nous cherchons également à indiquer l'ordre entre les items et définissons le marquage explicite à partir de ce critère. La notion d'indices complémentaires complète ce schéma des marqueurs explicites. De fait, comme nous le montrons dans les sections suivantes, ces signes jouent également un rôle dans l'introduction des séries.

Enfin, la façon dont nous avons identifié les items non marqués constitue la dernière question. Avons-nous pris en compte uniquement les séries facilement identifiables ? Ou avons-nous tenté d'inclure également dans l'analyse des structures moins cohérentes ? La réponse affecte inévitablement les séries repérées car, si nous prenons uniquement en compte les séries facilement identifiables et explicitement signalées, notre hypothèse sur la fréquence du marquage explicite et sa nature soigneuse se confirme d'elle-même par la façon dont nous menons l'analyse.

Les items non marqués sont au centre de l'analyse de la section 7.3, dans laquelle nous examinons les indices complémentaires et leur fréquence ainsi que les items dans lesquels même ceux-ci ne sont pas utilisés. Nous admettons cependant que l'identification des items non marqués et l'inclusion des séries moins cohérentes dans l'analyse sont difficiles à mener de façon neutre à l'aide de nos méthodes. En principe, nous avons pris en compte

toutes les séries qui remplissent les critères posés ci-dessous et dans lesquelles nous avons pu identifier au moins deux items. Or, si nous n'avons identifié qu'un seul item ou si la distinction des items nous a paru autrement impossible, nous n'avons pas inclus la série dans l'analyse. Une telle série, sans structure apparente, est en effet très difficile à analyser. Ce choix restreint déjà inévitablement le nombre de séries incohérentes incluses dans l'analyse et influence ainsi nos résultats. Nous le considérons néanmoins comme la seule option possible, les méthodes que nous appliquons ne permettent pas de considérer les séries sans structure interne, divisible en items. '

5.2.2.2 Classes des composants des séries linéaires

Lors de l'annotation, nous distinguons les items, les amorces et les clôtures. De plus, nous précisons manuellement la classe du marqueur utilisé pour chacun de ces éléments. Ces classes peuvent ensuite être quantifiées et examinées lors de l'analyse.

Comme nous l'avons précisé dans la section précédente, les deux propriétés principales dans notre classification des marqueurs sont le type de marqueur utilisé et l'ordre exprimé par celui-ci. En tout, pour les marqueurs d'item proprement dits, notre annotation distingue 13 classes différentes, et pour les amorces et clôtures, quatre classes respectivement. Dans ce qui suit, nous définirons les classes de marqueurs que nous avons utilisées dans l'annotation. Nous commencerons par les marqueurs d'item proprement dits. Dans ceux-ci, nous distinguons encore les marqueurs utilisés dans les items explicitement signalés et les indices complémentaires qui peuvent contribuer à l'identification des items que nous analysons comme non marqués. Ensuite, nous continuerons par présenter les types d'amorce annotés et finirons par les types de clôtures.

MARQUEURS D'ITEM EXPLICITES

Connecteurs et expressions pouvant adopter une fonction cadrative

1. MIL exacts indiquant la place exacte de l'item dans la série :
premièrement, d'abord, deuxièmement, finalement, enfin, d'une part – d'autre part, etc.
2. MIL relatifs indiquant l'addition de l'item dans la série :
puis, ensuite, aussi, etc.
3. Adverbiaux détachés à l'initiale de la phrase (non MIL) exacts :
dans la section trois, dans le dernier exemple, etc.
4. Adverbiaux détachés à l'initiale de la phrase (non MIL) relatifs :
dans une autre section, dans l'autre exemple, etc.
5. Connecteurs adversatifs :
mais, par contre, etc.

6. Marqueurs fonctionnant en paires :
tantôt – tantôt, soit-soit.

Expressions référentielles et syntagmes nominaux fonctionnant comme marqueur d'item

7. Expressions anaphoriques :
l'un – l'autre, le premier, le troisième, le dernier, etc.
8. Constructions numéro + nom exactes et verbes exacts :
le premier exemple, la dernière section, le troisième mot, commencer, finir, etc.
9. Constructions numéro + nom et verbes relatifs :
l'autre exemple, une autre section, continuer, etc.
10. Et : *Et* est annoté séparément à cause de sa fréquence. Cette étiquette couvre également les cas rares où d'autres conjonctions de coordination, en général *ou*, sont utilisées pour marquer un item.

ITEMS NON MARQUÉS ET INDICES COMPLÉMENTAIRES

11. Répétition : *médecin – médecine, parole – parler, exemple – exemple, etc.*
12. Autres connecteurs (qu'adversatifs ou MIL) : *en ce qui concerne, quant à, etc.*
13. Non marqué proprement dit : les items ne présentant aucun des indices des classes précédentes.

Dans ces marqueurs et indices complémentaires, certains détails nécessitent encore quelques précisions. Tout d'abord, la conjonction de coordination *et* est annotée séparément car elle est très fréquente. Cette conjonction, qui, selon Adam et Revaz (1989), marque souvent le dernier item, peut également assumer d'autres fonctions (voir Suomela-Salmi 1993). Nous l'avons pourtant annotée uniquement lorsqu'elle fait partie de la série linéaire annotée. Les *constructions numéro + nom*, en outre, ne sont pas dénommées expressions référentielles ni anaphoriques car leur caractère référentiel possible n'est pas analysé lors de l'annotation. Dans ce cas, la distinction entre les marques référentielles et les autres ne se fait pas pour des raisons pratiques : la nature référentielle peut être une question complexe présentant des cas limites difficiles à déterminer.

Certaines précisions sur les indices complémentaires et le non marquage s'imposent également. Tout d'abord, sur le plan général, nous considérons comme non marqués tous les items qui ne sont pas introduits explicitement par un marqueur explicite d'ordre ou d'addition, tel que défini précédemment. Toutefois, les items non marqués peuvent être facilement identifiés par des indices complémentaires. La catégorie *autres connecteurs* inclut les connecteurs qui ne signalent aucun ordre ou addition en faisant partie des indices complémentaires et, plus généralement, du non marquage. Comme indices de répétition,

nous analysons tous les items qui incluent un même mot répété dans un autre item de la série. Le mot répété peut également se présenter sous une autre forme grammaticale, voire appartenir à une autre catégorie syntaxique : il suffit que les items répétés partagent le même morphème de base (cf. *complex lexical repetition* dans Hoey 1991 : 52-68). Enfin, la répétition des noms différents est aussi incluse dans cette catégorie.

AMORCES (adaptées à notre travail d'après Tadros 1994)

1. **Énumération exacte** spécifie le nombre d'items de la série :
Cette série consiste en trois parties...
2. **Énumération vague** ne spécifie pas le nombre d'items de la série mais indique que le segment textuel à suivre est organisé en items :
Cette série consiste en quelques parties...
3. **Question** :
Comment la série est-elle construite?
4. **Implicite** :
Dans ce qui suit, nous allons... / Nous allons...

L'amorce implicite, selon Tadros (1994), annonce seulement que l'auteur s'engage à accomplir à un acte de discours : *Ces résultats nous amènent à distinguer...* Pour autant, puisque cette annonce peut se réaliser pour différents types d'amorces qui ne sont pas nécessairement évidents à identifier dans un texte où les phrases sont, par défaut, liées les unes aux autres, leur annotation risquerait d'être peu systématique et leur analyse quantitative peu fiable. Pour cette raison, nous avons décidé d'annoter seulement les amorces implicites incluant une expression explicite de l'intention de l'auteur dans le texte à venir. En pratique, ces expressions sont du type *dans ce qui suit, ci-dessus, le chapitre suivant, nous allons*, faciles à distinguer dans le texte. Il est vrai que cette décision exclut de l'analyse un nombre important d'amorces. Le nombre des séries exclues est cependant probablement moins significatif. De fait, la présence des amorces est décisive uniquement quand la série en question inclut des items non marqués.

CLOTURES

1. **Connecteurs conclusifs ou explicatifs** :
Donc, bref, ainsi, etc.
2. **Autres connecteurs** :
Cependant, par conséquent, etc.
3. **Reprises anaphoriques** :
Ces parties... Cette distinction...
4. **Reprises anaphoriques spécifiant le nombre d'items de la série** :

Ces trois parties...

Les clôtures implicites, qui ne font pas partie des catégories définies mais, clairement, ferment la série, sont également relativement fréquentes. Ces expressions ne sont pourtant pas en général annotées à cause de leur variation qui risquerait de produire des annotations peu fiables. Comme seule exception à cette règle, nous annotons les clôtures implicites dans les séries dénommées *argumentatives*, dont le premier item correspond à la thèse première, le second aux données (prémises) et la clôture à la nouvelle thèse (cf. Toulmin 1958) ; Adam 1992 (cf. section 6.1). Sans cette exception, une partie importante de ces structures resterait non annotée bien qu'elles attestent clairement du même type de séquence.

Tadros (1994) notait déjà que plusieurs amorces peuvent s'enchaîner. Il en va de même pour les marqueurs d'item et les clôtures. Dans ce cas, un ordre de préférence des types annotés s'impose. Notre liste de préférence s'étend du plus spécifique au plus imprécis et implicite. De cette manière, le marquage le plus spécifique est annoté.

Pour les items, l'ordre est le suivant :

1. marqueurs exacts (*premièrement, le premier, enfin...*)
2. marqueurs relatifs (*puis, ensuite, autre*)
3. *et*
4. connecteurs en paire (*tantôt – tantôt*)
5. connecteurs adversatifs (*mais*)
6. autres connecteurs (*en ce qui concerne*)
7. répétition
8. marquage implicite

Pour les amorces :

1. énumération exacte
2. énumération vague
3. question
4. amorce implicite, « advanced labelling »

Pour les clôtures :

1. reprises spécifiant le nombre d'items
2. autres reprises, connecteurs conclusifs et reformulatifs

Pour finir, nous devons souligner que toutes les classes que nous venons d'énumérer ne sont pas considérées dans toutes les analyses que nous présenterons dans les sections à venir. Par exemple, les adverbiaux détachés à l'initial de la phrase sont souvent analysés ensemble avec les connecteurs, et aussi la présence des indices complémentaires dans les items non marqués est souvent dépassée et les items analysés simplement comme non marqués. La distinction de ces classes lors de l'annotation permet pourtant leur analyse quand nécessaire. Par exemple, bien que les adverbiaux détachés à l'initial de la phrase

soient en général analysés ensemble avec les connecteurs, dans la section 7.2.3 sur l'usage des marqueurs d'item individuels nous gardons les deux types de marqueurs à part.

5.2.2.3 Critères d'annotation des séries linéaires simples

Afin d'être incluses dans l'analyse, les séries doivent remplir les trois critères suivants :

1. La longueur minimale de la série est de deux phrases, commençant par une majuscule et se terminant par un point. Les amorces et clôtures possibles sont comprises dans cette longueur.
2. Au moins un item de la série doit être explicitement introduit par un marqueur d'addition ou d'ordre (La conjonction de coordination *et* ne suffit pas). Si le marqueur ne signale que l'ordre relatif entre les items et pas sa place exacte, la série doit également être précédée par une amorce ou suivie par une clôture pour remplir nos critères.
3. Si tous les items de la série sont introduits non marqués (ou le seul marqueur utilisé est *et*), le nombre d'items de la série doit être précisé, avant la série, par une amorce ou, après la série, par une clôture.

La longueur minimale est importante à définir. Sans elle, la phrase *J'ai deux chiens, un noir et un blanc* pourrait aussi être interprétée comme une série linéaire avec une amorce et deux items non marqués. La limite minimale de deux phrases nous a paru justifiée car l'objet de notre étude est l'organisation du texte cohérent, tandis que les phrases individuelles demeurent organisées par la syntaxe. Cette limite est également facile à appliquer. Une autre option aurait été d'exiger que chaque item inclue un verbe. Ce critère aurait pourtant posé des difficultés : comment faire, ensuite, si un item mais pas tous les items incluent des verbes ? Et que faire des verbes infinis et des subordonnées ?

Les critères 2 et 3 servent à inclure dans l'analyse les séries organisant des items au moins partiellement ordonnés et à exclure celles pour lesquelles l'ordre n'est pas du tout un critère d'organisation. Le schéma à thèmes dérivés du modèle de la progression thématique de Danes (1974), composé d'un hyperthème et de sous-thèmes dérivés de celui-ci (cf. 4.2.1) offre un exemple d'une telle structure, ignorant l'ordre comme critère mais semblable aux séries linéaires. Lorsque les marques de surface de ces structures sont insuffisantes selon les critères que nous venons d'indiquer, les structures ne sont pas analysées. Notre choix d'exclure les structures ayant un marqueur relatif comme seul signal explicite, en revanche, a pour but de bannir de l'analyse les usages individuels des marqueurs. Pour illustrer ce critère, la série introduite par *Cet article comprend quelques éléments importants... 0... De plus...* peut être incluse dans l'analyse tandis qu'une série dont le seul marqueur explicite est *de plus* en est exclue. La série introduite par *0... 0... Enfin...* peut toutefois être prise en compte car le connecteur exprime un ordre absolu.

La fin des séries constitue également un aspect important de l'annotation. Comme le note Charolles (1997), la fermeture du cadre et la fin de l'univers de discours se sont pas

toujours aisément identifiables et peuvent résulter d'un faisceau d'indices. D'après notre expérience, il en va de même pour la fin des séries linéaires. En effet, l'étendue du dernier item ou de la clôture est, dans plusieurs cas, difficile à déterminer étant donné que, dans un texte cohérent et organisé, les phrases sont normalement liées les unes aux autres. Dans les séries déséquilibrées, le dernier item risque également de paraître excessif en regard des autres items : il peut notamment initier une chaîne de référence finissant par embrasser un segment plus vaste que toute la série antérieure.

Malgré les difficultés, l'annotation fiable et systématique des longueurs des séries linéaires est importante afin qu'elles ne soient pas arbitraires et puissent être analysées. Dans une bonne partie des séries, la fin est identifiable par des repères naturels, tels que la fin du paragraphe ou de la section, également des réalisations des techniques de fermeture et d'encadrement chez Goutsos (1996). Nous avons interprété ces indices comme des signes de la fin de série également lorsqu'elle risque d'être difficile à prouver. Autrement dit, si la fin est difficile à déterminer mais proche, par exemple, d'un changement de paragraphe, nous le marquons comme la fin de la série. Dans les cas où un repère naturel ne peut pas terminer la série, nous avons décidé de marquer la fin de la série plus tôt que plus tard. Par exemple, si la thématique discutée dans le dernier item semble se poursuivre et rendre la longueur de l'item disproportionnée en regard de celle des autres items, nous avons décidé d'interrompre simplement la discussion après quelques phrases ou en fonction de la longueur des autres items de la série.

5.2.2.4 Critères d'annotation des séries enchaînées

Les séries linéaires peuvent également se combiner et former des structures plus complexes. Nous annotons séparément ces *séries enchaînées*, composées de plusieurs séries simples, puisque leur marquage et leur longueur diffèrent de ceux des séries simples. Par ailleurs, les critères d'annotation appliqués à ces structures varient selon différents aspects. Dans cette section, nous présentons brièvement ces structures et leur mode d'annotation.

Les séries enchaînées sont toujours composées d'au moins deux séries indépendantes. Les structures plus longues sont également possibles. Dans ces séries, ce qui constituerait une amorce dans une série simple énumère déjà les items de la série et devient ainsi une série en soi. Les items de la deuxième série suivent la première et se réfèrent souvent aux mêmes objets que la précédente. Un exemple simplifié d'une série chaîne est fourni ci-dessous.

Exemple 18

- 0 J'ai *deux chiens*.
- 1 Le noir est un berger
- 2 *et* le blanc est un chien de garde.
- 1 *Le premier* aime la viande.
- 2 Le blanc, par contre, ne l'aime pas.
- Cl. *Les deux* aiment pourtant le poulet.

Nous appliquons certains critères particuliers pour l'annotation des séries enchaînées. Ces critères concernent surtout la longueur et le marquage minimal des séries inférieures composant la série enchaînée : comme les séries enchaînées incluent au moins deux séries, les critères appliqués pour une seule structure sont parfois trop stricts. Tout d'abord, pour la longueur minimale, le critère de deux phrases pour les séries individuelles n'est pas pris en compte. La longueur d'une série faisant partie d'une série enchaînée peut donc être inférieure à celle-ci, tant que l'ensemble de la structure couvre cette longueur. Pour le marquage également, une seule des séries de la structure doit remplir les critères de marquage indiqués précédemment. L'introduction explicite des items des deux séries (ou plus) n'est pas nécessaire car elle est de toute façon explicite dans l'autre série.

Ces extensions des critères d'annotation augmentent naturellement le nombre total de séries incluses dans le corpus. En effet, les séries enchaînées peuvent inclure des structures qui ne seraient pas analysées en tant que séries simples. En même temps, elles permettent le repérage des structures fonctionnant comme des séries linéaires. Comme nous le constatons dans l'Exemple 18 ci-dessus, les séries enchaînées, sans respecter les critères des séries individuelles, organisent toutefois le texte d'une manière très similaire à celle des autres séries linéaires.

5.2.3 QUELQUES REMARQUES FINALES SUR LA METHODE ET LES CRITERES D'ANNOTATION

Avant de passer à la partie analyse, certains aspects de notre méthode méritent encore une remarque. Tout d'abord, nous voulons souligner que la construction et l'application d'un schéma d'annotation ne vont jamais sans compromis. Bien que s'imposant comme la meilleure option dans les circonstances, les critères de surface ne suffisent pas toujours pour la définition de séries reposant, malgré tout, parfois sur des aspects sémantiques. En conséquence, des structures très similaires aux séries linéaires sont exclues de l'analyse et d'autres très différentes sont incluses. Une structure prédite par *quelques* avec des items non marqués ne remplissant pas les critères établis peut, par exemple, parfois fonctionner comme une série plus typique qu'une série incluse dans l'analyse. De même, notre définition des amorces exclut sans doute des structures qui devraient être analysées comme des séries linéaires. Des séries avec des items non marqués et une amorce exacte, définissant le nombre d'items, sont en revanche facilement incluses dans l'analyse bien que, dans la pratique, elles forment plutôt de simples reprises renvoyant à la phrase précédente.

D'après nous, la difficulté à définir avec précision et exhaustivité les séries linéaires est due à une caractéristique fondamentale de la langue : comme remarqué lors de la discussion sur les grammaires de texte (cf. 3.1), la langue s'intègre difficilement dans les modèles préconstruits du fait de sa variation, de ses possibilités infinies de reformulation et des finesses de sens. Ces propriétés sont plutôt inscrites dans un continuum que délimitables par des critères exacts qui, même dans le meilleur des cas, sont en quelque sorte artificiels.

Nous considérons utile une autre remarque finale sur l'objectivité des méthodes quantitatives incluant une phase d'annotation manuelle. En section 2.1, nous avons constaté que les méthodes de corpus sont plus indépendantes de l'analyste, et ainsi plus objectives, que les méthodes fondées sur un nombre limité d'exemples ou sur l'introspection du chercheur. Quoique ce constat soit généralement correct, il faut ajouter que l'annotation manuelle du corpus implique nécessairement également un processus d'interprétation du chercheur malgré la définition exacte du schéma d'annotation (cf. Bestgen *et al.* 2006). Pour les séries linéaires également, l'annotation et l'identification des items non marqués et les fins des séries impliquent, surtout, des jugements en définitive subjectifs. Cette subjectivité est un écueil pourtant difficile à éviter sans restreindre l'objet de l'étude aux marqueurs explicites. Dans cette optique, il est plus productif de traiter avant tout les méthodes de contrôle de l'interprétation et de l'annotation afin de garantir les meilleurs résultats possibles. Comme nous l'avons mentionné, nous avons à cet effet analysé les cas difficiles conjointement avec notre directeur de thèse et fait précéder du chiffre XX les marqueurs possibles de sériation.

PARTIE EMPIRIQUE

6. SERIES LINEAIRES DANS LE CORPUS : FREQUENCES, LONGUEURS ET MARQUAGES

Le corpus de notre étude se compose de 796 072 mots. Selon la mise en page appliquée à ce texte, ce total s'élève à quelque 1 600 pages. Pour les séries linéaires, nous avons repéré, en tout dans ce corpus, 748 structures suivant les critères présentés auparavant, séries simples et séries enchaînées incluses. En moyenne, cette fréquence correspond à une série pour un segment de 1 064 mots, soit deux pages environ, ou neuf séries linéaires dans chaque article de recherche. Nous pouvons donc déjà remarquer le rôle important de ces structures dans l'organisation textuelle des articles de recherche.

Comme exposé en section 1.3, nos questions de recherche concernent le fonctionnement des séries linéaires dans l'organisation du texte, leur marquage et non marquage ainsi que leur variation entre les trois disciplines scientifiques examinées. Dans cette section, nous formulerons une première réponse à ces questions et explorerons les séries repérées dans le corpus afin d'en décrire les propriétés principales : fréquences, longueurs et modes de marquage. Cette description forme également la base des analyses plus détaillées.

6.1 FREQUENCES ET FONCTIONS SPECIFIQUES

La fréquence des séries linéaires est indicative de leur rôle dans l'organisation textuelle ainsi que dans la formation du texte et de la cohérence. Dans cette section, nous commencerons par examiner cela. Dans un second temps, nous examinerons la variété des séries repérées en abordant la présence et la fréquence de quelques fonctions spécifiques des séries que nous avons identifiées lors de l'annotation. Ces informations fournissent une première perspective à l'usage de ces structures dans notre corpus.

Dans son ensemble, comme le montre le Tableau 2 ci-dessous, le corpus comprend 748 séries linéaires, dont 130 enchaînées et 618 simples. Les fréquences des amorces et des clôtures sont aussi indiquées dans le tableau. Nous pouvons énoncer que les structures subdivisent, approximativement, en groupes de séries relativement réguliers : celles précédées d'une amorce, celles suivies d'une clôture, celles incluant les deux et celles qui ne se sont attachées ni à l'une ni à l'autre. La seule exception semblerait être la fréquence des séries enchaînées suivies par une clôture, moins fréquentes que les séries simples similaires. Pourtant, cette différence peut être liée au double rôle des clôtures dans les séries enchaînées ; elles peuvent simultanément fonctionner comme clôture de la première série inférieure et comme amorce de la deuxième. En définitive, la majorité des structures du corpus comprend une amorce, une clôture ou les deux. En effet, seulement 26% des séries simples et 30% des séries enchaînées ne sont ni précédées d'une amorce, ni suivies d'une clôture.

	Fréquence	
TOTAL DES SERIES LINEAIRES	748	
Séries simples	618	
avec amorce	142	23 %
avec clôture	186	30 %
avec amorce et clôture	127	21 %
sans amorce et clôture	163	26 %
Séries enchaînées	130	
avec amorce	41	31 %
avec clôture	17	13%
avec amorce et clôture	33	26%
sans amorce et clôture	39	30%

Tableau 2 Amorces et clôtures dans le corpus.

Bien que les séries linéaires repérées soient très variées, l'Exemple 19 ci-dessous en illustre un cas. En l'occurrence, la série est composée de deux items, introduits explicitement par les expressions anaphoriques *le premier* et *le second*. De plus, la structure est précédée d'une amorce à laquelle les deux items se réfèrent et qui définit également le nombre d'items de la série. L'article dont le texte est tiré aborde la construction du discours autour de la puissance paternelle dans le système judiciaire québécois aux XIX^{ème} et XX^{ème} siècles. Le style en gras du texte est de nous.

Exemple 19

... Bien que le Code favorisait le père lors d'une séparation de corps, à partir des années 1890 nous remarquons un changement assez important dans les propos des juges et leurs décisions. Ce changement de perception était en fait orienté vers l'enfant et les besoins nouveaux qu'on lui découvrait.

0 Ce changement de cap s'ouvrait sur **deux discours**.

1 **Le premier**, le plus acharné à contester le principe de l'autorité paternelle, était fondé sur l'idée que la femme était naturellement plus douée que l'homme pour élever des enfants.

2 **Le second** tenait pour acquis que, malgré leurs désirs ou leurs droits, les pères étaient physiquement incapables de combler les besoins affectifs de leurs enfants.

Fish

Les séries linéaires simples similaires à l'Exemple 19 précédent s'ajustent bien à notre définition : elles structurent le texte en items au moins partiellement ordonnés. Il en va de même pour les séries enchaînées. Ces structures ressemblent aux séries simples mais sont composées de plusieurs séries inférieures. Ainsi, ce qui dans une série simple constituerait une amorce peut, dans une série enchaînée, correspondre à une première série énumérant déjà ses items au lieu d'en indiquer seulement le nombre. Les items de la deuxième série sont généralement coréférentiels avec ceux de la première série.

Le corpus comprend 130 occurrences de séries enchaînées. Ci-dessous, l'Exemple 20 en est un exemple. Cette structure se compose de deux séries, le cas le plus fréquent. La première inclut une amorce annonçant le nombre de ses items, *deux types de compléments*, ainsi que deux items courts présentant ces deux compléments prédits. Les items ne sont pas explicitement introduits, à l'exception de la conjonction *et*, au début du second item. Outre la présentation de deux types de compléments annoncés par l'amorce, les items incluent également des références cataphoriques aux items de la série suivante qui sont, en fait, des exemples numérotés 10 et 11 dans le texte original. Les références, (*cf. 10*) et (*11*), fonctionnent donc comme une sorte d'amorces aux items à venir. La seconde série est à nouveau terminée par *ces deux types de pronominalisation*, une clôture exacte qui rappelle le nombre d'items de la structure.

Exemple 20

- 0 En outre, en linguistique française la notion de datif s'est imposée, depuis les études de Kayne (1977) et de Leclère (1976) et (1978), pour faire la distinction entre **deux types de compléments** indirects en à,
- 1 les datifs cliticisables en lui/leur (*cf. (10)*)
- 2 **et** les non-datifs, non-cliticisables en lui/leur, comme le complément de penser dans (*11*):
- 1 (10) a. Pierre donne un livre à Paul. b. Pierre lui donne un livre.
- 2 (11) a. Pierre pense à Marie. b. Pierre pense à elle. c. *Pierre lui pense.
- Cl. L'opposition entre **ces deux types de pronominalisation** et le fait que les clitiques lui/leur s'opposent aux clitiques nominatifs il/elle et accusatifs le/la/les permettent donc d'affirmer que le français dispose d'un marquage morphologique du datif au niveau des pronoms clitiques.

Van Peteghem

Lors de l'annotation, outre les séries simples et enchaînées typiques illustrées ci-dessus, nous avons identifié quelques types de série additionnels dont les caractéristiques les distinguent des cas typiques. Certaines de ces structures en diffèrent de par leur introduction tandis que d'autres structurent le texte d'une façon spécifique. Ces propriétés pouvant influencer le marquage et la longueur de ces séries, nous leur avons attribué des étiquettes afin de les identifier lors de l'analyse. Grâce aux étiquettes attribuées, nous pouvons prendre en compte ces spécificités automatiquement lors de l'analyse. Le Tableau 3 ci-dessous présente les fréquences de toutes les séries linéaires, simples et enchaînées, en incluant les séries avec une étiquette spécifique. Pour les séries simples, les étiquettes annotées sont *série argumentative*, *série accumulative* et *série section*. Pour les structures enchaînées, nous distinguons *série méta* et *série à deux temps*. De plus, les deux séries incluent des structures *temporelles*. Ces séries étiquetées représentent ensemble 139 occurrences. Les structures typiques forment ainsi la majorité, soit 609 occurrences et 81 % du corpus.

	Fréquence
TOTAL DES SERIES LINEAIRES	748
Séries simples	618
dont séries simples typiques	518
séries argumentatives	27
séries accumulatives	8
séries section	23
séries temporelles	42
Séries enchaînées	130
dont enchaînées proprement dites	91
séries méta	31
séries à deux temps	6
séries temporelles	2

Tableau 3 Séries simples, enchaînées et à étiquettes particulières.

Pour les séries simples, le premier type de série avec une étiquette est *la série argumentative*, déjà définie dans Jackiewicz (2005 : 108). Caractéristique de cette série, le marquage d'un des items, normalement du deuxième, par un marqueur adversatif et sa correspondance avec le modèle d'argumentation de Toulmin (1958, retravaillé dans Adam 1992). Ce modèle décrit l'argumentation en y distinguant les étapes de la thèse originale, les données, ou l'antithèse, et la conclusion, qui représente souvent une nouvelle thèse. Dans les séries argumentatives, les items correspondent à ces étapes, ce qui influence la façon dont ces séries structurent le texte. En tout, le corpus comprend 27 séries argumentatives de ce type. L'Exemple 21 suivant, tiré de la conclusion d'un article traitant de la notion de « formation discursive », illustre ces séries. Il commence par une amorce du type *question*. Le premier item, introduit par le MIL *dans un premier temps*, présente une réponse positive à cette question, c'est-à-dire la thèse originale selon le modèle de Toulmin. Le deuxième item, signalé par le connecteur adversatif *mais*, expose une opinion contraire, autrement dit une réponse négative à la question posée dans l'amorce. Cet item peut être interprété comme l'antithèse. La clôture de l'exemple, enfin, présente la conclusion, soit une nouvelle thèse inspirée des arguments présentés.

Exemple 21

- 0 Au sein d'un tel espace de co-partage, la notion de formation discursive peut-elle encore conserver une place ?
- 1 **Dans un premier temps**, nous avons eu tendance à considérer sa disparition sur le devant de la scène discursive comme définitive, tout en laissant ainsi une place vide sans cesse remplie par quelque chose qui existe et quelqu'un qui parle – en l'occurrence l'existence d'un sujet empirique – à l'horizon d'une situation sociale donnée.
- 2 **Mais**, à bien y réfléchir, la formation discursive peut désigner ce quelque chose en tant que genre discursif le plus élevé, dans la mesure où ce quelque chose s'avère être un sujet approprié pour la pensée et le discours, la réalité et le langage, donc s'inscrit dans un horizon donné, en l'occurrence la quête sociale de l'émancipation humaine.
- Cl. Plus simplement, la notion de formation discursive renverrait, dans une perspective nominaliste (Kaufmann & Guilhaumou, 2003), à la nécessaire médiation de l'ordre du discours entre la réalité et la pensée : un ordre du discours qui marque ainsi fortement sa présence au sein du lien entre la réalité et l'esprit.

Guilhaumou

La *série accumulative* est le deuxième type d'étiquette que nous avons attribuée aux séries simples. L'alternance de deux types d'items semble caractériser ces structures ; les items typiques aux séries linéaires énumèrent, par exemple, des idées ou des arguments et les items introduits par des marqueurs adversatifs présentent des contre arguments ou des idées opposées. La série peut également inclure des items explicatifs, justifiant le propos antérieur, ou des items concessifs, présentant une concession à l'énoncé d'un item précédent. La série résultante est souvent similaire à une chaîne de raisonnements avec arguments, contre arguments, explications et conclusion(s).

Les séries accumulatives sont plutôt rares : seulement huit occurrences dans le corpus. Étant relativement variées, il n'est pas possible de les définir aussi exhaustivement que les séries argumentatives. Cela ne serait pas non plus nécessaire pour cette étude étant donné leur fréquence très faible dans notre corpus. Toutefois, certains traits importants de ces séries méritent une remarque, le plus notable étant peut-être la présence d'au moins un item introduit par un marqueur adversatif ou concessif, tel que *mais*, *cependant* ou *pourtant*. La longueur et la présence des items non marqués constituent d'autres caractéristiques. Contrairement aux séries argumentatives, composées de deux ou trois items, ces séries comprennent toujours au minimum quatre items et semblent également inclure au moins un item non marqué. L'Exemple 22 ci-après, composé de six items et d'une amorce, illustre ces séries accumulatives.

Exemple 22

- 0 Les élèves en difficulté ont besoin d'un encadrement plus important que les autres élèves qui participent à ces programmes.
- 1 Les rencontres hebdomadaires peuvent bien combler les besoins de ces jeunes.
- 2 Plusieurs rencontres individuelles devront **cependant** avoir lieu tout au long de l'année.
- 3 Ces élèves présentent souvent des problèmes de motivation et sont portés à démissionner au premier obstacle. De ce fait, un système de récompenses peut stimuler la motivation, le temps de permettre à l'élève de retirer un autre type de gratification de son expérience de médiateur.
- 4 **Cependant**, si l'élève n'obtient pas de satisfaction à pratiquer la médiation, les récompenses n'arriveront pas à le motiver à venir en aide à ses pairs. Dans cette étude, les élèves ont placé les récompenses (argent scolaire) très loin derrière les gratifications sociales engendrées par la pratique de la médiation. Ces sentiments de valorisation semblent en effet fournir la motivation nécessaire aux élèves pour persévérer dans leur tâche.
- 5 **De plus**, il importe de susciter l'intérêt des parents afin qu'ils valorisent et encouragent eux aussi la participation de leur enfant au programme de médiation.
- 6 **Finalement**, la promotion continue du programme dans l'école permet aux autres élèves de ne pas oublier l'existence des médiateurs et les incitent à s'adresser plus souvent à leurs services. Sans cet élément promotionnel, les médiateurs pratiquent très peu de médiations et se désintéressent rapidement de leur rôle.

Beaumont

Dans l'Exemple 22, l'amorce est implicite. Au lieu d'indiquer le nombre d'items de la série, elle définit simplement la thématique des composants énumérés : l'aide dont les élèves en difficulté ont besoin. Le premier et le deuxième items présentent des moyens d'aide possibles, soit des rencontres avec les élèves, le deuxième item précisant le premier. Le troisième item expose un autre aspect de ces moyens, un système de récompenses. Pourtant, au lieu de l'indiquer immédiatement au début, cet item commence par une description du comportement des élèves, le système de récompenses apparaissant uniquement dans la deuxième phrase de l'item. Bien que cette description justifie le moyen d'aide présenté, elle rend également l'organisation interne de l'item différente des deux premiers et risque de compliquer la lecture de l'ensemble de la série. Le quatrième item fournit encore un contre argument à celui-ci en définissant des conditions dans lesquelles le système de récompenses ne fonctionne pas. Les items 5 et 6, enfin, énumèrent également d'autres moyens d'aide possibles.

La lecture de l'Exemple 22 prouve que les séries accumulatives ont plusieurs traits que nous avons, en sections 4.2.1 et 4.2.4, associés aux séries non parallèles : marqueurs différents, items non marqués et de natures variées. Les natures différentes des items, notamment, rendent la série plus difficile à interpréter ; les items du début et de la fin énumèrent des moyens d'aide possibles, mais ceux du milieu présentent des contre arguments voire incluent des parties descriptives. De plus, même les moyens d'aide énumérés ne sont pas tout à fait équivalents. Le deuxième item, par exemple, s'affirme plutôt comme une précision ou une élaboration du premier. Par conséquent, il est loisible de se demander si cette série est facilement lisible, claire et logique comme segment textuel ou si elle est trop complexe pour clairement argumenter et convaincre le public. Quelle que soit la réponse, le corpus inclut huit structures similaires, prouvant ainsi l'utilisation de cette façon d'organiser du texte dans le discours de recherche.

En définitive, les séries argumentatives et accumulatives diffèrent des séries typiques selon l'inclusion des contre arguments ou idées opposées grâce auxquels la série résultante adopte un fonctionnement particulier. Dans cette perspective, la troisième étiquette attribuée aux séries simples diffère des deux premières en ce sens qu'elle concerne uniquement la façon dont la série est signalée. Dans ces *séries section*, les items correspondent à des sections titrées des articles de recherche examinés. Ainsi, outre les marqueurs d'item proprement dits, les items sont introduits par les titres de section contribuant également au marquage de la série. À cause de ces titres et de la longueur de ces séries, leur distinction des autres séries est importante. Le corpus comprend 23 séries de ce type.

Enfin, la dernière étiquette spécifique attribuée aux séries simples a la particularité de s'appliquer aussi aux séries enchaînées. Au lieu d'organiser le texte en soi, ces *séries temporelles* se réfèrent au monde extérieur et aux événements ayant eu lieu dans le monde décrit par le texte. En section 2.2.2, discutant du métadiscours, nous avons souligné la différence entre les expressions internes, se référant au texte, et les expressions externes,

renvoyant au monde réel, en notant que les séries linéaires peuvent appartenir aux deux groupes et que la distinction entre ces deux groupes risque d'être difficile à établir. Nonobstant, nous avons décidé de rendre cette distinction explicite puisqu'elle révèle, entre autres, des caractéristiques des trois sous-corpus examinés (cf. section 8). Ainsi, les séries linéaires faisant partie du métadiscours et remettant uniquement à la structure du texte sont définies comme le défaut, sans étiquette particulière. L'étiquette *temporelle*, en revanche, est attribuée aux séries linéaires décrivant des événements ayant eu lieu dans le monde réel. L'étiquette temporelle est adéquate car, en pratique et eu égard au critère d'ordre entre les items, ces séries sont toujours temporelles et décrivent des événements s'étant déroulés l'un après l'autre. Dans ces séries, la structure du texte revêt également un rôle important, les marqueurs d'item se référant simultanément au texte même et aux événements extra textuels. Les structures dans lesquelles l'organisation du texte est dépourvue de rôle et les marqueurs d'item ne présentent aucun lien avec le texte, tels que *la première guerre mondiale... la deuxième guerre mondiale...*, sont exclues de l'analyse.

Ci-dessous, l'Exemple 23 illustre les séries temporelles comptant en tout 44 occurrences dans le corpus : 42 séries simples et 2 séries enchaînées. L'Exemple 23, une série enchaînée, décrit les entrevues réalisées dans une étude sur les effets de l'entraînement à l'improvisation pour la gestion des imprévus par les enseignants. Les items des deux séries inférieures se réfèrent à deux entrevues menées, la deuxième une semaine après la première. Par conséquent, l'ordre entre les items et les entrevues est de nature temporelle et la série étiquetée comme *temporelle*. La série commence par une amorce désignant le nombre d'items des séries : *deux entrevues*. La première série présente ces entrevues brièvement et en indique les durées. Les items sont introduits par les anaphoriques *une première* et *une seconde* qui, outre la référence à l'amorce, trouvent leur source dans le monde réel et dans l'ordre temporel de déroulement des entrevues. Les items de la seconde série, introduits par des *constructions numéro+nom*, offrent un complément d'information sur ces deux événements et sont coréférentiels avec les items de la première série.

Exemple 23

- 0 **Deux entrevues** semi-dirigées ont été réalisées avec chaque participant à la recherche :
- 1 **une première** d'une durée de 60 minutes
 - 2 **et une deuxième**, de 30 minutes.
- 1 **La première entrevue** a permis d'explorer des situations de gestion d'imprévus à partir d'une description de ceux-ci et des actions prises, puis une explication des liens possibles entre leur entraînement à l'improvisation et ces actions.
- 2 **La deuxième entrevue** s'est déroulée une semaine plus tard et elle a servi à valider l'analyse faite après la première entrevue.
- Cl. **Ce moyen de triangulation** a permis d'assurer ce que Savoie-Zajc (1996, p. 9) nomme l'acceptation interne qui désigne « le degré de concordance et d'assentiment qui s'établit entre le sens que le chercheur attribue aux données recueillies et sa plausibilité telle que perçue par les participants à l'étude.»

Jonnaert

Il nous reste à présenter les étiquettes attribuées aux séries enchaînées, *séries méta* et *séries à deux temps*. Les séries *méta* combinent les séries enchaînées et les séries section ; elles sont composées d'une série *métadiscursive* qui présente d'une façon cataphorique les items de la seconde série, elle-même une série section dont les items correspondent aux sections titrées. Nous pourrions donc considérer l'organisation pourvue par ces structures comme très typique des articles de recherche.

Les séries méta affichent 31 occurrences dans le corpus. Le deuxième composant de ces structures couvrant automatiquement plusieurs sections, elles sont, par défaut, relativement longues et englobantes. Les séries longues étant en général moins nombreuses que les séries plutôt courtes (voir section suivante), nous pouvons donc les considérer comme relativement fréquentes. De plus, de par leur longueur, ces structures jouent inévitablement un rôle important dans l'organisation globale de l'article quand elles sont utilisées. Il est également intéressant de noter le nombre plus élevé de séries méta que de séries section. Il révèle que l'organisation des sections est plus fréquemment annoncée explicitement de façon cataphorique avant les sections par une autre série qu'uniquement dans les sections par des marqueurs individuels. L'Exemple 24 ci-dessous illustre une série méta, composée d'une première série cataphorique ou métadiscursive et d'une seconde série couvrant plusieurs sections.

Exemple 24 Les marqueurs explicites sont en gras et les indices lexicaux répétés en italiques.

... Or, le conseil municipal n'est pas l'unique lieu d'intervention des élites locales. Différents groupes d'intérêts, particulièrement la Chambre de commerce, la Ligue des propriétaires et l'Association des manufacturiers canadiens, participent aussi à la mise en place des nouveaux mécanismes de l'assistance publique en collaborant de diverses façons avec le conseil de ville.

- 1.1 Voyons **d'abord** comment les municipalités héritent de pouvoirs aussi importants en matière d'assistance *durant les années 1930*.
- 1.2 Nous soulignerons **ensuite** l'intérêt particulier que revêt *Drummondville comme terrain d'enquête*.
- 1.3 **Puis**, pour mieux comprendre les liens que tissent *le conseil municipal* et le monde associatif *dans le dossier de l'aide aux chômeurs*, nous mettrons en relief les responsabilités qu'assume *la municipalité de Drummondville* dans ce domaine, ainsi que la composition des organisations impliquées dans la mise en œuvre des mesures d'assistance et ce qui favorise ou entrave leur collaboration.

2.1 LA MISE EN PLACE DES POLITIQUES SOCIALES DURANT L'ENTRE-DEUX-GUERRES

Avant l'adoption de la Loi de l'assistance publique (1921), il n'existait pratiquement pas de mesures destinées aux pauvres gérées par l'État central. Les mesures étaient déployées localement et relevaient du domaine de la charité privée. En 1910, à Montréal, un projet de municipalisation de l'assistance aux indigents est étudié par une Commission spéciale. L'un des membres, Albert Chevalier, propose alors de créer un bureau de placement municipal pour les chômeurs désirant travailler et un refuge municipal pour protéger les individus préférant « vivre d'expédients ». [...]

2.2 DRUMMONDVILLE COMME TERRAIN D'ENQUÊTE

L'intérêt de la ville de Drummondville pour cette recherche est multiple. À la différence de Montréal, Québec, Sherbrooke et Verdun, sur lesquelles ont porté les études de cas sur la Dépression et les politiques sociales, Drummondville ne dépasse le cap des 10 000 habitants qu'au début des années 1940. Le choix d'une petite localité permet d'avoir une meilleure prise sur les réalités locales et de cerner les caractéristiques d'une plus forte proportion des acteurs municipaux et des bénéficiaires des mesures d'assistance. [...]

2.3 LE GOUVERNEMENT LOCAL ET LES MESURES D'AIDE AUX CHÔMEURS

Face au chômage de masse que produit la crise des années 1930, des sociétés charitables comme la Société de Saint-Vincent-de-Paul (SSVP), conçues essentiellement pour venir en aide aux personnes inaptes au travail ou pour le chômage de courte durée, sont insuffisantes. Au début de la crise, cet organisme de bienfaisance est néanmoins perçu par les autorités municipales comme le mieux à même de distribuer les secours aux personnes sans emploi. [...]

Roux-Pratte

Dans l'Exemple 24, les deux séries formant la série méta sont segmentées en trois items. La première série présente les thèmes traités dans les trois sections à venir, chaque item correspondant à un item-section de la seconde série. Les items de la première série sont introduits d'une façon prototypique du métadiscours : *d'abord, ensuite, puis*. Les items de la seconde série sont, en revanche, non marqués. Des instances de répétition lexicale rattachent cependant ces items à leurs corrélats explicitement signalés. *Drummondville comme terrain d'enquête* du deuxième item de la première série est répété dans le titre du deuxième item de la série section et *aide aux chômeurs* est également répété dans les derniers items des deux séries. L'expression *durant les années trente* du premier item de la première série devient pourtant *durant l'entre-deux-guerres*. À notre avis, ce changement ne perturbe pourtant pas la lecture puisque, dans ce contexte, elles peuvent facilement être interprétées comme quasi-synonymes grâce au savoir extra linguistique. Il est à noter également que, dans l'article dont l'exemple est extrait, les sections ne sont pas numérotées. Cette habitude semble relativement fréquente dans les articles de recherche en histoire et reflète sans doute l'organisation des textes dans ce domaine.

Contrairement aux séries méta, les séries à deux temps sont plutôt rares avec seulement six occurrences dans le corpus. Nous avons déjà brièvement présenté ces structures en section 4.2.1, en discutant de travaux antérieurs sur les séries linéaires. Comme toutes les séries enchaînées, ces structures se composent de plusieurs séries combinées. Pourtant, dans celles à deux temps, les composants ne sont pas égaux. En effet, les items de la deuxième série dépendent thématiquement de ceux de la première dans laquelle ils sont introduits pour la première fois. Par conséquent, la deuxième série ne pourrait pas être utilisée seule, de façon autonome. Comme les définit Porhiel (2007 : 119) dans son article sur ces séries :

Dans le premier temps, les items sont nouveaux et forment avec l'amorce une structure textuelle indépendante et référentiellement autonome. Les items du second temps sont, quant à eux, connus et dépendent référentiellement des items introduits dans le premier temps.

Porhiel (2007 : 119)

Ce déséquilibre entre les deux séries inférieures transparaît également dans leur marquage : la première série est introduite d'une façon typique pour les séries linéaires, tandis que les items de la deuxième série sont signalés par des marqueurs thématiques comme *quant à* et *pour ce qui est de*. L'Exemple 25 ci-dessous illustre ce marquage. Il est extrait d'un article du domaine de l'éducation, traitant de l'usage de l'entraînement à l'improvisation dans la gestion de classe par des enseignants. La première série présente brièvement les items auxquels ceux de la seconde série se réfèrent. De plus, la première série est précédée d'une amorce indiquant le nombre d'items de la série à venir : *deux principales composantes de l'entraînement à l'improvisation*.

Exemple 25

- 0 Les **deux principales** composantes de l'entraînement à l'improvisation identifiées sont
- 1 la rapidité d'exécution
- 2 **et** la vigilance.
- 1 **En ce qui a trait** à la composante « vigilance », les résultats vont dans le sens des travaux de Lobman (2005) qui identifie clairement la vigilance, plus précisément l'écoute, comme une composante développée lors d'un entraînement à l'improvisation théâtrale chez des enseignantes et des enseignants du préscolaire et qui favorise l'écoute en classe.
- 2 Non seulement la vigilance apparaît-elle comme une composante importante de l'entraînement à l'improvisation, mais elle est intimement liée à la rapidité d'exécution, qui représente l'autre composante active dans la gestion des imprévus.
- Cl. L'entraînement à l'improvisation aiderait **donc** à rendre plus efficace le processus de gestion des imprévus en classe en stimulant davantage la vigilance, en augmentant la rapidité du processus de réflexion-en-action et en permettant de développer de nouvelles stratégies ou actions à prendre pour gérer certains imprévus.

Pelletier

La première série de l'exemple est très courte : les items sont tous énumérés en une seule phrase. Comme la plupart des composants des séries enchaînées, cette série ne serait pas analysée comme une série simple car elle ne couvre qu'une seule phrase. Le premier item de la série est non marqué, tandis que le second est introduit par la conjonction de coordination *et*. Les items de la seconde série se réfèrent à ces deux items bien que, d'une façon intéressante, dans l'ordre inverse : l'item 1 de la première série correspond au deuxième item de la seconde série et *vice versa*. De plus, la deuxième série est plus longue que la première et ses items développent la thématique présentée auparavant. De surcroît, seul le premier item est introduit par un marqueur thématique : *en ce qui a trait à*. Le second item, se référant au premier item de la première série, s'avère plus difficile à identifier et surtout à délimiter. Selon le modèle structurel des séries à deux temps, cet item devrait traiter de la rapidité d'exécution, comme annoncé dans le premier composant. Ce

thème est bien discuté dans la phrase suivant le premier item, mais seulement à la fin. Cette organisation interne combinée à l'ordre inverse des items et au non marquage rendent ce dernier item de la structure moins typique.

Pour conclure, nous avons exposé dans cette section une variété de séries linéaires différentes. Dans l'ensemble, les séries linéaires se présentent comme des structures variées qui, outre leur fonction typique de segmentation du texte en items se suivant et, au moins partiellement, ordonnées, peuvent aussi adopter d'autres fonctionnements plus spécifiques. Certaines, telles que les séries section et méta, diffèrent des cas typiques par leur marquage et d'autres, telles que les séries accumulatives, par leur fonctionnement. Pour une partie des séries, telles que les structures accumulatives et argumentatives, nous pouvons même questionner leur statut de série linéaire : peut-on dire qu'elles structurent le texte en items qui se suivent dans le texte ? La réponse dépend surtout de la précision appliquée dans la définition. Nous considérons qu'il est plus avantageux d'inclure dans l'analyse une variété de structures différentes afin de mieux en décrire formes et fonctionnements.

6.2 LONGUEURS

La longueur d'une série, soit le nombre de mots qu'elle couvre, est naturellement une question importante de l'étude du fonctionnement de ces structures. De fait, une série de 50 mots joue un rôle très différent dans l'organisation textuelle en regard d'une série de 5 000 mots. D'ailleurs, la longueur d'une série ne signifie pas que les séries longues soient supérieures aux séries courtes mais plutôt que leurs rôles et fonctions diffèrent. La fréquence des séries courtes ou des séries longues peut indiquer un mode d'organisation typique du texte examiné. Ainsi, les séries longues, telles que les séries sections et méta, sont effectivement utilisées dans des articles de recherche alors qu'elles sont difficilement envisageables dans un texte narratif, par exemple.

La longueur minimale des séries simples annotées est donc de deux phrases. Comme nous l'avons mentionné dans la section précédente, nous n'avons pas défini de longueur maximale. La longueur des articles de recherche examinés impose pourtant ses limites : les articles sont plus longs dans la revue *Marges Linguistiques* avec une moyenne de 13 963 mots et plus courts dans la revue *Spirale*, en moyenne 4 831 mots. Naturellement, les séries organisant le texte ne peuvent pas excéder la longueur de l'article. Il est pourtant possible que la série englobe presque tout l'article.

Outre le nombre de mots, la longueur d'une série linéaire peut être mesurée par le nombre d'items qu'elle inclut ainsi que par le nombre de séries combinées dans le cas d'une série enchaînée. La présence d'une amorce ou d'une clôture dans la structure influence aussi la longueur finale. En théorie, le nombre d'items dans une série ou le nombre de séries dans une structure enchaînée n'est pas limité. La pratique, pourtant, impose évidemment des limites.

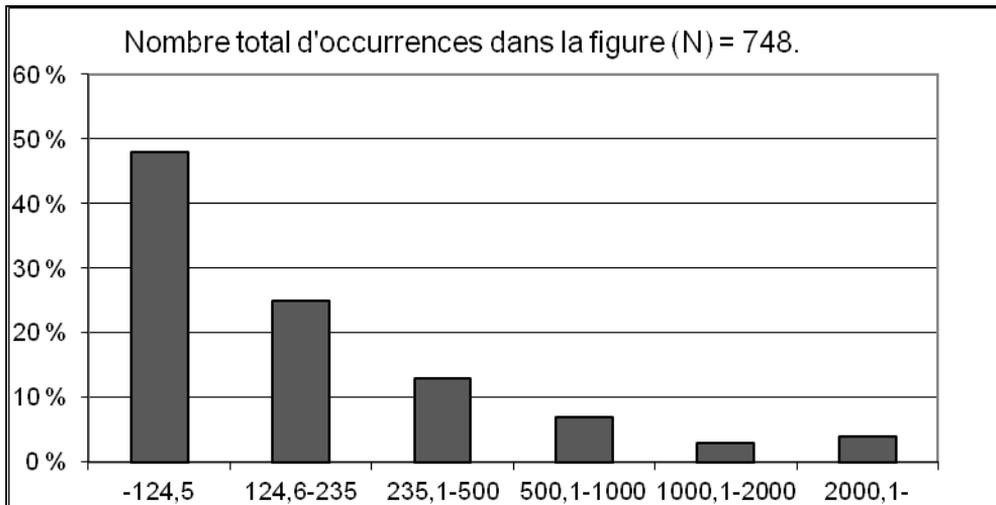


Figure 6 Longueurs de toutes les séries linéaires du corpus.

La Figure 6 ci-dessus présente brièvement les longueurs des séries linéaires dans le corpus. Comme nous pouvons le constater, la majorité (près de la moitié) des structures sont très courtes et couvrent moins de 124,5 mots. Les séries de 124,6 à 235 et de 235,1 à 500 mots sont encore relativement fréquentes tandis que celles dépassant 500 mots se limitent à 14 % du total des séries. Ce pourcentage ne doit pourtant pas non plus être sous-estimé. En effet, il correspond à 111 séries linéaires, dont 34 (4 %) dépassent même 2000,1 mots, la plus longue atteignant 12 860 mots. Ainsi, bien que les structures soient majoritairement très courtes, le corpus inclut aussi des structures plus longues, voire extrêmement longues.

À la suite, nous présentons un exemple des séries plutôt courtes, donc des plus fréquentes. L'exemple est tiré d'un article sur la pédagogie dans un environnement numérique. Cette série couvre 50 mots et est composée de deux items, le premier formant une et le second deux phrases. Les items présentent respectivement un aspect positif et un aspect négatif du progrès de l'enseignement dans le domaine discuté. Afin de bien distinguer les arguments et de souligner le contraste entre eux, les items sont introduits par les MIL *d'une part* et *d'autre part*. Comme il n'y a cependant que deux arguments qui ne sont, en outre, traités que brièvement, la série reste plutôt courte.

Exemple 26

- 1 **D'une part**, il est bon que les établissements d'enseignement supérieur rattrapent leur retard sur les plus grandes firmes de ce monde.
- 2 **D'autre part**, ce nouvel environnement numérique ne va pas nécessairement ou magiquement élever notre niveau d'intelligence ou d'efficacité collective. Cela va dépendre de ce que nous en ferons.

Demailly

La longueur de la majorité des séries linéaires simples du corpus est donc inférieure à 124,5 mots. Or, comme nous l'avons montré dans la section précédente, les séries repérées sont relativement variées et de natures différentes. Par conséquent, il importe d'étudier les longueurs également en fonction du type de la série repérée. Le Tableau 4 suivant décrit les longueurs et les nombres d'items des séries examinées pour chaque type de série. Nous notons que les longueurs sont décrites par deux chiffres, la médiane et la moyenne. La moyenne sert souvent à illustrer un ensemble de valeurs mais elle est très facilement influencée par les valeurs extrêmes du corpus. Dans notre cas, ce sont surtout les séries très longues qui, bien que peu fréquentes, influencent ce chiffre. La médiane, en revanche, indique la valeur centrale des occurrences en divisant le corpus en deux parties dont chacune contient le même nombre de valeurs. Elle constitue, par conséquent, un élément descriptif plus stable. Cette différence entre les deux chiffres est facilement repérable dans le tableau ci-dessous, dont les moyennes s'avèrent souvent beaucoup plus élevées que les médianes. Cet écart entre les deux valeurs indique la présence d'occurrences extrêmes, c'est-à-dire de séries très longues bien que minoritaires dans le corpus.

	Médiane	Moyenne	Nombre d'items	Nombre de séries
TOTAL DES SERIES LINEAIRES	130	433	3	-
Séries simples	125	271	2,6	-
séries simples typiques	119	196	2,6	-
séries argumentatives	127	181	2,2	-
séries accumulatives	280	376	4,8	-
séries section	1 458	2187	3,1	-
séries temporelles	109	188	2,9	-
Séries enchaînées	201	1201	5	
dont enchaînées proprement dites	118	213	4,6	2,1
séries méta	4 004	4250	5,8	2,0
séries à deux temps	340	770	7,7	2,3
séries temporelles	592	592	4	2

Tableau 4 Longueurs des séries en nombres de mots, d'items et de séries inférieures.

La longueur moyenne de l'ensemble des séries du corpus est donc de 433 mots et la valeur médiane de 130 mots. Ces chiffres sont la démonstration de notre explication de la Figure 6 ; la plupart des séries sont très courtes bien que le corpus inclue aussi des structures très longues. La médiane de la longueur des séries simples typiques, sans étiquettes, est 125 mots et leur moyenne de 271 mots. Les séries les plus fréquentes sont donc aussi, le plus souvent, plutôt courtes.

Sur un plan global, les séries simples sont évidemment plus courtes que les séries enchaînées. Ces chiffres sont pourtant fortement influencés par les séries section et méta. En particulier, les séries méta sont très étendues, étant donné qu'avec la présente mise en page leur longueur moyenne (4 250 mots) couvre huit pages environ. En revanche, si les séries section et les séries méta sont exclues de l'analyse, nous notons que les séries enchaînées proprement dites présentent environ la même longueur que les séries simples

typiques. Les séries à deux temps sont de longueur intermédiaire avec une médiane de 340 mots.

Quant aux autres types de séries simples, leur longueur ne diffère pas de manière importante de celle des séries simples typiques. Nous pouvons cependant constater que les moyennes des séries argumentatives et temporelles sont plutôt basses, indiquant que les valeurs extrêmes, soit les séries longues, sont plus rares dans ces types de série. Seule exception notable dans les longueurs des séries simples, les séries accumulatives sont un peu plus vastes, avec une médiane de 280 et une moyenne de 376 mots. Cette longueur est due au grand nombre d'items, caractéristique de ces séries. Comme le montre le Tableau 4, ces séries se composent de 4,6 items en moyenne, tandis que les séries simples en général incluent 2,6 items.

Le nombre d'items d'une série peut donc affecter la longueur, ainsi pour les séries accumulatives, ou ne pas l'influencer, comme pour les séries enchaînées. Comme nous venons de le mentionner, les séries simples incluent environ 2,6 items par série, les séries accumulatives étant exceptionnelles dans ce groupe. Le nombre maximum d'items pour les séries simples dans notre corpus s'élève à 8. Pour les structures enchaînées, ce nombre est évidemment plus important avec une moyenne de cinq items et une valeur maximale de 18 items. Dans ce groupe, la moyenne la plus élevée (7,7) est pourvue par les séries à deux temps.

Le nombre de séries composant les séries enchaînées constitue un dernier facteur d'influence de la longueur finale de la structure. Comme le Tableau 4 l'illustre, la majorité des structures enchaînées est composée de deux séries linéaires environ. En tout, seules sept structures incluent trois et deux structures comprennent quatre séries. Du fait de leur rareté, ces séries linéaires ont peu d'effets sur l'ensemble du corpus. Pourtant, elles offrent de nouveaux points de vue sur les usages et combinaisons possibles des séries enchaînées. L'Exemple 27 présente l'illustration d'une série enchaînée composée de trois séries inférieures. Tiré d'un article d'histoire, le texte discute des usages et façons dont le mot *quartier* est compris au XIX^{ème} siècle. Comme c'est souvent le cas pour les séries enchaînées composée de plus de deux séries, l'exemple est relativement court et ressemble plutôt à une chaîne de référence avec deux sources. Par ailleurs, les marqueurs d'item sont des expressions anaphoriques.

Exemple 27

- 0 L'usage du « quartier » dans les dossiers de Juin 1848 renvoie donc à **une double capacité** :
 - 1 pouvoir désigner sans pour autant identifier ;
 - 2 pouvoir définir qualitativement sans pour autant connaître.
- 1 **La première** résulte de l'interconnaissance
- 2 **la seconde** de la circulation des informations
- 1 L'une
- 2 **et l'autre** sont évidemment liées, comme le montre la déposition du prévenu Cocqueret : « Le propos qui m'a engagé à l'acte que vous me reprochez a été tenu en présence de M. Bornier, propriétaire, 81 rue du Faubourg-Saint-Denis, je ne sais pas le nom de l'homme qui le tenait ».

Clavier

La série linéaire enchaînée de l'Exemple 27 commence par une série présentant les futures zones de référence et est également précédée d'une amorce, annonçant le début de la série : *une double capacité*. Les items indiquant les deux sens du mot sont non marqués, séparés par un point-virgule. Les items des deuxième et troisième séries, introduites par *la première*, *la seconde* et par *l'une*, et *l'autre*, se réfèrent respectivement à ces deux items de la première série. En effet, dans la troisième série, les items sont pratiquement combinés car seul le deuxième item comporte du contenu autre que la référence elle-même. Ainsi, cette série ne serait pas analysée individuellement et nous pourrions même nous demander s'il ne s'agit pas plutôt d'une clôture que d'une série en soi. Nous l'avons pourtant analysée comme une série courte faisant partie d'une structure supérieure parce qu'elle inclut deux références individuelles renvoyant, au moins en principe, à des sources différentes au lieu de se référer aux items de la série comme un tout.

Pour terminer la discussion sur les longueurs des séries, nous examinerons encore une caractéristique des séries longues qui mérite une remarque. Comme nous l'avons démontré, la longueur de 111 séries, soit 14 % du total, excède 500 mots. En fait, sur ces séries, 23 sont des séries section et 30 des séries méta, ce qui montre que les séries de cette longueur sont en général segmentées en items titrés. Cela est évident, vu que dans les articles de recherche, la longueur des sections et des sous-sections n'est pas nécessairement supérieure à ce chiffre. Les titres participent donc forcément au marquage des structures de cette longueur.

Le corpus inclut pourtant aussi 58 séries linéaires de plus de 500 mots qui ne sont pas étiquetées méta ou section. Or, comme seule une de ces séries dépasse 2 000 mots, nous pouvons avancer que les séries très longues ne couvrant pas de sections entières, sont extrêmement rares. Outre la segmentation du texte en sections, nous sommes en droit de nous demander si ce résultat reflète une limite de portée des connecteurs et d'autres indices textuels employés comme marqueurs.

À la suite, nous illustrons les séries relativement longues ne couvrant pas des sections entières par l'Exemple 28 qui comprend en tout 1 072 mots ainsi que trois items et une

clôture qui ne correspondent pas aux sections titrées. Cet exemple provient de l'introduction d'un article en éducation sur la violence en milieu scolaire. Le texte organisé en série linéaire présente les objectifs du travail et les justifie. La présentation se termine par la clôture qui résume ces objectifs en les reformulant en questions de recherche.

Exemple 28

- 1 En plus d'être rares, les études systématiques disponibles sur l'ampleur réelle des phénomènes de violence présentent plusieurs limites méthodologiques importantes (Charlot et Emin, 1997 ; Debarbieux et Blaya, 2001). **Premièrement**, [...]. **L'objectif de la présente étude est de combler en partie ces lacunes** et de dresser un tableau le plus fidèle possible des phénomènes de violence dans l'enseignement secondaire de la Communauté française de Belgique. [...]
 - 2 **Au-delà de la nature et de l'ampleur des phénomènes de violence, une autre question importante** consiste à identifier à quels facteurs sont associés ces phénomènes. Mieux connaître les facteurs reliés positivement ou négativement à l'ampleur des victimations est une étape fondamentale dans l'élaboration d'actions de prévention ou d'interventions, en ce qui concerne la violence à l'école. [...]
 - 3 **Enfin, une autre question** qui se pose est celle de l'impact de différentes formes de violence. Toutes les formes de violence à l'école ont-elles le même retentissement dans la vie des personnes ? [...]
- Cl. **En résumé**, la présente enquête vise à répondre aux questions suivantes, tant du point de vue des élèves que du point de vue des enseignants : (1) [...].

Galant

Selon nos critères d'analyse, le premier item est non marqué puisqu'il n'est signalé par aucun indice d'ordre ou d'addition. L'item commence par noter le manque d'études systématiques sur l'ampleur des phénomènes de violence et les limites méthodologiques des travaux dans ce domaine. Cet item inclut également une série enchâssée, énumérant des limites rencontrées et commençant par *premièrement*. Après la série enchâssée, l'item se termine par la présentation explicite du premier objectif, simultanément une conséquence des insuffisances exposées : *l'objectif de l'étude est de combler ces lacunes...* En tout, l'item couvre 481 mots.

Le deuxième item commence par se référer au premier item et par répéter l'essentiel pour l'énumération : *au de-là de la nature et de l'ampleur des phénomènes de violence...* Ensuite, l'item continue en présentant le deuxième objectif du travail : *une autre question [...] à identifier...* En fait, ce n'est qu'à partir de la mention *autre question* que la série en soi est formée : dans le premier item, rien n'indique la présence d'une série. En l'occurrence, c'est notamment le marqueur additif *autre* qui la révèle. Finalement, le troisième item reprend le nom *question* du deuxième item pour introduire encore un nouvel objectif : *enfin une autre question...* Le statut de cet item comme étant le dernier est encore renforcé par le connecteur *enfin*.

Vu la longueur de la série, l'organisation de l'Exemple 28 en série linéaire n'est pas très explicite bien que seul son premier item soit non marqué. La série inclut également plusieurs traits des structures non parallèles, discutés en sections 4.2.1 et 4.2.4. Tout

d'abord, durant toute la lecture du premier item de 481 mots, soit presque une page, le lecteur ne sait pas comment le texte à suivre est organisé. De plus, le nom *question* fonctionnant comme principe d'énumération est répété uniquement dans les deuxième et troisième items, rompant le parallélisme entre le premier composant et les suivants. La longueur du premier item souligne encore le non parallélisme : comprenant 481 mots, il est bien plus long que les deux autres items s'élevant, ensemble, à 490 mots. De même, la structure thématique du premier item diffère de celle des deux autres : au lieu d'indiquer directement l'un des objectifs du travail comme le font les autres, il commence par présenter des lacunes des travaux antérieurs puis, seulement, se termine en exposant le premier but. La description de ces lacunes forme toutefois partie intégrante du premier item, le deuxième item y faisant référence.. Pour cette raison, elle ne peut pas être exclue de l'item.

6.3 MARQUAGES

Le marquage des séries linéaires et de leurs items est l'un des intérêts principaux de ce travail. Le marquage inclut plusieurs aspects. D'une part, il s'agit de l'usage des marqueurs différents, tels que les connecteurs et les expressions anaphoriques, pour introduire les items des séries linéaires et marquer l'organisation textuelle. D'autre part, le marquage se réfère à l'introduction explicite ou implicite des items des séries linéaires, notamment à l'usage ou non des marques de cohésion. De plus, le marquage concerne les types d'amorce et de clôture utilisés avant ou après la série proprement dite, également susceptibles de participer à l'introduction explicite de ces structures. Par exemple, une amorce définissant le nombre d'items de la série peut influencer la façon dont ceux-ci sont signalés. Enfin, les marqueurs différents et le non marquage sont également centraux dans l'étude des séries linéaires en tant que structures organisant typiquement le texte en items parallèles. En effet, ces deux propriétés sont souvent associées au non parallélisme des items.

Pour examiner le marquage et le non marquage des items ainsi que les marqueurs utilisés, nous avons classifié les séries linéaires du corpus selon la présence d'items non marqués et de marqueurs différents dans une seule série. Dans les structures enchaînées, les catégories sont créées à partir du marquage des séries simples formant les structures supérieures. À la suite, nous présenterons ces catégories et leurs fréquences, d'abord pour les séries simples et ensuite pour les enchaînées. En ce faisant, nous explorerons également les types de marqueurs utilisés dans ces séries. Enfin, dans un troisième temps, nous aborderons les types d'amorces et de clôtures repérés.

6.3.1 SERIES SIMPLES

Dans cette section, nous commencerons par les catégories de série analysées et les types de marqueurs y utilisés. Ensuite, à la fin, nous réfléchirons également sur l'influence que le non marquage et l'usage des marqueurs différents peuvent avoir sur le parallélisme et le non parallélisme de la série.

1. Séries homogènes (HOMO). Dans ces séries, tous les items sont introduits explicitement par le même type de marqueurs. Les types de marqueurs distingués dans la classification sont les connecteurs, les expressions anaphoriques et les constructions numéro+nom (cf. 5.2.2, dorénavant Num+N), les connecteurs incluant également les expressions cadratives. Vu la similarité de ces deux groupes de marqueurs (cf. la définition des organisateurs textuels par Schneuwly *et al.* (1989) en 3.4.1 et la discussion en 4.3.1), nous commençons par les analyser ensemble afin de simplifier l'analyse. Dans les sections précédentes, certains des exemples présentés illustrent déjà la catégorie de séries linéaires homogènes ; l'Exemple 19 à la page 111 est composé de deux items introduits par des expressions anaphoriques, et l'Exemple 26 à la page 121 par les MIL *d'une part* et *d'autre part*.

2. Séries mixtes (MIX). Tous les items de ces séries sont introduits explicitement mais, contrairement aux séries homogènes, les marqueurs utilisés ne sont pas du même type. Ainsi, les expressions anaphoriques, les connecteurs et les constructions Num+N peuvent être utilisés pour introduire les items d'une seule série.

L'Exemple 29 ci-dessous présente une série MIX. De même que l'Exemple 19 à la page 111, cette série est extraite d'un article sur la construction du discours autour de la puissance paternelle aux XIX^{ème} et XX^{ème} siècles. La partie organisée en série traite, comme l'annonce l'amorce, des conceptions différentes sur la partie qui devrait avoir la garde légale de l'enfant après le divorce des parents. *Le principe d'énumération (PE*, cf. section 4.2.4) de la série, *trois perspectives différentes*, est déjà présenté par elle. L'amorce forme également l'objet de référence pour le premier item. Celui-ci est introduit par la simple anaphorique *la première*. Le marquage du deuxième item suit le modèle établi par l'amorce et le premier item et leur est parallèle : *la seconde perspective*. Le dernier item, enfin, brise ce parallélisme. Au lieu de répéter le principe d'énumération d'origine, cet item utilise un autre, *législation*. De plus, il n'est pas non plus introduit par une expression référentielle mais par le pronom autre : *d'autres législations*. Cette expression est encore suivie par le connecteur exact *enfin*.

Exemple 29

- 0 Les recherches portant sur les jugements rendus en matière de garde légale au xix^e siècle identifient **trois perspectives différentes**.
- 1 Très présente à cette époque dans plusieurs États américains ainsi qu'en France et en Angleterre, **la première** privilégiait « les meilleurs intérêts de l'enfant » en leur donnant préséance sur les droits concurrents des parents. Selon S. Boyd, les lois canadiennes s'inscrivaient dans ce courant à la fin du xix^e siècle.

- 2 **La seconde perspective** recourait plutôt à l'argument dit « de l'âge tendre » en catégorisant les étapes du développement infantile selon la nécessité d'une présence maternelle ou d'une influence paternelle. Toujours selon Boyd, ce type d'argument se développe aussi au Canada, mais seulement à partir du début du xxe siècle.
- 3 **D'autres législations enfin**, tenant compte du comportement conjugal des époux, attribuaient la garde des enfants à la partie lésée.

Fish

3. Séries partiellement non marquées (PNM). Dans ces séries, au moins l'un des items est non marqué et un item est explicitement introduit³⁶. L'Exemple 28, dont le premier item est non marqué, illustre les structures non parallèles où le principe d'organisation varie. La série accumulative de l'Exemple 22 est également partiellement non marquée : les premier et troisième items ne sont pas explicitement introduits. L'Exemple 30 suivant présente encore une autre structure partiellement non marquée, traitant du changement de terminologie en Bordelais des différentes parties de la charrue utilisée en viticulture. Les deux items de la série décrivent deux périodes du développement de la charrue pendant lesquelles la terminologie est restée inchangée : le premier se réfère à la période pendant laquelle le fer est devenu le matériau des charrues et le deuxième à une période plus tardive qui a vu le remplacement des bœufs par les chevaux. Ils sont donc organisés selon un axe temporel. En ce qui concerne le marquage des items, le premier est non marqué, bien que la construction *verbe+inchangé* soit répétée dans les deux items. Le deuxième est explicitement introduit par *dans un deuxième temps*, signalant que les événements décrits sont postérieurs à ceux du premier item. La série se termine encore par une clôture, marquée par le connecteur *ainsi*.

Exemple 30

... Ainsi a-t-on pu voir en Médoc l'ancien araire en bois, la corba (Viaut 1992 : 132-134), utilisée pour chausser les rangs de vigne, devenir entièrement métallique, durant la première moitié du XXe siècle, sans changer fondamentalement de forme, tant que cette spécialisation des instruments aratoires s'est avérée fonctionnelle et que la traction s'est faite au moyen de paires de boeufs.

- 1 La terminologie traditionnelle désignant les différentes parties de cette charrue restèrent, durant cette période de transformation, inchangées, seul le matériau, en passant du bois au fer, ayant en fait évolué.
- 2 Elle demeura **encore** inchangée sur le fond, **dans un deuxième temps** (en gros, durant la première moitié du XXe siècle), lorsque l'on passa des boeufs aux chevaux (Viaut 1992 : 67-69), l'efficacité reconnue de cette méthode de travail du sol ayant pu transcender le changement de mode de traction.

Cl. Il peut **ainsi** y avoir transition, par adaptation tout au moins partielle, du savoir ancien au nouveau, sans disparition, par conséquent, de l'un au profit de l'autre.

Viaut

4. Séries non marqués (NM). Contrairement aux séries partiellement non marquées, aucun des items des séries non marquées n'est explicitement introduit par un marqueur.

³⁶ Dans les séries PNM, la conjonction de coordination *et* ne suffit pas comme seul marqueur. Si c'est le cas, la série est analysée comme NMet.

Comme le spécifient les critères d'annotation définis en section 5.2.2, ces séries doivent, pour être incluses dans l'analyse, être précédées d'une amorce exacte et / ou suivies par une clôture exacte indiquant le nombre d'items de la structure.

Ci-dessous, l'Exemple 31 est une structure non marquée typique. L'article dont la série est tiré discute de l'utilisation du patrimoine dans le cadre d'activités de découverte du monde en maternelle. Dans la série, l'auteur pose des hypothèses d'apprentissage des élèves dont certains ont été « sensibilisés de manière explicite au cadre de vie urbain médiéval » et d'autres ne l'ont pas été. La série commence par l'amorce qui définit le principe d'énumération et le nombre d'items de la série : *deux hypothèses*. Les items proprement dits sont formés par ces deux hypothèses, sous forme de question. Malgré le non marquage, ils sont facilement identifiables grâce à l'équivalence créée par l'amorce et la correspondance des items aux hypothèses, comme annoncé par celle-ci. La série est encore close par une expression qui ne prend pas position sur les questions posées mais présente une nouvelle affirmation : *On peut au moins affirmer une chose*. Bien que la clôture ne soit pas introduite par un connecteur conclusif typique, l'adverbe *au moins* semblerait indiquer dans cette série une transition vers une nouvelle thématique et, ainsi, la fin de la série.

Exemple 31

- 0 A ce stade, **deux hypothèses** peuvent être avancées :
- 1 Est-ce le côté inhabituel ou « exotique » du jeu « mât de cocagne » qui a retenu toute l'attention des élèves en difficultés ?
 - 2 Est-ce le fait d'y avoir réellement joué ?
- Cl. On peut au moins affirmer une chose : c'est bien parce que ce jeu n'était pas connu des élèves de maternelle, qu'après l'indispensable explicitation orale puis écrite (élaboration d'une fiche technique), la maîtresse leur a proposé de l'appréhender en y jouant réellement.

Loison

5. Séries NMet. Dans ces séries, le seul marqueur explicite utilisé est la conjonction de coordination *et* (dans quelques cas *ou*). Bien qu'en théorie ces structures puissent être analysées comme partiellement non marquées, dans la pratique, elles ressemblent plutôt aux séries non marquées. Ce sont souvent, en effet, des listes d'items très courts. De plus, de même que les séries non marquées, ces séries sont soit précédées, soit suivies d'une expression définissant le nombre d'items. L'item introduit par la conjonction est presque toujours le dernier bien que, dans quelques cas exceptionnels, il corresponde à l'avant-dernier de la série

Cette catégorie comprend donc uniquement les structures dans lesquelles la conjonction de coordination est le seul marqueur présent. Si la série incluait également d'autres marqueurs, elle serait analysée comme partiellement non marquée ou comme une structure mixte. Or, il semble que les conjonctions de coordination sont presque toujours utilisés comme unique marqueur explicite de la série. Nous avons donc décidé d'analyser ces structures séparément, comme une catégorie indépendante.

Ci-dessous, l'Exemple 32 est tiré d'un article intitulé « Les composantes de l'entraînement à l'improvisation actives dans la gestion des imprévus en salle de classe du niveau secondaire ». Comme annoncé par l'amorce, la série énumère trois formes d'entraînement à l'improvisation. Le deux premiers items sont non marqués et correspondent simplement aux syntagmes nominaux. Le troisième item est très similaire à ceux-ci mais il est, de plus, signalé par la conjonction de coordination *et*.

Exemple 32

- 0 Bien que les stages de formation pratique soient évidemment des occasions de développer la capacité à improviser puisque les étudiantes et étudiants doivent faire face à de nombreux imprévus lors de leurs activités de prise en charge d'une classe, notre analyse porte plutôt sur **les trois formes d'entraînement** précédemment mentionnées :
- 1 improvisation théâtrale,
 - 2 cours universitaire
 - 3 **et** animation de groupes.
- Cl. **En effet**, à cause des lacunes observées dans la gestion de classe des enseignants novices malgré le fait qu'ils aient bien réussi leurs stages, il est important de mettre en évidence des formes d'entraînement à l'improvisation complémentaires à ceux-ci. Ainsi, l'expérience relative à l'une ou l'autre de ces trois formes d'entraînement a servi de critère au recrutement des enseignantes et des enseignants qui ont participé à la recherche.

Pelletier

En définitive, nous venons de définir cinq catégories de séries différentes afin d'examiner la façon dont les séries linéaires simples sont marquées. La Figure 7 illustre leurs fréquences. Comme nous le notons, les plus fréquentes sont les séries partiellement non marquées formant presque la moitié (47 %) des structures simples. Les séries homogènes représentent presque un tiers (28 %) et les séries mixtes 9 % de ces séries. Ensemble, les séries non marquées et les séries NMet comprennent 16 % du total.

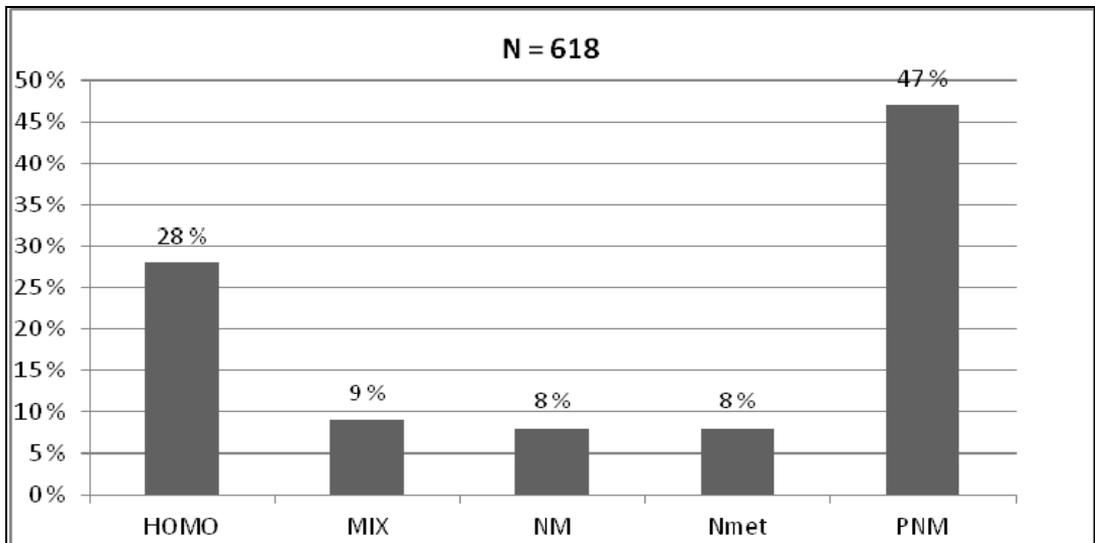


Figure 7 Proportions des catégories des séries linéaires simples.

Les fréquences des catégories de séries indiquent donc que les séries homogènes introduites par le même type de marqueurs, qu'on pourrait imaginer les plus fréquentes, forment dans la pratique seulement un tiers des structures du corpus. Les plus utilisées, en revanche, sont les séries avec au moins un item non marqué, ce qui permet d'avancer que le marquage partiel de l'organisation textuelle est le plus fréquent, tout du moins dans le cas des séries linéaires. La fréquence du non marquage et la variété des séries linéaires sont reflétées également par les séries non marquées et NMet. De plus, ces structures attestent de l'importance des amorces et des clôtures dans le marquage de l'organisation textuelle et la structuration du texte.

Outre les catégories que nous venons de présenter, le marquage des séries peut aussi être examinée à partir des types de marqueurs et de combinaisons de marqueurs utilisés. La Figure 8 illustre la fréquence des marqueurs différents utilisés dans une seule série³⁷. Comme nous le constatons, les séries dont les connecteurs sont les seuls marqueurs sont les plus fréquentes. Ces séries sont soit homogènes, soit partiellement non marquées. D'autres structures similaires sont étiquetées par *L* et *A*. Leurs marqueurs sont des constructions Num+N ou des expressions anaphoriques. La fréquence de ces dernières - 4 % - nous semble quelque peu inattendue car, intuitivement, nous les imaginions plus courantes.

³⁷ Pour la simplicité, la catégorie *C* inclut également les expressions cadratives.

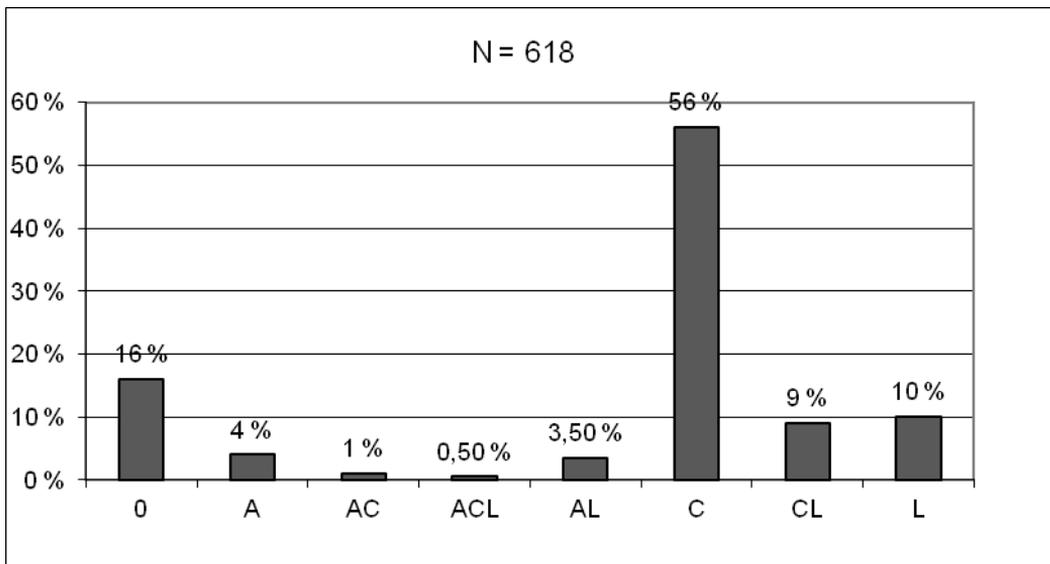


Figure 8 Fréquences des marqueurs et des combinaisons de marqueurs dans les séries simples.

Les deux caractéristiques des séries linéaires que nous venons d'examiner, la présence des items non marqués et leur introduction par des marqueurs différents, sont souvent associés au non parallélisme de ces structures³⁸. Dans ce qui suit, nous discuterons cette problématique et réfléchirons sur la véracité de cette hypothèse dans le cadre de notre corpus.

Tout d'abord, si nous considérons que le non parallélisme est une conséquence directe du non marquage d'un ou plusieurs items de la série, les résultats que nous venons de présenter soutiennent la fréquence des séries non parallèles, constatée dans des travaux antérieurs discutés dans 4.2.1. Or, d'après notre analyse, ce n'est pas nécessairement le cas parce que les items de la série peuvent être équivalents même quand non marqués.

Déjà en définissant le schéma d'annotation, nous avons discuté des indices complémentaires susceptibles de contribuer à l'identification des items sans faire partie des marqueurs explicites. Les titres et la répétition des mêmes items lexicaux, entre autres, appartiennent à ce groupe. De même, nos exemples incluant des items non marqués signalent que l'équivalence structurelle des items, comme leur correspondance aux questions (cf. l'Exemple 31) ou au principe d'énumération annoncé par l'amorce (cf. l'Exemple 32), contribue à l'identification des items non marqués. Ainsi, les structures avec des items non marqués ne nous semblent pas nécessairement mal formées, déviantes ou difficiles à interpréter. Pour ces raisons, leur analyse comme non parallèles ne nous semble pas justifiée. En présence d'indices complémentaires ou autres caractéristiques des séries parallèles, ces séries ressemblent plutôt aux énumérations étudiées dans le projet Annodis (Ho Dac *et al.* 2009), par exemple. Elles incluent dans leur analyse des structures

³⁸ Cf. section 4.2.1.

non ordonnées dont les items sont identifiés grâce à leur équivalence ou, selon nos termes, parallélisme. L'Exemple 32 précédent, notamment, est similaire à ces énumérations : non marqués, les items y suivent le principe d'énumération établi dans l'amorce et sont ainsi parallèles.

Si le non marquage en soi n'a pas pour effet direct le non parallélisme de la série, l'identification automatique du parallélisme, ou non, des items d'une série s'avère très ardue avec nos méthodes³⁹. Les types de marqueurs utilisés dans les séries offrent cependant un aspect pour l'examen de ce phénomène. Surtout, les structures signalées par des combinaisons de marqueurs sont potentiellement non parallèles, comme noté déjà par Hempel & Degand (2008). Ces structures font partie des catégories de série mixtes ou partiellement non marquées et sont, dans la Figure 8, étiquetées AC, ACL, AL ou CL. Les CL, les plus fréquentes en couvrant 9 % du total, sont divisées régulièrement entre structures mixtes et partiellement non marquées.

Afin d'examiner le parallélisme des séries signalées par des combinaisons de marqueurs plus en détail, nous avons mené une étude manuelle sur ces structures. Cette analyse indique que ce sont surtout les séries CL, AC et ACL qui sont non parallèles alors que les AL sont plutôt parallèles. Notre examen montre que les indices souvent associés au parallélisme, tels que la répétition du principe d'énumération dans chaque item de la série et la similarité structurelle des items (cf. l'Exemple 31 à la page 129), sont très rares dans les séries CL signalés par des connecteurs et des constructions Num+N. Seules neuf des 56 séries examinées incluent l'un de ces indices. Dans toutes les autres, la répétition ne couvre pas tous les items ou les items ne sont pas structurellement équivalents. Très souvent, le principe d'énumération n'est répété que dans une partie des items et leur longueur ainsi que leur statut syntaxique varient aussi (cf. l'Exemple 28 à la page 125). Par exemple, l'un des items peut être présenté dans une subordonnée et un autre couvrir plusieurs phrases. Le déséquilibre structurel des items des séries CL est encore souligné par la présence de marqueurs différents, certains étant adverbiaux et d'autres sujets ou objets de la phrase, notamment. Par conséquent, les structures syntaxiques des items varient aussi, ce qui renforce le non parallélisme de la structure.

L'Exemple 33 ci-dessous illustre une autre série étiquetée CL. Il s'agit d'une structure mixte avec une amorce introduisant le principe d'énumération *différentes propriétés linguistiques* et une clôture qui se réfère aux exemples présentés dans les items tout en indiquant la fin de la structure. Comme dans l'Exemple 28 à la page 125, pour cette série aussi le non parallélisme des items est souligné par les longueurs variées des items : tandis que les longueurs des autres items avoisinent 150 mots, le quatrième item en couvre 659.

³⁹ Lors de l'annotation, nous avons identifié quelques caractéristiques des items non marqués qui peuvent être associées aux séries parallèles, ainsi des liens de cohésion lexicale reliant les items non marqués. Comme nous le montrerons en section 7.3, ces indices sont pourtant insuffisants à eux seuls pour décider du parallélisme ou non de la série. Par conséquent, la distinction automatique des séries dont les items non marqués causent le non parallélisme et celles pour lesquelles ce n'est pas le cas n'est pas possible.

De plus, le principe d'énumération annoncé dans l'amorce n'est pas répété mais varie entre les items. Comme indiqué pour ces séries, les marqueurs sont des connecteurs et des constructions Num+N. Les premier et troisième items sont introduits par *d'abord* et *de même* et les autres items par des constructions composées du pronom *autre* et d'un nom variant dans chaque item : *un autre cas*, *un autre phénomène* et *un autre critère*.

Exemple 33

- 0 Je voudrais maintenant parcourir les différentes **propriétés linguistiques** que j'ai pu relever dans mon échantillon d'occurrences de la séquence « , et ».
 - 1 Il y a **d'abord** des cas où la virgule permet de distinguer des niveaux de parathésie, lorsque et est répété : [...]
 - 2 **Un autre cas** de figure susceptible de motiver la présence d'une virgule devant et correspond à un décalage syntaxique entre les éléments conjoints. Par exemple dans : [...].
 - 3 **De même**, les grammaires citées suggèrent que la virgule apparaît devant et quand le sujet change. Par exemple, dans Grevisse (1969, p. 1146) : [...]
 - 4 **Un autre phénomène** syntaxique semble jouer souvent, c'est la complexité ou la longueur du membre conjoint introduit par et. A ce propos, Drillon (1991, p.180) cite Tassis (1859) qui, selon lui, « ne recommande la virgule que lorsque les propositions sont longues (et qu'on risque d'oublier le verbe ou le subordonnant qui a commandé la première). [...]
 - 5 **Un autre critère** semble jouer un rôle, c'est le statut prédicatif de l'élément introduit par et. Par exemple, en cas d'ellipse, la virgule est très courante : [...].
- CL. **Ces exemples** sont à la fois importants et intéressants parce qu'ils montrent, à mon sens, que la question n'est pas directement syntaxique : la virgule n'étant pas une marque linguistique, mais une marque du code graphique, ne peut pas en fait avoir une fonction syntaxique.

Lambert

Comme la présence des connecteurs et des marqueurs pouvant fonctionner comme sujet ou objet de la phrase semble être un des facteurs influençant le non parallélisme des séries CL, nous présumons que les quelques séries incluant des anaphoriques et des connecteurs (AC et ACL) le sont aussi. Or, notre analyse indique également que ces séries non parallèles ne sont ni difficiles à identifier, ni incohérentes. Le non parallélisme provient plutôt de la différence entre les items, atypique pour les structures dans lesquelles l'équivalence entre les items est souvent considérée comme une valeur de défaut. De plus, nous devons constater que le parallélisme, ou non, n'est pas absolu ; la classification en séries parallèles et non parallèles n'est pas possible. Il s'agit davantage d'un trait relatif qui caractérise mieux certaines structures que d'autres.

Contrairement aux séries CL, AC et ACL, les séries AL, en revanche, sont plutôt parallèles. Déjà, l'analyse automatique révèle que toutes sont mixtes et dépourvues d'items non marqués. De plus, une analyse manuelle révèle qu'elles suivent très majoritairement le modèle illustré par l'Exemple 34 ci-dessous. Le principe d'énumération change au milieu de la structure dans seulement trois des 22 séries. Ainsi, bien que les marqueurs soient en principe annotés différemment, ils sont, dans la pratique, très similaires et la série ne nous

paraît pas non parallèle. Dans l'Exemple 34, l'amorce indique le nombre d'items et le principe d'énumération, *deux remarques*, auquel les deux items se réfèrent. Le premier item est signalé uniquement par l'anaphorique *la première* tandis que, dans le second, l'anaphorique est combinée avec le nom *remarque*, annoncé dans l'amorce.

Exemple 34

- 0 J'ajouterais à ce propos **deux remarques**.
- 1 **La première** concerne la direction de la connexion. Cette dernière peut en effet s'effectuer, non plus du sujet vers le verbe, mais du verbe vers le sujet, comme le montre la comparaison de l'exemple d'Augustin avec le célèbre exemple de Donnellan (1966, pp. 285-286) : [...].
- 2 **La seconde remarque** sera relative, non plus aux rôles respectifs du sujet et du prédicat dans l'expression de la connexion, mais au type même de la connexion. Il semble en effet qu'il en existe de deux sortes : celle que nous venons de voir et qui est de l'ordre de la conséquence ; mais aussi celle que réalise (9) : [...]

Carel

Pour conclure cette sous-partie sur le marquage des séries simples, notre analyse met en valeur surtout la fréquence des séries partiellement non marquées et donc du marquage partiel de l'organisation textuelle. Les séries homogènes, dont les items sont signalés par le même type de marqueurs, ne forment qu'un tiers du total des structures repérées. Les types de marqueurs les plus fréquents sont les connecteurs, les seuls marqueurs utilisés dans 56 % de l'ensemble.

De plus, les résultats de cette section prouvent que les items non marqués ne rendent pas la série inévitablement non parallèle et qu'ils peuvent, malgré le non marquage, suivre le principe d'énumération établi par l'amorce et être facilement repérables. Dans ce cas, la série ressemble davantage aux énumérations non ordonnées examinées entre autres par Ho Dac *et al.* (2009), pour lesquelles l'ordre entre les items n'est pas un aspect important. Nous pouvons estimer que c'est le cas également pour les séries de notre corpus, incluant des items non marqués mais autrement parallèles. De plus, il est possible que l'auteur ait simplement souhaité éviter la répétition des marqueurs devant tous les items. Cependant, il est important de noter que toutes les séries incluant des items non marqués ne sont pas parallèles ou faciles à appréhender. Le déséquilibre entre les items et, par conséquent, le non parallélisme de la série peuvent être imputables, notamment, au changement du principe d'énumération au milieu de la structure.

Enfin, l'analyse de cette section propose également que l'usage des différents marqueurs dans une série peut également influencer le parallélisme ou le non parallélisme. D'après notre analyse, les structures introduites par plusieurs types de marqueurs avec des connecteurs semblent surtout plutôt non parallèles. Outre les marqueurs différents, c'est la conséquence du changement de principe d'énumération au milieu de la série, d'une part, et du déséquilibre syntaxique et fonctionnel des items introduits par ces marqueurs, de l'autre. En revanche, les structures signalées par des anaphoriques et des constructions

numéro+nom sont relativement parallèles grâce à la répétition du principe d'énumération dans les items.

Bref, bien qu'il soit profitable de présenter un chiffre exact de fréquence des séries non parallèles dans notre corpus, cette caractéristique n'est pas automatiquement repérable avec nos méthodes, surtout du fait des items non marqués qui peuvent toutefois être parallèles. En outre, comme le parallélisme semble plutôt relatif qu'absolu, il n'est pas toujours possible de séparer catégoriquement les séries qui le sont de celles qui ne le sont pas. Selon notre analyse manuelle, nous pouvons sans doute avancer que, sauf quelques exceptions, les séries étiquetées AC, ACL et CL sont plutôt non parallèles, surtout à cause du déséquilibre des items. Ces séries couvrent 14 % des séries simples. Le caractère des autres structures requiert pourtant des analyses plus détaillées.

6.3.2 SERIES ENCHAINEES

Le marquage des séries enchaînées diffère de celui des structures simples en ce que, dans les enchaînées, les séries inférieures ne sont pas autonomes et le marquage d'une telle série peut influencer celui des autres. Le schéma d'annotation tient également compte de cet aspect. Comme nous l'avons exposé, les critères de marquage d'une série simple faisant partie d'une structure supérieure ne sont pas aussi sévères que ceux des séries simples individuelles.

Comme point de départ de l'examen du marquage des structures enchaînées, nous utilisons les catégories de séries simples présentées à la section précédente. Les catégories de séries enchaînées sont créées en combinant celles des séries inférieures. Le nombre de catégories risquant de s'élever du fait de la multitude de combinaisons possibles, nous avons combiné les séries NM et NMet pour toutes les étiqueter NM.

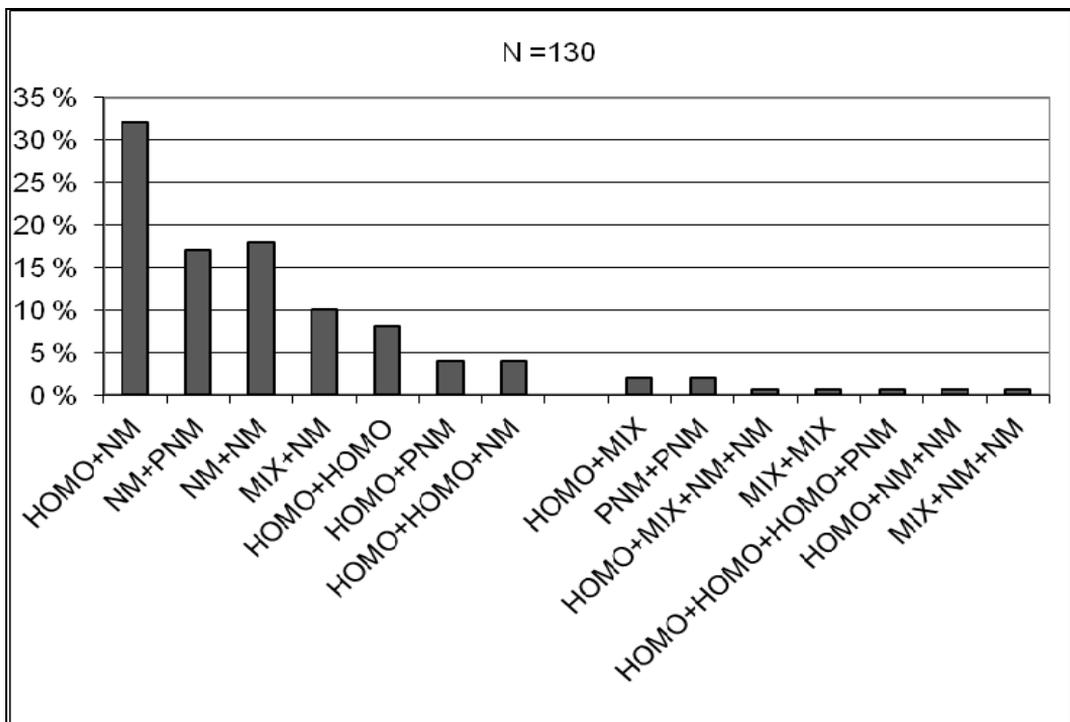


Figure 9 Fréquences des catégories de séries enchaînées.

Les fréquences des catégories de séries enchaînées sont présentées ci-dessus en Figure 9. Comme nous le notons, la distribution des structures est relativement dispersée malgré la fusion des séries NM et NMet. La figure révèle que les plus utilisées sont les séries enchaînées composées de séries homogènes et de séries non marquées. Les structures dont toutes les séries inférieures sont soit non marquées, soit homogènes sont aussi relativement fréquentes. Celles introduites par une structure non marquée et une partiellement non marquée couvrent également 18 % du total et celles composées de non marquées et mixtes 10 %. Les autres catégories sont minoritaires.

L'examen du marquage des séries enchaînées peut aussi passer par l'étude des fréquences des catégories de séries simples dans les structures supérieures. La Figure 10 illustre les proportions de séries dans lesquelles les catégories différentes sont utilisées. Ainsi, nous notons que les séries non marquées sont très majoritairement présentes, dans 82 % du total, et qu'aussi les séries homogènes sont utilisées dans 52 % et les partiellement non marquées dans 24 % des séries. Les séries mixtes, enfin, sont plutôt rares et font partie de seulement 15 % des structures supérieures.

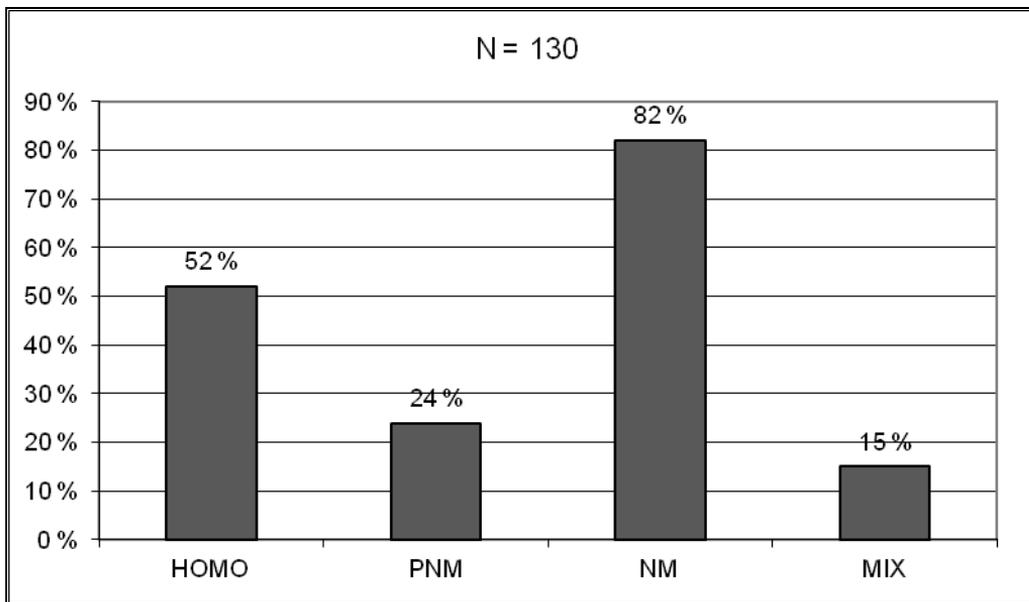


Figure 10 Proportions des catégories de séries simples dans les structures enchaînées.

Comparées aux séries simples dont 47 % sont partiellement non marquées, 28 % homogènes et 16 % non marquées (NM et incluses), les séries enchaînées sont très fréquemment composées de structures non marquées et de séries explicitement introduites par le même type de marqueurs. Les enchaînées sont donc souvent explicitement introduites. Par conséquent, il nous semble que notre décision initiale de ne pas les considérer comme non parallèles (cf. 4.2.4) est justifiée. En revanche, l'effet répétitif potentiel du marquage de tous les items de chaque série inférieure peut être une raison de la fréquence des séries non marquées. De plus, le fonctionnement des séries non marquées comme sources pour les autres séries de la structure doit expliquer leur usage fréquent; elles seules introduisent les items auxquels se réfère une autre série de la structure en explicitant également l'ordre entre eux. Dans l'Exemple 27 à la page 124, notamment, la première série, non marquée, introduit les items auxquels les séries suivantes se réfèrent par des expressions anaphoriques. En revanche, l'Exemple 24 à la page 117, dont le premier composant est homogène et le second non marqué, est plutôt cataphorique et commence par une présentation brève des items pour se poursuivre par leur énumération proprement dite.

Bien que moins courantes que celles composées de séries non marquées et homogènes, les combinaisons de séries non marquées et partiellement non marquées couvrent 18 % des séries enchaînées. L'Exemple 35 ci-dessous illustre un cas extrait d'un article du domaine d'éducation.

Exemple 35

0 Un certain nombre **de raisons ou d'obstacles** est alors invoqué.

RAISONS ET OBSTACLES INVOQUÉS

0 Les raisons ou obstacles signalés sont de **quatre ordres** :

1 épistémologique,

2 didactique,

3 pédagogique

4 **ou** matériel.

1 La principale raison épistémologique avancée concerne la structuration progressive de la temporalité qui, d'après les instructions officielles de 2002, doit être nourrie par le récit d'événements du passé. [...]

2 L'obstacle didactique le plus souvent énoncé est celui de la faisabilité des programmes. On estime en effet que les élèves de cycle 1 sont trop jeunes pour percevoir la profondeur du passé, [...].

3 **Le troisième type d'obstacle** concerne le champ pédagogique – investi par le « pédagogue » qui, au sens étymologique du terme, « accompagne l'élève » –, et pose la question de la formation des enseignants dont certains avouent ne pas maîtriser ce type d'approche articulant utilisation du patrimoine et constitution d'un premier capital historique.

4 **Les dernières raisons** invoquées sont relativement classiques pour ne pas dire traditionnelles puisqu'elles sont d'ordre matériel. A ce niveau, on signale le manque de moyens financiers [...].

Cl. **Pour clore** ce chapitre, nous évoquerons la question de la pertinence de l'orientation des instructions officielles de 2002 qui recommandent de mettre en oeuvre des activités à caractère patrimonial, jugées par bon nombre d'enseignants comme étant prématurées pour ne pas dire inefficaces voire inutiles. [...]

Loison

Dans l'Exemple 35, la série couvre une sous-section et commence dès la fin de la section précédente par une phrase introductrice, fonctionnant également comme amorce et définissant le principe d'énumération qui inclut, en fait, deux composants différents : *raisons ou obstacles*... L'amorce se poursuit au début de la section, après le titre, et définit le nombre d'items de la structure. Les items suivant l'amorce sont non marqués, à l'exception du dernier qui est introduit par la conjonction de coordination *ou*. Par ailleurs, les items sont courts et formés par des adjectifs qualitatifs décrivant les principes d'énumération annoncés par l'amorce. Les items de la deuxième série, en revanche, sont plus longs et développent la thématique des items de la première série. De ces items, le troisième et le quatrième sont introduits explicitement par des constructions numéro + nom, *le troisième type d'obstacle* et *les dernières raisons*. Les items non marqués suivent également les principes d'énumération annoncés et répètent même les formes lexicales utilisées dans la première série. Pourtant, malgré cette cohésion lexicale, la série n'est peut-être pas très typique à cause des deux principes d'énumération. De plus, bien que les items soient facilement repérables et de longueurs similaires, leurs marquages différents conjointement avec l'amorce atypique rendent cette série plutôt non parallèle.

Pour les types de marqueurs, les distributions des séries enchaînées sont encore plus dispersées que pour les catégories de série. La Figure 11 et la Figure 12 ci-dessous présentent les fréquences des combinaisons les plus courantes ainsi que les nombres de séries dans lesquelles les marqueurs et leurs combinaisons sont utilisés. La fréquence des séries homogènes et non marquées paraît dans ces figures sous la forme des séries étiquetées 0 et introduites par un seul type de marqueur. Ces types couvrent la majorité des cas.

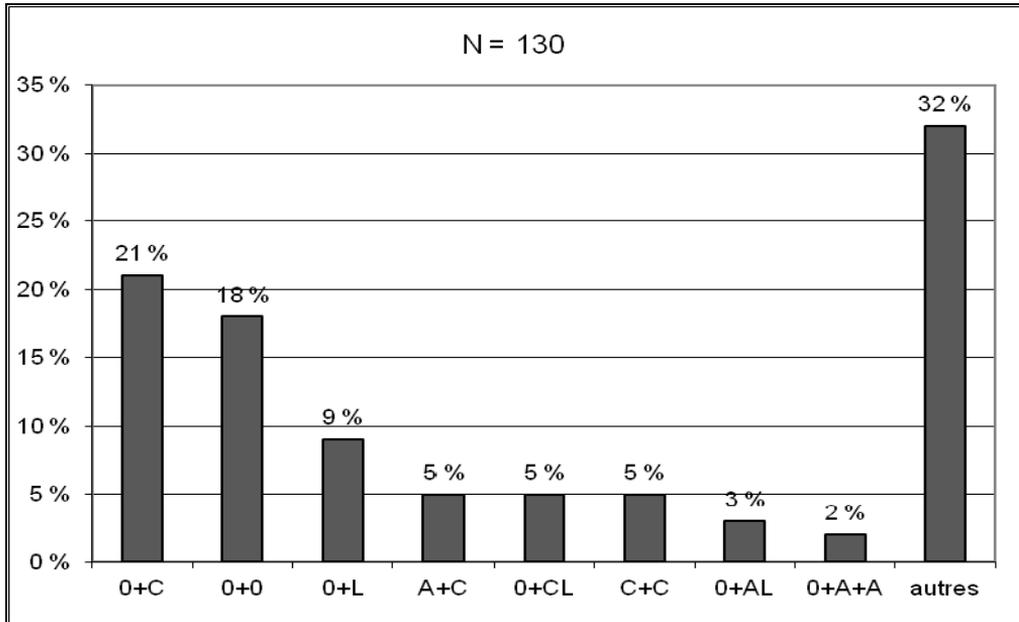


Figure 11 Fréquences des combinaisons de types de marqueurs dans les structures enchaînées⁴⁰.

⁴⁰ C = connecteur, L = construction Num+N, A = expression anaphorique.

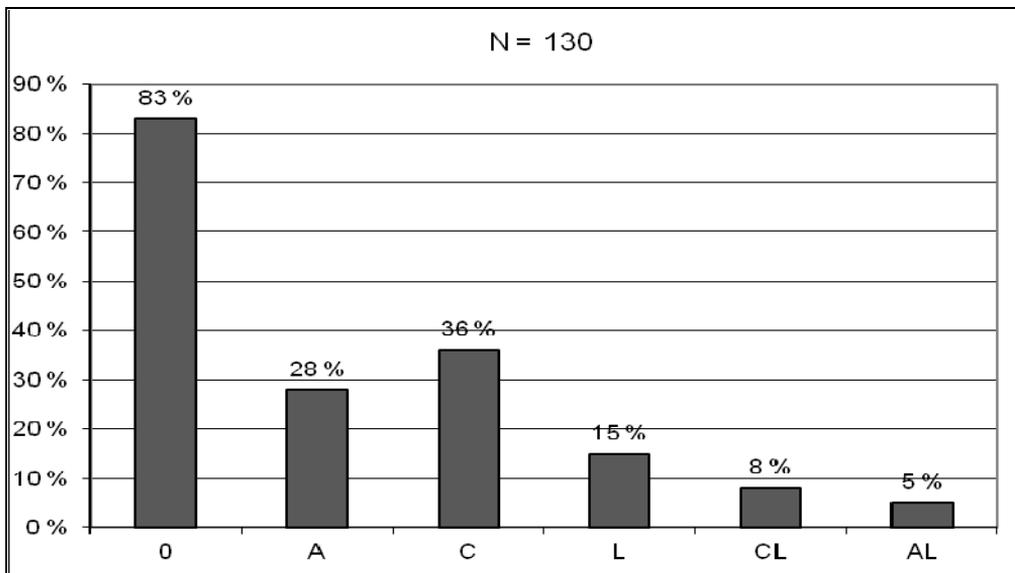


Figure 12 Proportions des séries enchaînées dans lesquelles ces combinaisons de marqueurs sont présentes.

Comparés aux types de marqueurs utilisés dans les séries simples, nous relevons que les connecteurs sont également toujours les marqueurs les plus utilisés dans les structures enchaînées. Les expressions anaphoriques, en revanche, sont beaucoup plus manifestes que dans les séries simples. Un autre signe de la fréquence des séries enchaînées référentielles. Les combinaisons de marqueurs différents sont également plutôt rares comme nous avons déjà pu le déduire de la rareté des séries mixtes. Pareillement, les constructions numéro+nom sont plus rares que dans les séries simples.

6.3.3 AMORCES ET CLOTURES

Les amorces et clôtures participent également au marquage des séries linéaires, comme le montrent principalement nos exemples des séries non marquées et NMet. En section 6.1, nous avons brièvement présenté leurs fréquences dans le corpus en général : un tiers environ des séries simples et des séries enchaînées sont sans amorce et sans clôture et deux tiers incluent au moins l'un de ces composants. Dans cette section, nous traitons ces composants plus en détail en présentant les fréquences de leurs différents types.

La Figure 13 présente les fréquences des types d'amorce dans le corpus et le Tableau 5 les abréviations utilisés dans la figure. Les types appliqués ont été définis en section 5.2.2, lors de la discussion sur le schéma d'annotation. Comme nous le notons, les amorces exactes⁴¹ indiquant le nombre d'items de la série sont le type d'amorce plus fréquent, formant 24 %

⁴¹ Du type *Cette série comprend deux items...*

des séries simples et 48 % des enchaînées. Les amorces vagues⁴² représentent 11 % et 3 % des cas et les autres types sont encore plus minoritaires.

La distribution des types d'amorce montre que, outre le marquage des items, la structure de la série est souvent annoncée avant même la série et que les amorces ont un rôle notable dans l'introduction de ces structures. De plus, nous relevons que les amorces exactes sont encore beaucoup plus fréquentes dans les séries enchaînées que dans les simples. C'est l'indication que les séries enchaînées sont introduites plus exactement que les séries simples. La rareté relative des autres types d'amorce dans les structures enchaînées conforte cette supposition ; tandis que 20 % des séries simples sont précédées d'une amorce ne spécifiant pas son nombre d'items, le chiffre chute à 9 % pour les structures enchaînées.

Types d'amorce	
0=0	Pas d'amorce
0=IMP	Amorce implicite
0=NUM	Énumération exacte
0=VAG	Énumération vague
0=Q	Question
Types de clôture	
R=0	
R=CON	Connecteurs conclusifs / explicatifs
R=ACON	Autres connecteurs
R=ANA	Reprises anaphoriques
R=NUM	Reprises anaphoriques spécifiant le nombre d'items de la série

Tableau 5 Abréviations utilisées.

⁴² Du type *Cette série comprend quelques items...*

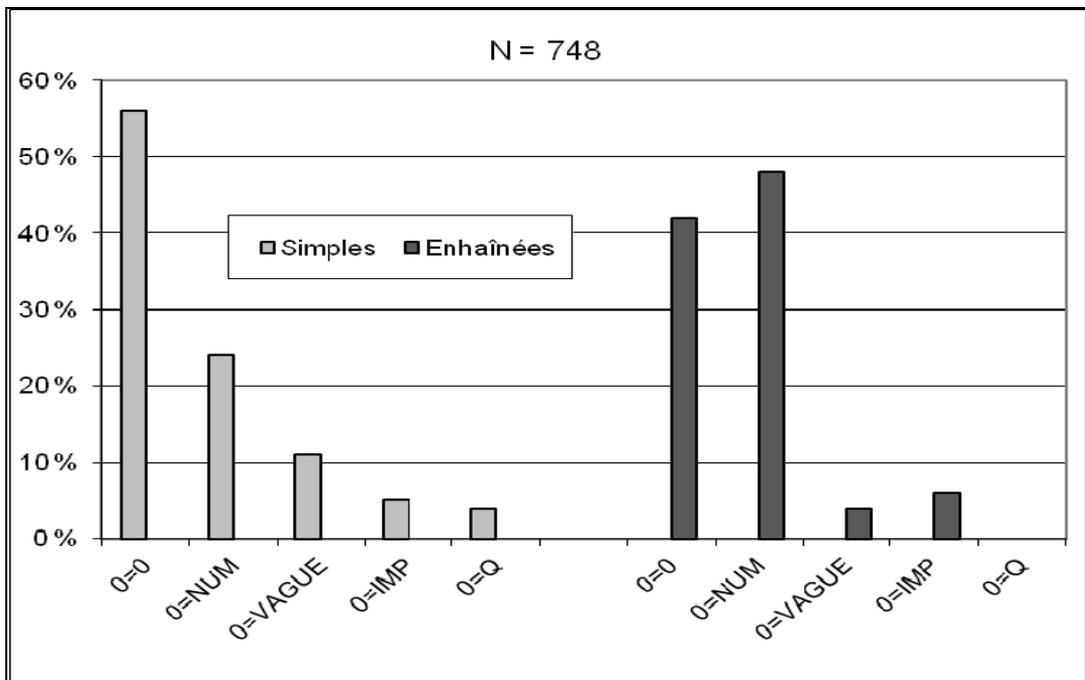


Figure 13 Amorces dans les séries simples et dans les séries enchaînées.

Nous avons déjà présenté plusieurs séries illustrant les amorces exactes et vagues. Ci-dessous, nous décrivons une série précédée d'une amorce-question, extraite d'un article en linguistique sur l'origine des langues. Dans l'exemple, l'amorce pose une question sur l'absence des études créoles sur l'origine des langues dans les années 1880. Les items, tous introduits par des MIL, énumèrent des raisons pour lesquelles on aurait dû l'examiner. Contrairement à ce que l'on pourrait attendre, ils ne répondent pourtant pas directement à la question posée. L'exemple reflète bien la façon dont l'amorce peut annoncer la série linéaire et guider le lecteur dans le texte sans préciser le principe d'énumération ou le nombre d'items de la structure. Évidemment, le marquage explicite des items y contribue aussi.

Exemple 36

- 0 Les études créoles démarrent dans les années quatre vingt du XIXe siècle (voir Adam (1883) et les textes rassemblés dans Schuchardt 1922). **On pourrait se demander pourquoi elles ne sont pas, dès cette époque, intervenues directement dans la question de l'origine des langues.**
- 1 **D'abord**, les créoles sont considérés comme des langues formées ;
 - 2 **ensuite**, ils invalident les hypothèses biologiques sous la forme raciale (une langue = une race) spécifique de l'époque, puisqu'on parle à leur propos de langues hybrides (Adam) ou mixtes (Schuchardt) ;
 - 3 **enfin**, ils ne respectent pas le modèle arborescent (1 langue mère, n descendantes par divergence), en mettant en lumière le schéma inverse (n « langues mères », 1 descendante par convergence). Au mieux, ils posent des problèmes au modèle génétique du comparatisme (voir Baggioni (1988), sur la polémique Meillet/Schuchardt), tout en renforçant l'aspect historique et contingent de l'évolution

linguistique (Schuchardt finira par admettre que toutes les langues sont des langues mixtes).

Auroux

La Figure 14 présente les fréquences des types de clôture dans les séries simples et enchaînées. De même que pour les amorces, les clôtures spécifiant le nombre d'items de la série sont les plus courantes, 20 % environ des totaux des deux types de série. Les clôtures introduites par des connecteurs conclusifs ou explicatifs représentent également 20 % des séries simples tandis que dans les structures enchaînées elles se limitent à 8 % de l'ensemble.

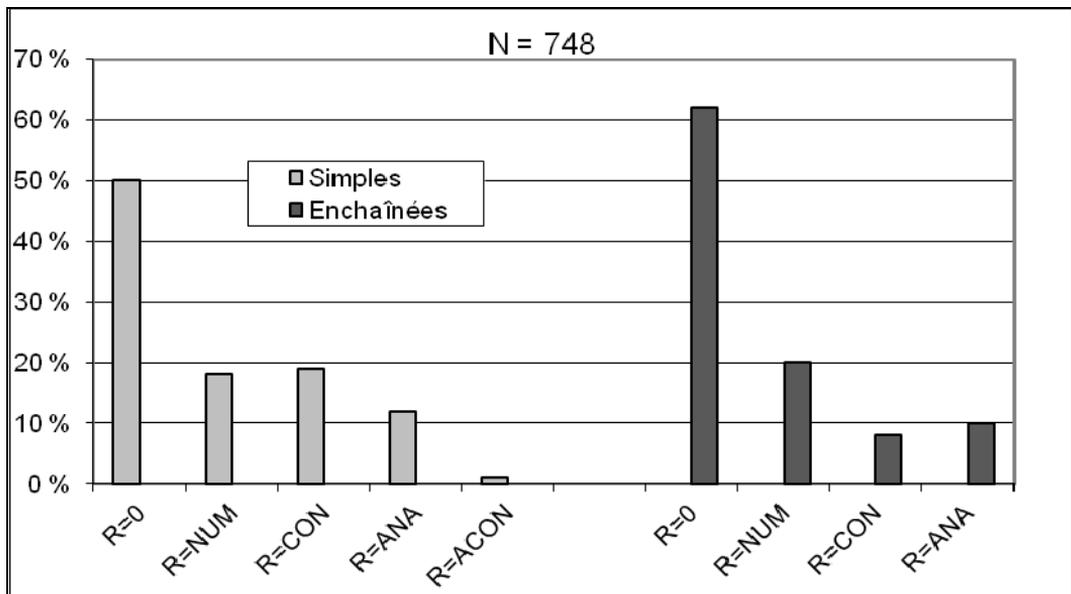


Figure 14 Clôtures dans les séries simples et dans les séries enchaînées.

En définissant les clôtures dans la section 4.2.4, nous avons noté qu'elles sont souvent négligées par les travaux antérieurs. De plus, nous avons cité Sinclair (2004 : 88) qui affirme que les amorces l'emportent naturellement sur les clôtures car apparaissant avant les items. Ainsi, dans les séries précédées d'une amorce, ce serait elle qui guide le lecteur dans l'interprétation du texte tandis que, dans les séries (surtout non marquées) sans amorce, l'effort incombe au lecteur. Poursuivant, Sinclair mentionne que la clôture peut confirmer le résultat de cet effort mais, comme c'est le cas uniquement rétrospectivement, son rôle de guide du lecteur serait moins important que celui des amorces. Nous partageons effectivement la prémisse de Sinclair : les amorces peuvent guider le lecteur plus efficacement que les clôtures car elles les précèdent dans la série. Cependant, les clôtures peuvent tout de même se révéler aussi essentielles dans la formation et l'interprétation des séries linéaires. Les séries dont le nombre d'items n'est défini ni par l'amorce, ni par le marquage des items en sont un exemple. En l'occurrence, c'est la clôture qui, souvent, indique au lecteur la fin de la structure susceptible, autrement, de rester floue et d'exiger

un effort d'interprétation considérable. Cette annonce est importante surtout dans les séries longues dont les items peuvent être non parallèles et la structure globale plus imprécise. Dans ces séries, le rôle du lecteur et de son interprétation est peut-être plus important que dans les séries homogènes et explicitement introduites.

Ci-dessus, l'Exemple 33 à la page 134 illustre une série linéaire dont la clôture revêt un rôle important. Cette structure couvre 1 300 mots et 2,5 pages environ. De plus, comme le marqueur introducteur du dernier item ne précise pas sa place exacte dans la série, le lecteur ne peut pas le reconnaître en lisant le texte. Il doit continuer jusqu'à la clôture afin de comprendre que la série est terminée. Si la clôture était absente, le lecteur devrait interpréter la fin à partir d'autres indices textuels qui, dans cette série, sont moins implicites. De fait, les items sont relativement différents et même leur longueur varient considérablement. Par conséquent, c'est la clôture qui, dans ce cas, a le rôle crucial de guidage du lecteur, lui facilite l'interprétation et rend la structure plus facilement repérable.

L'Exemple 37 illustre un autre cas dans lequel la clôture est importante. Il est extrait d'un article sur la puissance paternelle dans le système judiciaire québécois. Les items listent des travaux sur le rôle des pères au Québec.

Exemple 37

... Pourtant, ceux qui ont tenté d'examiner l'introduction du masculin dans le quotidien domestique nous ont légué des études qui apportent des nuances importantes à notre compréhension de la réalité des hommes dans leur vie privée.

- 1 S'appuyant sur les manuels domestiques publiés pendant l'entre-deux-guerres, l'étude de C. Comacchio souligne que les pères étaient censés jouer un rôle passif et détaché.
- 2 L'étude de R. Rutherford sur la consommation des pères de famille dans les années 1950 démontre qu'ils étaient présents dans la culture de consommation des biens domestiques.

Cl. **Ces deux études** sur les pères de famille au Canada anglais soulignent les courants parfois contradictoires entre l'idéal et ce qui se passait au quotidien.

Fish

Dans l'Exemple 37, la clôture se réfère aux items de la série en indiquant leur nombre et la fin de la structure. La série est également précédée d'un segment pouvant être interprété comme une amorce implicite car il annonce le principe d'énumération de la structure. Selon notre schéma d'annotation, il n'est pourtant pas annoté.

6.4 BILAN : SERIES LINEAIRES, SERIES VARIEES

Cette section avait pour but la présentation générale des séries linéaires du corpus : leurs fréquences, longueurs et modes de marquage. L'examen de ces structures montre, tout d'abord, qu'elles sont un moyen de structuration du texte courant dans les articles de recherche. Nous trouvons, en moyenne, une série linéaire dans un segment de deux pages ou neuf séries linéaires dans un article. Amorces et clôtures sont aussi relativement fréquentes et seul un tiers des séries linéaires du corpus n'inclut ni l'une ni l'autre.

Notre examen montre également la variété des différentes séries linéaires. Dans notre corpus, nous procédons à une distinction importante entre les séries simples, très majoritaires, et les séries enchaînées, composées de plusieurs séries simples. De plus, une partie des séries analysées affichent des traits particuliers pour leur marquage ou la façon dont elles structurent le texte. Dans les séries *méta* et *section*, les titres participent au marquage des items. Dans les structures *argumentatives* et *accumulatives*, ce sont les connecteurs adversatifs qui jouent un rôle important. Les séries *temporelles* se réfèrent au monde extérieur et aux événements s'y produisant. Enfin, les séries *à deux temps* se composent de séries inférieures dont la seconde est thématiquement dépendante de la première. Bien que ces types particuliers forment seulement 19 % des structures repérées, leur identification nous semble importante pour la description de l'ensemble des séries linéaires, d'une part, et pour les analyses menées dans les sections suivantes, d'autre part. Ils reflètent bien également les façons dont l'auteur les utilise afin d'organiser son texte.

La longueur des séries linéaires démontre aussi leur variété : bien que la majorité des structures soit plutôt courte et que la moitié ne dépasse pas 130 mots, le corpus inclut toujours 111 séries de plus de 500 mots. L'organisation du texte en séries linéaires se produit donc pour les segments textuels très courts mais aussi pour des passages très longs. Ces structures longues sont majoritairement des séries méta et section composées de sections titrées. Un élément logique étant donné la longueur des textes du corpus et le fait qu'ils sont tous organisés en sections. Or, les séries longues ne couvrant pas de sections titrées ne sauraient non plus être ignorées. Ces structures ne dépassent pourtant pas, en règle générale, 2 000 mots. Cette limite reflète sans doute un aspect de la structuration des articles de recherche : ils sont segmentés en sections titrées d'une longueur en général supérieure à 2 000 mots.

A l'exception des séries composées de sections entières, les autres types de série particuliers sont généralement utilisés pour organiser des segments textuels similaires aux séries typiques sans étiquette particulière. Parmi les séries simples, seules les séries accumulatives s'étendent sur des segments un peu plus longs. Pour les séries enchaînées, ce sont les structures à deux temps et séries temporelles qui dépassent la longueur moyenne. Les séries enchaînées typiques sans étiquette particulière, en revanche, ne sont pas plus longues que les séries simples typiques bien que leur nombre d'items soit, évidemment, plus grand.

Les séries linéaires sont également très variées dans leur marquage. En général, les résultats indiquent que le marquage partiel de l'organisation textuelle est le plus fréquent, du moins pour les séries linéaires simples. De même, le non marquage complet des items et leur annonce par une amorce ou clôture sont fréquents. L'introduction des items par le même type de marqueurs, couramment utilisée pour illustrer les séries linéaires, n'apparaît pourtant que dans un tiers des séries simples du corpus. De plus, le marquage des items proprement dits est soutenu par les amorces et les clôtures : 24 % des structures simples

incluent une amorce précisant le nombre d'items de la structure et 17 % sont suivies par une clôture exacte. Les types de marqueurs les plus usités sont les connecteurs, les seuls marqueurs dans 56 % des séries simples.

Comparées aux séries simples, les séries enchaînées sont signalées de façon plutôt explicite et exacte. Ce peut être la résultante de la structuration plus complexe de ces séries ; comme elles ne sont pas plus longues que les séries simples sans étiquette particulière mais que leur nombre d'items est supérieur, elles peuvent nécessiter un guidage plus explicite afin de demeurer cohérentes et facilement lisibles. Ainsi, les séries enchaînées sont majoritairement composées de structures non marquées et de structures homogènes dont tous les items sont introduits par le même type de marqueurs. La fréquence du non marquage s'explique par la nature référentielle de ces structures ; les séries non marquées sont souvent des sources pour d'autres séries de la structure. Pour cette raison, les expressions anaphoriques y jouent également un rôle beaucoup plus important que dans les structures simples. Enfin, le marquage exact des structures enchaînées se reflète également dans la présence usuelle des amorces exactes définissant le nombre d'items de la série. Tandis que les amorces exactes précèdent seulement 24 % des structures simples, elles sont utilisées dans 48 % des structures enchaînées.

Enfin, bien que les structures enchaînées soient plus exactement et explicitement introduites que les simples, notre analyse décrit globalement les séries linéaires comme des structures relativement variées dans lesquelles le non marquage et le marquage partiel sont très courants. Ainsi, nos résultats soutiennent ceux de Hempel et Degand (2008) discutés dans la section 4.2. Or, notre analyse a également montré que le non marquage ne complique nécessairement l'identification des items ni ne rend la série non parallèle ; les items peuvent demeurer relativement équivalents et les indices complémentaires de sériation, tels que la répétition des items lexicaux ou d'autres structures similaires, peuvent contribuer à l'identification des items. Dans ces séries, plutôt que de rendre les items difficiles à identifier, le non marquage semble simplement indiquer que l'ordre entre les items n'est pas un critère d'organisation important. Par conséquent, il n'est pas souligné par le marquage explicite.

Le non marquage ne suffisant pas à lui seul pour rendre la série non parallèle, l'identification automatique des structures non parallèles se révèle très laborieuse avec nos méthodes. Lors de l'examen des marqueurs utilisés, à partir d'une analyse manuelle sur les structures introduites par une combinaison de marqueurs, nous avons cependant noté que les séries introduites par plusieurs types de marqueurs, dont au moins un connecteur, sont plutôt non parallèles. Ces séries incluent très fréquemment des changements du principe d'énumération au milieu de la série et des items autrement déséquilibrés. Ces structures couvrent 14 % des séries simples du corpus. Les séries introduites par des expressions anaphoriques et des constructions numéro+nom, en revanche, sont toutes plutôt parallèles malgré la combinaison de marqueurs.

Notre analyse avance donc que le non marquage ne rend pas les séries directement non parallèles mais que c'est le cas pour le changement du principe d'énumération et la différence structurelle des items, provoquée entre autres par leurs structures syntaxiques ou thématiques différentes. Pareillement, la variation importante de la longueur des items semble avoir ce même effet. Par ailleurs, notre analyse prouve également que les séries non parallèles ne sont pas, non plus, généralement difficiles à identifier ou incohérentes. Elles s'avèrent simplement atypiques en regard de leur définition répandue de structures composées d'items équivalents. Ce résultat est analogue au constat de Péry-Woodley (2000 : 139) selon lequel même les énumérations « déviantes » sont fonctionnellement acceptables.

7. MARQUAGE ET NON MARQUAGE DES SERIES LINEAIRES : FACTEURS TEXTUELS EXPLICATIFS

Dans la section précédente, nous avons montré que les séries linéaires sont des structures très variées : elles peuvent être introduites dans le texte de maintes manières et leur longueur affiche également des variations importantes. Dans cette section, nous analyserons le marquage des séries linéaires plus en détail, en privilégiant les facteurs susceptibles d'expliquer leur mode de marquage. Nous commencerons par examiner le degré d'exactitude de l'introduction des séries linéaires différentes et les contextes faisant varier ce degré d'exactitude. Pourquoi les marqueurs additifs, tels que *de plus* et *ensuite*, sont-ils très fréquents dans certaines séries tandis que dans d'autres les marqueurs exacts, tels que *premièrement* et *pour terminer*, sont seuls utilisés ? La seconde section proposera d'analyser le lien entre la longueur de la série et la façon dont elle est marquée. Outre la variation de la longueur des catégories de séries, cette analyse traitera du fonctionnement des types de marqueurs et des titres dans l'introduction des structures de longueurs différentes. Enfin, la dernière section portera sur les items non marqués et leur identification. Comment sont-ils identifiés ? La réponse à cette question pourra également révéler des raisons sous-jacentes de leur non marquage.

L'analyse de cette section repose sur les catégories de série présentées en détail dans la section précédente. Pour l'analyse de cette section, nous y avons pourtant apportées quelques modifications mineures. Dans la classification, le critère principal de distinction est la variation possible des types de marqueurs d'item et la présence possible d'items non marqués. Pour les séries simples, les séries *homogènes* (HOMO) sont composées d'items explicitement signalés par le même type de marqueurs tandis que, dans les séries *mixtes* (MIX), les types de marqueurs varient. Les séries *partiellement non marquées* (PNM) incluent des items non marqués et des items explicitement signalés. Les séries *non marquées* (NM) sont composées uniquement d'items non marqués. Les séries NMet diffèrent de ces dernières en ce qu'un des items est introduit par une conjonction de coordination. Les séries enchaînées sont classifiées en fonction de ces mêmes critères, selon les catégories d'appartenance des séries inférieures de ces structures.

De surcroît, pour l'analyse de cette section, nos résultats de la section précédentes nous a conduite à ajustifier les catégories analysées de deux manières. Tout d'abord, nos constats⁴³ sur la différence des séries mixtes introduites par des expressions anaphoriques et des constructions numéro+nom ainsi que celles signalées par une combinaison d'autres marqueurs nous ont encouragée à maintenir ces deux types de série à part. Ces deux sous-groupes sont intitulés respectivement MIXAL et MIXREST. La nature dispersée des séries

⁴³ Voir section 6.3, dans laquelle nous avons noté que la grande majorité des séries mixtes introduites par les expressions anaphoriques et les constructions numéro+nom sont plutôt parallèles tandis que les séries mixtes introduites par une combinaison d'autres marqueurs, dont au moins un connecteur, ont plusieurs traits typiques des séries non-parallèles.

enchaînées (cf. section 6.3.2) nous a conduite à simplifier leur classification. Pour ce faire, nous avons également fusionné les catégories de série NM et NMet composées respectivement d'items non marqués, d'une part, ainsi que d'items non marqués et de conjonctions de coordination, d'autre part. Enfin, nous avons également décidé de nous concentrer sur les catégories de série enchaînée les plus fréquentes. Ainsi, les catégories analysées et leurs pourcentages du total des séries linéaires enchaînées s'établissent comme suit : HOMO+NM (32 %), NM+PNM (18 %), NM+NM (17 %), MIX+NM (10 %) et HOMO+HOMO (8 %). En tout, ces séries recouvrent 85 % de toutes les séries enchaînées du corpus.

7.1 POUR UN GUIDAGE ADEQUAT DU LECTEUR : LE DEGRE D'EXACTITUDE DU MARQUAGE

Le marquage de l'organisation textuelle, dans notre cas les marqueurs d'item, les amorces et les clôtures, fonctionne comme un guide de l'organisation du texte pour le lecteur. Moins les séries linéaires sont explicitement introduites, plus le lecteur doit interpréter le texte afin de bien le comprendre. Outre la présence ou l'absence des marqueurs explicites, des amorces et des clôtures, la façon dont le marquage guide (ou non) le lecteur dans l'interprétation des séries linéaires dépend du degré d'exactitude des marqueurs utilisés. Les marqueurs *exacts*, tels que *premièrement* ou *le dernier aspect*, indiquent la place exacte des items dans la série. Les marqueurs *relatifs*, tels que les additifs *un autre exemple* ou *de plus*, signalent uniquement l'addition de l'item dans la série (cf. aussi Jackiewicz 2005 : 107). Nous rencontrons cette même distinction aux indices exacts et relatifs pour les amorces et les clôtures : tandis que les amorces et les clôtures exactes annoncent au lecteur le nombre d'items qu'il doit repérer dans le texte, les autres types laissent cette question ouverte. De même que les marqueurs d'item indiquant uniquement l'addition, nous nous référons à ces amorces et clôtures comme *marqueurs relatifs*.

Le rôle différent des marqueurs exacts et relatifs dans l'introduction des séries linéaires est surtout dû à la présence possible d'items non marqués. En exprimant la place de l'item, les marqueurs exacts indiquent au lecteur la présence possible des items non marqués dans la série. De la sorte, le lecteur ne doit pas la déduire lui-même. Les amorces et les clôtures exactes ont ce même effet. Les marqueurs relatifs, en revanche, nécessitent un effort d'interprétation plus important. En effet, le lecteur doit lui-même figurer le nombre final d'items à repérer.

Dans notre étude, les résultats indiquent que l'exactitude exprimée par les marqueurs d'item, amorces et clôtures varie selon la façon dont la série en question est autrement introduite. Les séries linéaires incluant des items non marqués sont plus souvent introduites par des marqueurs relatifs que celles dont tous les items sont explicitement signalés. La même variation est relevée pour les séries enchaînées, bien que leur marquage soit, dans l'ensemble, plus exact que celui des séries simples. À la suite, nous examinons ce

phénomène plus en détail en commençant par les amorces et les clôtures utilisées dans les catégories de série différentes et en continuant par l'examen des marqueurs d'item.

7.1.1 AMORCES ET CLOTURES

Ci-dessous, la Figure 15 présente les fréquences des amorces différentes dans chaque catégorie de série. Selon nos critères d'annotation, les séries non marquées doivent obligatoirement être précédées ou suivies d'une amorce ou clôture exacte. Elles sont donc absentes de la figure. Les types d'amorce sont présentés en détail en section 5.2.2 et dans la section précédente, dans le Tableau 5. La distinction la plus importante pour l'analyse du degré d'exactitude de cette section s'établit entre les amorces *exactes* (NUM) définissant le nombre d'items de la série et celles ne le définissant pas.

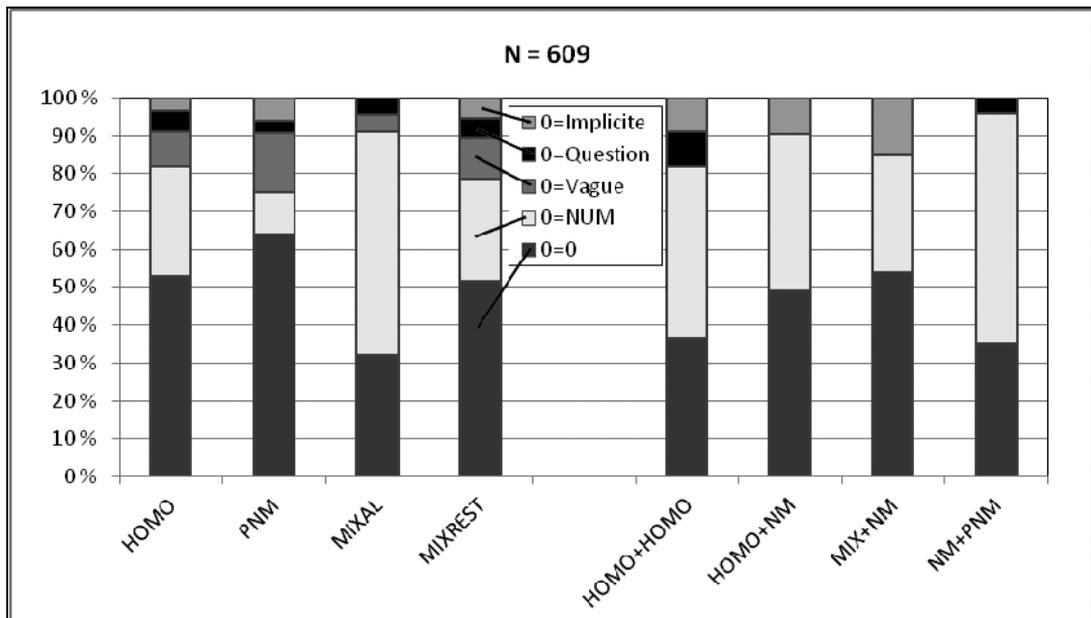


Figure 15 Amorces dans les catégories de série.

La Figure 15 indique, tout d'abord, que la présence des amorces varie selon les catégories de série. Les fréquences des séries sans aucun type d'amorce et de celles avec une amorce exacte indiquant le nombre d'items de la série ne sont notamment pas les mêmes. Pour les séries simples, les amorces séparent surtout les séries partiellement non marquées (PNM) des autres ; alors que près d'un tiers des séries homogènes (29 %) et des séries MIXREST (27 %) est également précédé d'une amorce indiquant le nombre d'items, ce chiffre n'est que 11 % pour les séries partiellement non marquées⁴⁴. De plus, la proportion des séries PNM sans aucune amorce est supérieure (64 %) à celle des séries homogènes (53 %) et des

⁴⁴ La différence de fréquence des amorces exactes dans les PNM et dans les HOMO et MIXREST est statistiquement significative avec le test X2, la valeur $p = 1,484e-06$, $X2 = 23,17$, $N = 499$.

séries MIXREST (51 %) ⁴⁵. Les MIXAL, en revanche, sont extrêmement souvent précédées d'une amorce exacte. Cette fréquence s'explique sans doute par les marqueurs utilisés dans ces séries ; les expressions anaphoriques nécessitent une zone de référence qui se réalise dans l'amorce.

Pour les séries enchaînées, nous avons déjà présenté à la section 6.3.2 certaines caractéristiques de l'ensemble de ces structures, l'une des plus importantes semblant être une certaine exactitude dans la façon dont elles sont signalées. Comparées aux séries simples, elles appartiennent plus souvent aux catégories de série homogènes et non marquées et sont plus fréquemment combinées à une amorce exacte. Cette exactitude transparait également dans la Figure 15. Les amorces exactes sont relativement fréquentes dans toutes les séries enchaînées, quoiqu'une différence semble s'établir entre les HOMO+HOMO et NM+PNM d'une part et les HOMO+NM et MIX+NM de l'autre.

En prouvant la rareté des amorces exactes dans les séries simples partiellement non marquées, la Figure 15 atteste déjà le lien entre le non marquage et les marqueurs relatifs. Une présentation plus générale de ce lien est illustrée dans la Figure 16, qui indique les proportions des séries précédées voire suivies d'une amorce ou d'une clôture exacte ⁴⁶. Ce mode de présentation optimise l'examen du degré d'exactitude surtout dans les séries MIXREST et dans les enchaînées, des catégories dont la nature dispersée complique l'analyse quantitative.

⁴⁵ La différence de fréquence des séries sans aucune amorce dans les PNM et dans les HOMO et MIXREST est statistiquement significative avec le X2 test, la valeur $p = 0,0147$, $X^2 = 5,96$, $N = 499$.

⁴⁶ Les fréquences des clôtures différentes dans les catégories de série sont relativement similaires. Pour cette raison, nous ne les présentons pas.

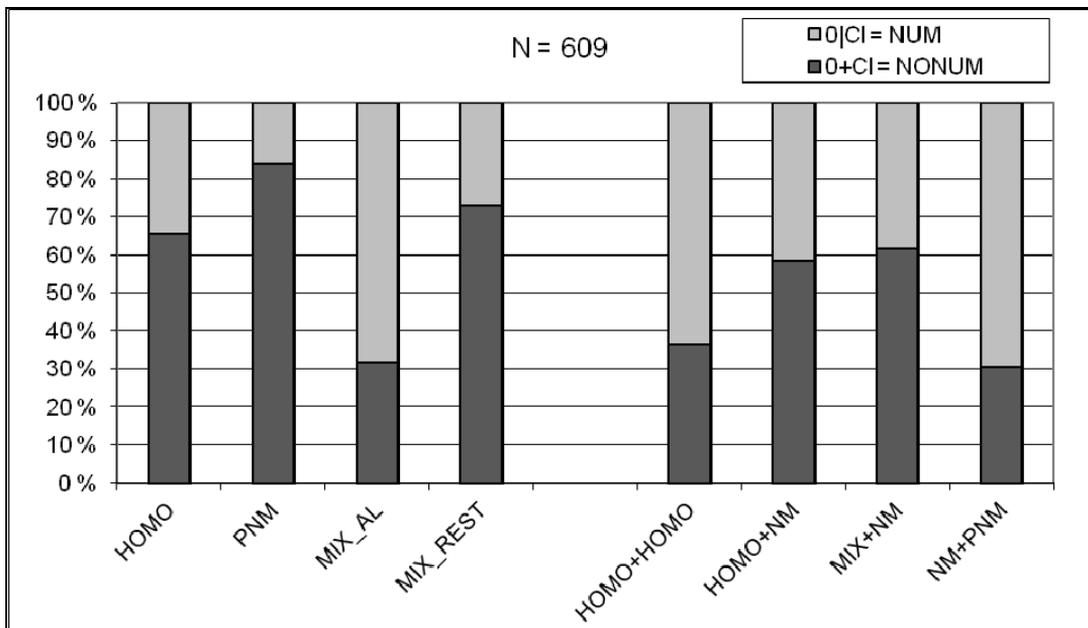


Figure 16 Amorces et clôtures exactes des catégories de série⁴⁷.

Pour les séries simples, la Figure 16 montre que la rareté des séries partiellement non marquées attachée à une amorce ou une clôture exacte devient encore plus évidente quand la présence des deux composants est examinée conjointement. De plus, elle révèle une différence invisible dans la Figure 15 ; les amorces et clôtures exactes semblent moins fréquentes dans les séries MIXREST que dans les séries homogènes. Dans les premières, un composant, soit l'amorce soit la clôture exacte, est présent dans 27 % des séries tandis que 34 % des séries homogènes sont combinées à l'un des deux. La différence comparée de ces deux catégories de série n'est pourtant pas statistiquement significative⁴⁸, ce qui peut être imputable au nombre limité d'occurrences de MIXREST (37). Rapportée à l'ensemble des séries PNM, MIXREST et homogènes, cette différence peut cependant être prouvée⁴⁹. Outre cet écart entre les catégories de séries simples, les deux groupes de séries enchaînées deviennent aussi plus aisément discernables dans la Figure 16 ; les séries HOMO+HOMO et NM+PNM sont plus fréquemment combinées à un composant indiquant leur nombre d'items que les séries HOMO+NM et MIX+NM⁵⁰.

Pour conclure nos résultats sur les séries simples, il semble que les amorces et clôtures n'y soient pas utilisés dans les séries présentant un marquage d'items partiel, afin de remplacer ce marquage manquant. En fait, la présence des amorces et clôtures suit plutôt la tendance générale de marquage de cette catégorie de série. Ainsi, dans les séries homogènes et

⁴⁷ NUM indique que le nombre d'items de la série est signalé au moins par une amorce ou une clôture et NONUM que ce n'est pas le cas.

⁴⁸ Avec le X2-test, la valeur $p = 0,495$, $X^2 = 0,46$, $N = 211$.

⁴⁹ Avec le X2-test, la valeur $p = 3,372e-05$, $X^2 = 17,19$, $N = 499$.

⁵⁰ Avec le X2-test, la valeur $p = 0,0251$, $X^2 = 5,02$, $N = 88$.

MIXREST, les amorces et clôtures répètent souvent la façon dont le segment est organisé, bien que les items eux-mêmes y soient déjà explicitement introduits. En revanche, dans les séries PNM, déjà moins explicitement signalées du fait de la présence d'items non marqués, les amorces et clôtures exactes indiquant au lecteur le nombre d'items sont moins fréquentes. Ces séries sont également laissées sans aucun type d'amorce plus souvent que celles dont tous les items sont explicitement signalés. L'Exemple 38 ci-dessous illustre une série dont le nombre de composants est signalé avant elle, bien que les items soient aussi explicitement marqués. La série commence par l'amorce annonçant le principe d'énumération ainsi que le nombre d'items de la série : *trois attitudes*. Les trois items suivent ce principe et sont, de plus, introduits par des connecteurs très typiques : *premièrement - ensuite - enfin*.

Exemple 38

- 0 Pour ce faire, **trois attitudes** sont au minimum nécessaires
- 1 L'enseignant doit **premièrement** partir des intérêts profonds de l'enfant et le laisser agir. Car comme l'affirme Cousinet (1949, p.21), « le premier contact des enfants avec l'enseignement est une inhibition. À ces êtres qui ne vivent que pour agir, et qui ne peuvent vivre qu'à condition d'agir, l'action est interdite ».
- 2 Il doit **ensuite** considérer l'enfant comme un être original et singulier, capable de construire son propre rapport au savoir. De la sorte, « la logique de l'apprentissage, n'est nullement la logique du programme. Non seulement l'apprentissage n'a pas besoin de ces artifices mais il en meurt. Pas d'horaires fixes par matière ou par sujet, pas de disciplines closes, pas de notes, pas de classements, mais la liberté, seulement la liberté d'apprendre ».
- 3 **Enfin**, il est nécessaire de réinvestir le milieu de vie de l'enfant dans ses apprentissages. L'enseignement classique, longuement observé par Cousinet lorsqu'il était inspecteur de l'Education Nationale, ne permet pas la rencontre de l'école et de la vie.

Vialle

Pour les séries enchaînées, en revanche, la distinction des catégories en deux groupes, illustrée dans la Figure 16, nécessite des examens supplémentaires. Pour ce faire, nous les avons analysées manuellement. Cet examen montre que c'est surtout le risque de répétition qui explique la rareté des amorces et clôtures exactes dans les HOMO+NM et MIX+NM. Dans les HOMO+HOMO et NM+PNM plus fréquemment attachées à ces composants, la nature des marqueurs et la structuration interne des items ont pour effet que l'effet répétitif ne s'y reproduit pas.

Dans les séries HOMO+NM et MIX+NM, les items non marqués servent souvent de points de référence pour ceux de l'autre série. Pour cette raison, comme nous l'avons exposé dans la section précédente, les anaphoriques sont très fréquents dans ces séries. De plus, les autres marqueurs, tels que les cadratifs, sont aussi fréquemment utilisés dans ces séries pour se référer à l'item de la série précédente. Dans l'Exemple 39 suivant, notamment, les items explicitement signalés par des adverbiaux à l'initiale de la phrase se réfèrent aux items non marqués de la première série. Ces derniers sont, en fait, uniquement des exemples numérotés de 34 à 36 par l'auteur de l'article.

Exemple 39 Les italiques ainsi que la numérotation des items sont de la version originale. Les caractères en gras, en revanche, sont de nous.

0 On peut citer les exemples **suivants** :

- 1 (34) *Une décision fut prise : dès son départ, il nous faudra nous inscrire au collège Mariette, et maman devra trouver une famille pour nous héberger.*
Boris Schreiber, *Un silence d'environ une demi-heure*, p. 228
- 2 (35) *Boria, tu n'as pas honte ? Pour Alexandre, et pour la guerre, et pour les morts ? Les atrocités ? Tout ça, ça ne compte pas pour toi ?*
Boris Schreiber, *Un silence d'environ une demi-heure*, p. 460
- 3 (36) *Maman à la fois approuvait et n'osait pas citer Riga, et craignait que Wladimir ne remarquât qu'elle n'osait pas.*
Boris Schreiber, *Un silence d'environ une demi-heure*, p. 702

1 **Dans le premier exemple**, il y a bien deux choses différentes à faire et non un ensemble cohérent de ces deux choses.

2 **Dans le second**, le *tout ça* est le résultat du décompte et donc de l'accumulation des *atrocités*. La répétition de *et* (ainsi que de la formule avec *pour*) rend plus sensible la valeur additive que la combinaison avec la virgule lui donne, quand la suppression de la virgule ici désautonomiserait les éléments conjoints en les intégrant dans un ensemble isotopique.

3 **Dans le troisième cas**, *à la fois* est structurellement ambigu mais porte en fait sur le second *et* : le contraste entre la continuité isotopique de *approuvait et n'osait pas* et la rupture entre cet élément et *craignait que...* est remarquable.

Cl. Si **ces analyses** vont dans le bon sens, je crois qu'on pourrait en tirer une forme de généralisation.

Lambert

En principe, le fait de fonctionner comme point de référence pour l'autre série ne saurait expliquer l'absence de l'amorce ou clôture exacte. Dans la pratique, pourtant, notre analyse manuelle montre que les items formant ces points de référence sont souvent très bien délimités et peuvent ainsi se passer d'amorce ou de clôture tout en demeurant facilement repérables. De plus, la présence d'un tel composant pourrait avoir un effet répétitif. Par exemple, dans les séries typiques sans titres, les items non marqués peuvent former des exemples numérotés, comme dans l'exemple ci-dessus. Ces items sont en général faciles à repérer, même sans marquage. Dans les séries méta HOMO+NM et MIX+NM, composées de deux séries inférieures, dont la seconde consiste en items-sections délimités et signalés par des titres, ce sont très majoritairement les parties non marquées qui correspondent aux sections titrées⁵¹.

L'Exemple 40 illustre une autre série enchaînée HOMO+NM dont les items non marqués sont facilement repérables. Dans cette série, les items sont extrêmement courts et correspondent aux mots individuels ou syntagmes nominaux. La première série, non marquée, comprend une seule phrase et les items énumérés correspondent en fait aux

⁵¹ Pour un exemple, voir l'Exemple 24 à la page 117.

adverbes *directement* et *indirectement*. Les items de la seconde série s'y réfèrent par des expressions cadratives.

Exemple 40

- 1 Ces angles d'analyse conduisent à une interprétation directement
 - 2 **ou**, le plus souvent, indirectement, en appui d'autres faits contextuels.
 - 1 **Dans le premier cas**, ils ont pour nous une fonction de « marqueur » (comme la présence de certains mots ou de certaines structures),
 - 2 **et dans le second** d'indice, ou de guide pour l'interprétation.
- Cl. Nous réservons les dernières parties pour discuter **ces points**, en nous centrant sur quelques exemples significatifs.

Lecolle

Comme pour la plupart des séries HOMO+NM et MIX+NM, l'Exemple 40 n'est ni précédé, ni suivi d'une amorce ou clôture exacte définissant son nombre d'items. En lisant la série, il nous paraît qu'il s'agirait d'éviter une redondance. Cet effet répétitif semblerait dû à la faible longueur et au repérage aisé des items, qui selon notre analyse semblent typiques de ces séries. L'évitement de la répétition se justifie par la maxime de quantité de Grice (1975) ; le lecteur doit être informé autant que nécessaire mais pas plus.

Pour les HOMO+HOMO et NM+PNM, en revanche, l'effet répétitif provoqué par l'amorce ou la clôture exacte ne semble pas se reproduire. Une des motivations pour son absence paraît être la longueur supérieure de ces séries. Cet aspect ne sera pourtant examiné que dans la section suivante. Une autre explication pourrait être le marquage de ces structures. En effet, notre analyse manuelle prouve que les connecteurs du type *d'une part – d'autre part* et *d'un côté – de l'autre* sont très fréquents dans les HOMO+HOMO. Bien que nous les ayons annotés comme exprimant un certain ordre entre les items et que Turco et Coltier (1988 : 58) indiquent que les MIL (dont *d'une part – de l'autre*) « perdent leur vocation sémantique primitive pour désigner la succession des constituants », l'ordre exprimé par ces marqueurs n'est peut-être pas aussi évident que celui signalé par les autres MIL ou des anaphoriques. Par conséquent, il se peut que l'effet répétitif des amorces et clôtures ne se reproduise pas une fois combinées à ces marqueurs.

L'Exemple 41 ci-dessous illustre les marqueurs fréquents dans les HOMO+HOMO et le fonctionnement de l'amorce / clôture exacte au milieu de la série. La série débute par une amorce implicite, indiquant que le texte à venir formera *une hypothèse*. Les items de la première série, relativement longs, sont signalés par *d'une part - d'autre part* et ceux de la seconde par les anaphoriques s'y référant. De plus, une amorce / clôture entre les deux séries répète le nombre d'items des séries : *ces deux types d'écoles*.

Exemple 41

- 0 Pour notre part, nous soumettons l'hypothèse **suivante** :
- 1 se pourrait-il que sous l'effet combiné de la décentralisation et de la concurrence, notre système éducatif primaire secondaire est en voie de se « hiérarchiser », comprenant **d'une part** des écoles ou des secteurs d'écoles, publiques et privées, performants et capables de produire les élèves que la demande sociale et les niveaux supérieurs du système réclament,
 - 2 **et d' autre part**, des écoles, surtout publiques, ou de grands pans d'écoles, regroupant des élèves faibles, en difficulté ou à risques ? Dans ces dernières écoles, les enseignants et le personnel professionnel et technique, partiellement soutenus par des programmes comme l'École Montréalaise (MÉQ, 1997b), gèrent au mieux un ensemble de problèmes sociaux plus larges que l'école. Ils font de leur mieux et même un peu plus, ce que le palmarès est incapable de prendre en compte, d'où leur frustration et le sentiment très répandu parmi eux de déconsidération sociale.
- 0/Cl. Entre ces **deux** types d'écoles,
- 1 entre **les premières** qui « instruisent »
 - 2 **et les secondes** qui tentent d'« éduquer », l'observation révèle l'existence d'écoles aspirant à monter dans la première catégorie, alors que d'autres luttent pour éviter la chute dans la dernière catégorie.
- Cl. **Ce schéma** est ici esquissé à gros traits, mais il exprime une logique d'évolution qui est forte.

Lessard

Selon notre analyse, l'amorce / clôture au milieu de l'Exemple 41 est dépourvue d'effet répétitif tandis que, si elle était placée dans l'Exemple 40, elle risquerait d'être répétitive. Ainsi, la nature des marqueurs, la longueur des items et la structuration des items non marqués sembleraient toutes contribuer à l'absence / présence des amorces et clôtures exactes dans les séries enchaînées.

Enfin, la nature des autres marqueurs et la longueur de la série sembleraient expliquer la fréquence des amorces et clôtures exactes dans les NM+PNM aussi. Comme nous l'avons constaté, la structure des séries enchaînées étant plutôt complexe, elles sont globalement signalées de façon exacte et explicite afin de préserver leur lisibilité (cf. section 6.3). Les marqueurs d'items proprement dits sont cependant plutôt rares dans les séries NM+PNM. Souvent, il s'agit d'un MIL comme *enfin* à la fin de la série. Pour cette raison, nous pouvons supposer que les amorces et clôtures exactes peuvent s'avérer nécessaires afin que la série, peu explicitement signalée par ailleurs, demeure cohérente. L'Exemple 42 ci-dessous présente une série NM+PNM dont le nombre d'items est annoncé par une amorce. Autrement, les seuls marqueurs explicites sont la conjonction *et* dans le dernier item de la première et le connecteur *enfin* dans le dernier item de la seconde série.

Exemple 42

0 Les « remèdes » à ces « anomalies sentimentales » sont de **trois ordres** différents :

1 physiques,

2 intellectuels

3 **et** psychiques.

1 Un examen médical pourra révéler, par exemple, un mauvais fonctionnement des glandes endocrines, un dérèglement du système nerveux, [...].

2 Entre autres remèdes intellectuels, l'ouvrage propose d'épargner aux enfants « la vue de tout désordre », [...].

3 **Enfin**, les remèdes psychiques tiennent essentiellement dans le contrôle de l'imagination. [...]

Hudon

Pour conclure, notre examen des amorces et des clôtures montre donc que leur usage dépend de la signalisation alternative de la série. Dans les séries simples, elles ne sont pas utilisées afin de remplacer le marquage manquant des items. Elles servent plutôt à souligner l'organisation des séries dont tous les items sont déjà signalés par des marqueurs explicites. Dans les séries incluant des items non marqués, le nombre d'items est très rarement signalé par une amorce ou clôture. Par ailleurs, les autres types de ces composants s'y font plus rares.

Comparées aux séries simples, toutes les séries enchaînées sont signalées dans le texte de façon explicite et exacte. En outre, les amorces et clôtures exactes y sont plus fréquentes dans toutes les catégories. Pour autant, notre analyse a relevé certaines caractéristiques qui influencent la présence des amorces et des clôtures dans ces séries. Il s'agit essentiellement d'un risque d'effet répétitif, possiblement lié à la nature des marqueurs explicites utilisés, à la structuration des items non marqués ainsi qu'à leur faible longueur. Par exemple, si les items non marqués correspondent uniquement aux mots individuels d'une même phrase ou sont délimités par des titres, les amorces et clôtures peuvent s'avérer inutiles.

7.1.2 MARQUEURS D'ITEM

La section précédente sur l'usage des amorces et clôtures vient de montrer que, dans les séries simples, le marquage exact du nombre d'items de la série et ainsi de son organisation est plus fréquent dans les séries dont tous les items sont explicitement signalés que dans celles présentant des items non marqués. Dans les séries enchaînées, en revanche, la précision du nombre d'items par une amorce ou une clôture est fréquente dans toutes les structures, bien que la faible longueur et le repérage facile des items puissent aussi motiver leur absence. Ces tendances de marquage transparaissent également à travers le degré d'exactitude des marqueurs d'item utilisés dans les catégories de série. À la suite, nous examinerons cette problématique plus en détail.

Ci-dessous, la Figure 17 et la Figure 18 présentent les degrés d'exactitude exprimés par les marqueurs d'items des différentes catégories de série. *Exact* indique que les marqueurs d'item exacts, tels que *premièrement* ou *enfin*, sont seuls utilisés tandis que *relatif* signale

que la série inclut exclusivement des marqueurs relatifs, tels que *de plus* ou *un autre* + *NOM*. *Exact+relatif* indique que les items sont signalés par les deux types de marqueurs.

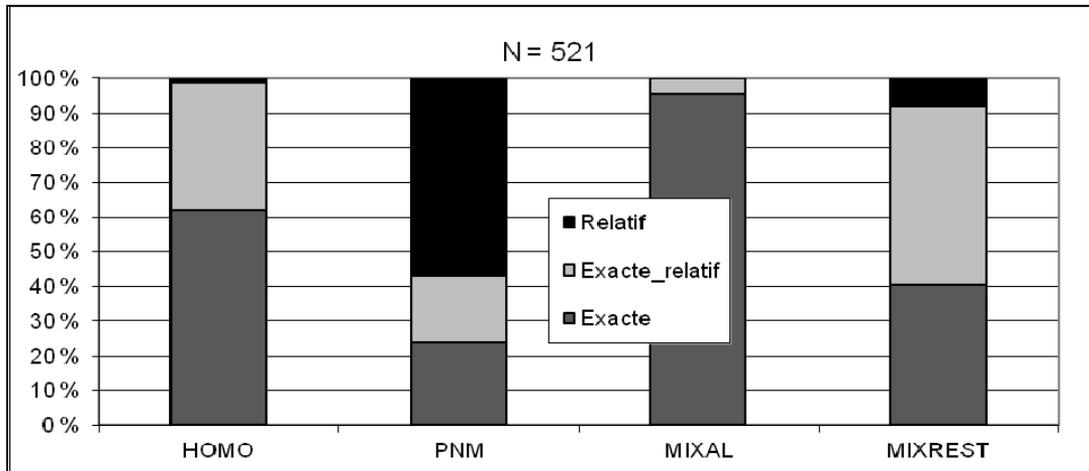


Figure 17 Degré d'exactitude des marqueurs des séries simples.

Pour les séries simples, la Figure 17 indique que les catégories de série se caractérisent par des degrés d'exactitude différents. Les marqueurs exacts sont les seuls utilisés dans presque toutes les séries MIXAL (95 %) et dans la majorité des séries homogènes (62 %). Les marqueurs exacts sont également souvent utilisés comme marqueurs uniques des MIXREST (41 %), bien que moins fréquemment que dans les homogènes⁵². En outre, la majorité (51 %) des MIXREST est signalée par la combinaison de marqueurs exacts et relatifs. Les marqueurs relatifs, en revanche, sont les seuls dans la majorité (57 %) des séries PNM.

⁵² Comme le nombre d'occurrences des séries uniquement signalées par des marqueurs relatifs est très bas, la comparaison statistique du degré d'exactitude du marquage des séries homogènes et MIXREST pose quelques problèmes. Une différence statistiquement significative s'établit pourtant entre les séries dans lesquelles les marqueurs exacts sont seuls utilisés et les autres ; les homogènes sembleraient être plus souvent signalées uniquement par des marqueurs exacts que les MIXREST. Avec le X2-test, la valeur $p = 0,0258$, $X^2 = 4,96$, $N = 211$.

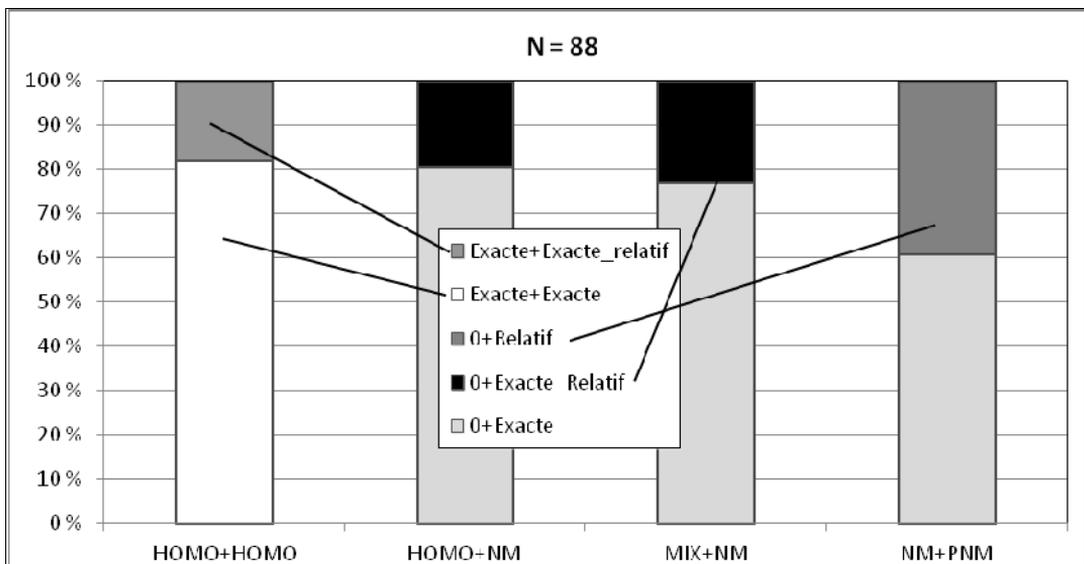


Figure 18 Degré d'exactitude des marqueurs des séries enchaînées.

En ce qui concerne les séries enchaînées, la Figure 18 montre que les marqueurs d'item exacts sont extrêmement fréquents dans toutes les catégories. Comme pour les séries simples, les marqueurs relatifs sont cependant souvent utilisés en tant que marqueurs uniques dans les NM+PNM. Un élément en faveur de notre hypothèse relative au besoin d'amorces et clôtures exactes dans cette catégorie. Comme les items de ces séries sont majoritairement non marqués et les marqueurs d'items relatifs sont fréquents, il se peut que la structuration des items proprement dits ne suffise pas pour rendre le texte cohérent. Ainsi, les amorces et clôtures exactes y seraient nécessaires pour assurer un guidage minimal du lecteur à travers une structure possiblement difficile à interpréter.

Bref, deux aspects des fréquences présentées dans la Figure 17 et la Figure 18 appellent un commentaire. Premièrement, la rareté des catégories autres que les PNM signalées uniquement par des marqueurs relatifs s'explique par la nature de ces marqueurs. Comme ils signalent généralement l'ajout d'un item à la série déjà existante, il est très rare qu'ils puissent introduire le premier item de la série. Par conséquent, les quelques séries homogènes et mixtes uniquement introduites par des marqueurs relatifs dans notre corpus sont des exceptions marginales dans lesquelles le premier item est signalé, entre autres, par *outré*. La fréquence extrême des séries uniquement signalées par des marqueurs exacts dans les séries MIXAL s'explique également par une raison similaire. La grande majorité de ces séries est introduite par les anaphoriques *le premier – le deuxième*, dont l'une est toujours combinée à un nom et annotée comme construction numéro+nom. La seule série MIXAL signalée par la combinaison de marqueurs exact et relatif est introduite par *la première – la caricature suivante*. La rareté extrême de ce type de combinaisons est due à la rareté des marqueurs relatifs annotés comme anaphoriques ou construction Num+N. Le

plus fréquent de ces marqueurs est sans doute la construction *autre* + *NOM*⁵³. Dans notre corpus, pourtant, ce marqueur n'est jamais utilisé conjointement avec une expression anaphorique⁵⁴. Par conséquent, les séries signalées par une expression anaphorique et une construction Num+N dont un marqueur serait de nature relative sont extrêmement rares.

Bien que les marqueurs relatifs soient difficiles à employer comme marqueurs uniques dans les autres catégories, rien n'empêche l'emploi de marqueurs exacts dans les séries partiellement non marquées. Selon notre analyse, cet usage est simplement moins fréquent. Par exemple, dans l'Exemple 43 ci-dessous, le quatrième item est signalé par le MIL exact *enfin*, soulignant la position finale de l'item. La série est également précédée d'une amorce exacte, *quatre degrés*, annonçant le nombre et le principe d'énumération des items ; chaque item correspond à un degré de tonalité, énumérés du plus positif au plus négatif.

Exemple 43

- 0 Nous avons retenu **quatre degrés** que nous allons illustrer à l'aide d'exemples.
- 1 La « tonalité très positive » mobilise des termes qui « renforcent » une tonalité déjà positive (très, beaucoup, important, etc.) : « Beaucoup de maturité, de réflexion et de concentration. C'est très bien ».
- 2 La « tonalité positive » ne recourt pas à ces renforcements : « Élève appliqué qui s'investit dans les activités de la classe ».
- 3 La « tonalité mitigée » introduit une réserve négative dans une appréciation au départ positive : « Élève consciencieuse et appliquée, Z. est bien entrée dans les apprentissages. Un peu moins dans l'esprit collectif d'une classe ».
- 4 **Enfin**, la « tonalité négative » est univoque : « S. est incapable de respecter les règles édictées par l'adulte. Incapable de concentration. Il lui est donc très difficile de rentrer dans les apprentissages ».

Marcel

La rareté des marqueurs exacts dans les séries PNM n'a donc pas de raison pratique. De même, les combinaisons de marqueurs exacts et relatifs pourraient être plus utilisées dans les séries HOMO et les séries enchaînées, et les marqueurs exacts être plus fréquents dans les MIXREST. Comme ce n'est pas le cas, nos résultats sur le degré d'exactitude du marquage sont très similaires à nos conclusions de la section précédente relative à l'usage des amorces et clôtures dans les séries différentes ; les séries différentes tendent simplement à guider différemment le lecteur. Ainsi, les séries affichant des marqueurs d'item exacts fréquents sont également souvent combinées à une amorce ou une clôture exacte. En revanche, les séries avec des marqueurs d'item relatif le sont plus rarement.

⁵³ Sans être combiné à un nom, *autre* signale (dans notre corpus) le dernier item de la série et est ainsi annoté comme exact.

⁵⁴ L'usage discursif de *autre* et *autre* + *NOM* et leur combinaison avec l'anaphorique *premier* sont discutés dans différents travaux (cf. Berrendonner & Reicher-Béguelin 1996 ; Van Peteghem 2000 ; Schnedecker 2006). Quoique la combinaison *le premier* – *l'autre* semble possible (Schnedecker 2006 :139 et Berrendonner & Reicher-Béguelin 1996 : 7), la combinaison *le premier* – *l'autre*+*NOM* n'a pas été mentionnée.

7.1.3 BILAN : STRATEGIES DE MARQUAGE D'ORGANISATION TEXTUELLE EXACT ET VAGUE

Cette section s'est proposé d'examiner le degré d'exactitude du marquage utilisé dans les séries linéaires différentes. L'analyse montre qu'il varie selon la façon dont la série est autrement introduite. Ainsi, dans les séries simples homogènes, composées d'items explicitement signalés par le même type de marqueurs, les items sont très fréquemment introduits par des marqueurs exacts. D'ailleurs, leur nombre est souvent encore souligné par une amorce et / ou clôture. En revanche, les marqueurs d'item exacts sont moins fréquents dans les séries MIXREST, composées d'items explicitement signalés mais par des marqueurs différents. Aussi, les amorces et clôtures exactes semblent moins fréquentes dans cette catégorie de série. Les séries incluant des items non marqués, enfin, sont rarement combinées à une amorce ou une clôture exacte, et les marqueurs d'item relatifs y sont très majoritaires.

Comparées aux séries simples, les séries enchaînées sont introduites de façon explicite et exacte ; les séries partiellement non marquées sont relativement rares et les marqueurs d'item de même que les amorces et les clôtures exacts sont fréquents dans toutes les catégories. Toutefois, dans certaines catégories dont la faible longueur et le repérage facile des items risquent de causer un effet répétitif, ces composants sont moins employés. De plus, les marqueurs d'item relatifs sont fréquents dans les séries NM+PNM de même que dans leurs variantes simples.

Nos résultats sur le marquage des séries linéaires de catégories différentes indiquent donc que certaines structures recherchent simplement une organisation textuelle moins explicite et exacte que d'autres. Surtout, les séries linéaires très typiques, explicitement signalées par le même type marqueurs, sont très exactes tandis que les séries aux traits plus atypiques, souvent associés au non parallélisme, tels que le non marquage partiel et l'usage des types de marqueurs différents, sont plus souvent signalées par des marqueurs relatifs. Comme le degré d'exactitude de signalisation des séries varie, nous pourrions avancer qu'elles peuvent réaliser des stratégies textuelles différentes⁵⁵. Autrement dit, l'auteur peut choisir d'utiliser des séries différentes afin que l'organisation du texte corresponde le mieux son but communicatif. Cela va dans le sens de nos notes de la section précédente sur les séries avec des items non marqués ; elles ne sont pas nécessairement mal formées mais plutôt le résultat du choix de l'auteur.

⁵⁵ Voir notre discussion en section 4.1, Virtanen (1992 : 51) ; Enkvist (1975 : 18-22, 1987).

7.2 LONGUEUR ET MARQUAGE DES SERIES LINEAIRES

Selon notre analyse, la longueur du segment textuel à organiser est un deuxième facteur influençant la façon dont les séries linéaires sont introduites dans le texte. Lors de la description générale du corpus, nous avons montré que les séries linéaires simples et enchaînées organisent des segments très courts et très longs. De plus, les travaux antérieurs sur les indices de cohésion différents indiquent que la capacité organisatrice des marques différentes varie. Par exemple, selon Schnedecker (2000 : 22), les sources des anaphoriques *l'un – l'autre* et *le premier – le second* se trouvent à quelques propositions de leurs corrélats. Ho Dac (2007) note aussi que les connecteurs sont généralement utilisés pour organiser des segments d'un ou deux paragraphes au maximum. Jackiewicz (2005 : 98) offre un aspect du marquage des séries linéaires des longueurs différentes en constatant que « plus une série est longue et / ou plus ses items sont de taille importante, plus le signalement discursif est saillant (y compris visuellement) et régulier ». Le modèle de Goutsos (1996) distingue, de même, les indices plutôt locaux, servant la stratégie de continuation, et les indices globaux, servant le déplacement. Selon ce modèle, par exemple, le pronom anaphorique *il* et les ellipses fonctionnent au niveau local alors que les reprises nominales, telles que *ces items*, fonctionnent au niveau global.

La théorie d'accessibilité d'Ariel (1990, 2001⁵⁶ 57) procure une autre perspective à l'examen du lien entre les marqueurs et la longueur du segment textuel qu'il introduit. Selon cette théorie fondée sur les processus mentaux du lecteur, les expressions référentielles se caractérisent par des degrés d'accessibilité mentale différents ; tandis que certaines sont plus facilement repérables et utilisées dans des contextes univoques, d'autres sont utilisées dans des contextes plus ambigus. La distance entre l'expression référentielle et sa source dans le texte constitue l'un des aspects influençant l'ambiguïté du contexte et, donc, le choix du marqueur. Les anaphoriques pronominaux, par exemple, sont retenus pour une distance plutôt courte et les syntagmes nominaux et les noms propres lorsqu'elle s'accroît. Ces résultats sont donc similaires à ceux de Schnedecker (ci-dessus) tout en pouvant également justifier l'usage et le fonctionnement des indices complémentaires dans les séries.

Cette section examinera comment le marquage des séries linéaires varie selon la longueur. La première partie de la section l'étudiera à partir des catégories de série, en analysant l'influence de la longueur du segment à organiser sur l'exactitude du marquage et le marquage explicite. Nos résultats indiquent que le marquage des séries plus courtes est plutôt explicite et exact et devient plus implicite et imprécis pour les segments plus longs. Nous commencerons par les séries longues et continuons par les courtes. Ensuite, dans la troisième section, nous nous étudierons les types de marqueurs et les marqueurs individuels tout en examinant leur fonctionnement dans l'introduction des items de

⁵⁶ Appliquée en français dans Suomela-Salmi (1997).

⁵⁷ Un autre modèle avec une échelle similaire mais une théorie de base différente est celui de Gundel *et al.* (1993), voir aussi Chafe (1976, 1994) et Givon (1983).

longueurs différentes. Enfin, la dernière section sera consacrée aux séries section et méta, exclues de l'analyse des trois premières sections. La longueur de ces séries étant très supérieure à celle des autres, leur inclusion dans celles-ci risquerait de déformer les résultats. De plus, le rôle des titres dans l'introduction des séries section mérite une analyse séparée.

7.2.1 LONGUE ET IMPRECISE...

La Figure 19 ci-dessous présente les longueurs des catégories de série différentes. En particulier, la figure montre que les séries des types de marqueurs différents et les partiellement non marquées (MIXREST, PNM, NM+PNM) organisent des segments plus longs alors que les séries non marquées et celles dont tous les items sont explicitement signalés par le même type de marqueurs (NM, NMet, HOMO) ainsi que leurs variantes enchaînées structurent des segments plus courts⁵⁸. Nous constatons ainsi un lien entre la façon dont les séries linéaires sont marquées et leur longueur.

Pour les séries simples, la longueur supérieure des séries PNM et MIXREST au regard des séries homogènes est statistiquement significative⁵⁹. Il est aussi démontré que les séries NMet sont plus courtes que les NM⁶⁰. La longueur des séries MIXAL, enfin, semble se situer entre les homogènes et les partiellement non marquées mais, du fait de leur nombre limité d'occurrences, leur analyse statistique est ardue. Pour les séries enchaînées, la longueur très supérieure des NM+PNM par rapport aux autres catégories de série enchaînée analysées est également statistiquement significative⁶¹. Or, la faible longueur des NM+NM, HOMO+NM et MIX+NM face aux HOMO+HOMO et NM+PNM est tout aussi réelle⁶².

⁵⁸ Comme nous l'avons décrit dans 6.3, les catégories de série sont formées à partir de la présence des items non marqués et des types de marqueurs différents. Les homogènes (HOMO) se composent d'items explicitement signalés par le même type de marqueurs. Dans les MIXREST, les types de marqueurs varient et au moins l'un des items est signalé par un connecteur. Pour les MIXAL, les items sont introduits par des expressions anaphoriques et par des constructions Num+N alors que, dans les PNM, au moins l'un des items est non marqué. Dans les NM, enfin, tous les items sont non marqués et leur nombre signalé par une amorce et / ou clôture. Les NMet leur sont similaires mais incluent également une ou plusieurs conjonctions de coordination.

⁵⁹ Entre les homogènes et les PNM, avec le Kruskal Wallis, la valeur $p = 1,134e-06$, $H(X2) = 23,69$, $N = 443$.

Entre les homogènes et MIXREST, avec le Kruskal Wallis, la valeur $p = 0,0081$, $H(X2) = 7,01$, $N = 206$.

⁶⁰ La différence de longueur est statistiquement significative avec le Kruskal Wallis, la valeur $p = 0,0101$, $H(X2) = 6,62$, $N = 96$.

⁶¹ Avec le Kruskal Wallis, la valeur $p = 2,903e-05$, $H(X2) = 17,48$, $N = 82$.

⁶² Avec le Kruskal Wallis, la valeur $p = 7,313e-06$, $H(X2) = 20,11$, $N = 82$.

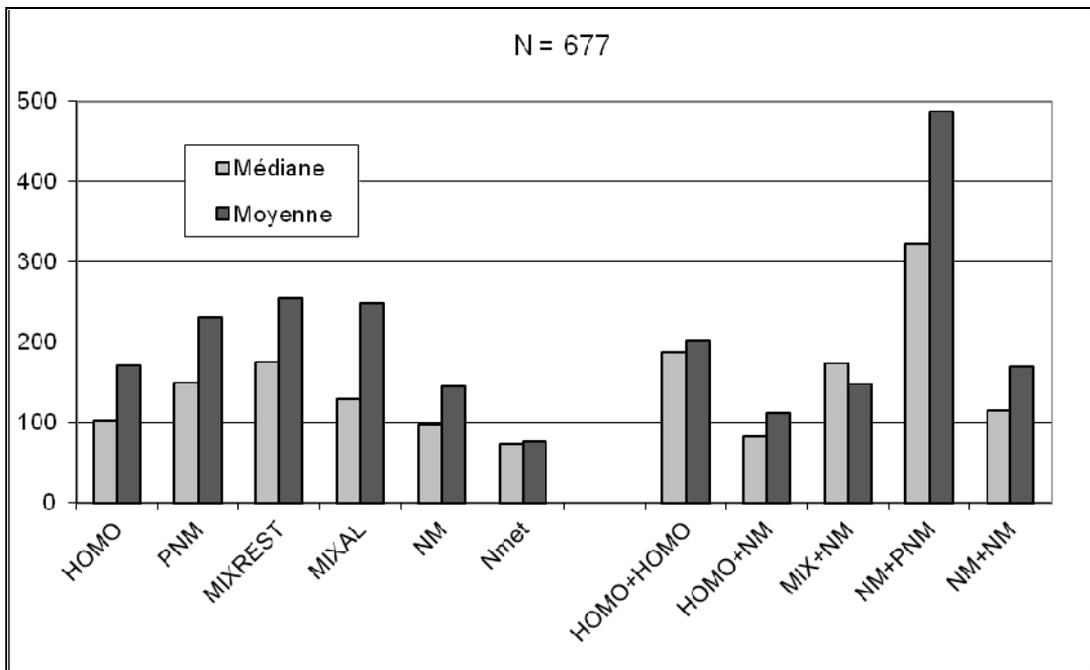


Figure 19 Longueurs des catégories de série, séries section et méta exclues.

Pour les séries enchaînées, la Figure 19 soutient nos deux hypothèses, avancées à la section précédente et relatives à la fréquence des amorces et clôtures dans les catégories différentes. Premièrement, la longueur attestée des NM+PNM corrobore notre hypothèse de nécessité des amorces et clôtures dans ces séries ; étant donnée leur complexité interne, leur marquage plutôt implicite et imprécis et leur longueur élevée, les amorces et clôtures peuvent s'avérer utiles pour garantir une lecture facile et cohérente. Deuxièmement, la Figure 19 prouve que les séries HOMO+NM et MIX+NM sont plus courtes que les autres. Ce résultat étaye notre résultat sur l'absence des amorces et clôtures exactes, expliqué par le risque d'un effet répétitif.

Le résultat le plus intéressant de la Figure 19 reste cependant la longueur plus élevée des séries PNM, MIXREST et NM+PNM. Dans la section précédente, nous avons montré que ces catégories de série sont souvent signalées par des marqueurs relatifs et que les amorces et clôtures exactes s'y montrent plus rares que dans les séries dont tous les items sont introduits par le même type de marqueurs⁶³. Ainsi, les résultats semblent proposer que le marquage de l'organisation textuelle, ou au moins des séries linéaires, est moins exact et moins explicite dans les séries longues que dans les courtes.

Le lien entre la longueur élevée et l'absence des marqueurs exacts mérite pourtant une analyse plus détaillée. Ci-dessous, la Figure 20 et la Figure 21 illustrent la variation de la

⁶³ Comme expliqué à la page précédente, la fréquence des amorces et des clôtures dans les NM+PNM forme une exception à cette tendance. Les marqueurs d'item relatifs y sont tout de même fréquents (cf. section précédente).

longueur selon le degré d'exactitude du marquage dans les catégories de séries simples différentes. Ces figures proposent que les séries introduites par des marqueurs d'items voire des amorces ou des clôtures exactes sont plus courtes que les autres séries et que cette variation est présente même au sein des séries de la même catégorie.

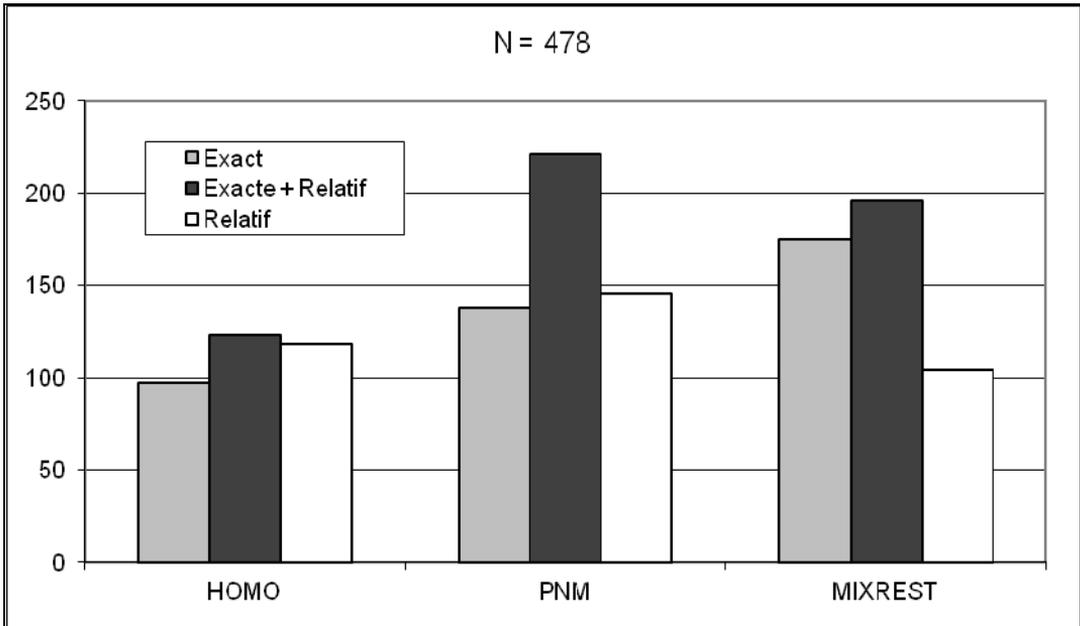


Figure 20 Variation de la longueur (médiane) selon le degré d'exactitude des marqueurs d'item.

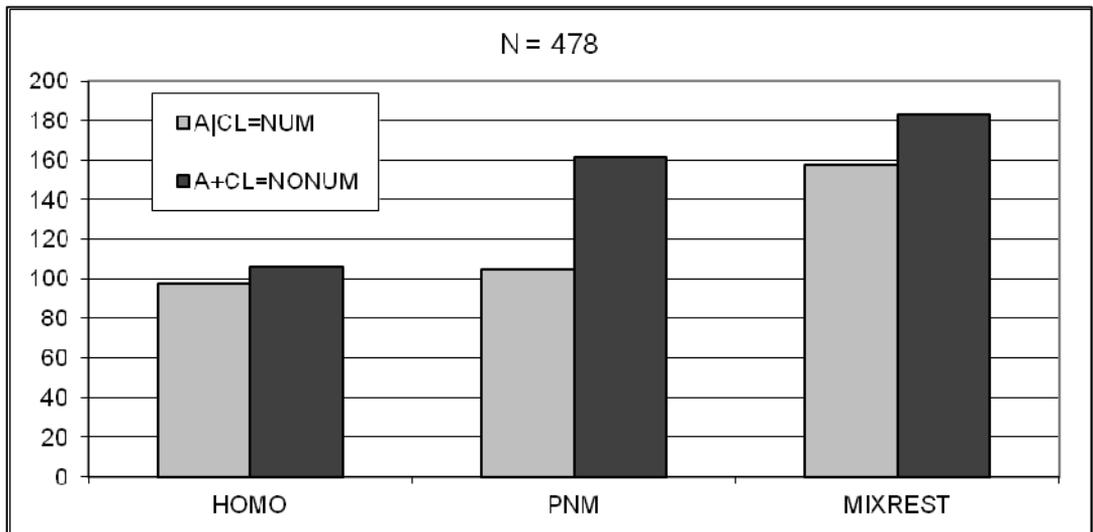


Figure 21 Variation de la longueur (médiane) selon le degré d'exactitude des amorces et clôtures.

Premièrement, selon la Figure 20, les séries introduites par la combinaison de marqueurs d'item exacts et relatifs pourraient être plus longues que celles introduites uniquement par des marqueurs d'item exacts. Dans les séries PNM, cette différence est également statistiquement significative⁶⁴. Pour les autres catégories, les différences sembleraient présenter la même tendance mais sans que le test puisse le prouver. Les longueurs des séries uniquement introduites par des marqueurs relatifs sembleraient similaires à celles des séries dont tous les marqueurs sont exacts. Cette différence est pourtant difficile à démontrer car les séries signalées uniquement par des relatifs sont fréquentes uniquement dans la catégorie PNM.

De même que la Figure 20, la Figure 21 semble aussi proposer que le marquage des segments plus longs s'avère moins exact que celui des segments courts. Selon cette figure, les séries sans amorce ou clôture exacte seraient plus longues que celles dont le nombre d'items est explicitement annoncé. À nouveau, la différence est statistiquement significative dans la catégorie des séries PNM⁶⁵. Par conséquent, il semblerait que la différence de longueur entre les séries introduites par des marqueurs exacts et les autres serait la plus importante dans ces structures.

Contrairement aux séries simples, l'analyse détaillée du lien entre le degré d'exactitude du marquage des séries enchaînées et leur longueur est plutôt difficile, surtout en raison du nombre limité d'occurrences et du marquage dans l'ensemble exact de ces structures. Les séries dans la catégorie NM+PNM sont toutefois plus longues que les autres séries et fréquemment signalées uniquement par des marqueurs relatifs. Ainsi, le lien entre les marqueurs relatifs, le marquage partiel et la longueur de série élevée est aussi prouvé pour les séries enchaînées. Par conséquent, nos résultats diffèrent de ceux de Jackiewicz (2005 : 98), selon qui le marquage des séries linéaires serait plus saillant dans les séries longues. L'Exemple 44 ci-dessous illustre une structure avec plusieurs caractéristiques typiques de ces séries plus longues : elle n'est ni précédée, ni suivie d'une amorce ou clôture exacte et les marqueurs relatifs sont les seuls utilisés. L'exemple couvre 221 mots.

Exemple 44

... Face à ces difficultés, nous proposons un recours aux Technologies de l'Information et de la Communication (TIC) qui pourraient aider les apprenants à surmonter les difficultés déjà mentionnées ci-dessus.

- 1 En fait, les TICE sont marquées par leur disponibilité spatiale dans la mesure où chaque apprenant a la possibilité de suivre la formation proposée en ligne où il veut sans avoir besoin de se déplacer pour assister à des cours en face à face.
- 2 **De plus**, elles permettent une souplesse temporelle d'autant plus que les apprenants peuvent suivre l'apprentissage selon leur emploi de temps. [...]
- 3 Soulignons **aussi** que les TICE constituent un moyen de communication qui peut se dérouler en synchrone (messagerie instantanée, VOIP Skype) ou en asynchrone (forum ou mël). De tels outils de communication favorisent à la fois la motivation [...].

⁶⁴ La valeur $p = 0,0024$ avec le test Kruskal Wallis, $H(X_2) = 9,25$, $N = 110$.

⁶⁵ Avec le Kruskal Wallis, la valeur $p = 0,0005$, $H(X_2) = 12,12$, $N = 272$.

- 4 Il est à préciser que les TICE sont une ressource importante voire inépuisable d'informations (F. Mangenot, 1998). Avec leurs sites spécialisés dans tous les domaines, les TICE sont un moyen considérable pour apprenants de réaliser les activités collaboratives proposées.
- 5 Reste à souligner qu'elles donnent la possibilité de proposer des documents écrits, sonores et audiovisuels. La diversité des documents proposés rend la formation à la fois riche, motivante et pratique.

Qotb

L'Exemple 44 comprend cinq items qui tous énumèrent des atouts des TIC (ou TICE), c'est-à-dire des recours aux Technologies de l'Information et de la Communication. Les seuls marqueurs explicites étant les deux relatifs *de plus* et *aussi*, l'ordre entre les items n'est sans doute pas un critère d'organisation très important. Dans cet exemple, les items sont plutôt organisés selon le modèle de Danes (1974, cf. 4.2.1) de progression à thème constant. Chaque item présente une qualité des TICE tout en étant le sujet grammatical de la phrase dont il fait partie. Ainsi, tous forment une chaîne de référence : *les TICE... elles... les TICE... les TICE... elles...* Outre cette continuité thématique, l'identification des items est facilitée par la similarité lexico syntaxique ; les structures des trois derniers items, surtout, sont très similaires, chaque verbe y décrivant l'action de l'auteur : *soulignons... il est à préciser... reste à souligner...*

7.2.2 ... ET COURTE ET EXACTE

Dans la section précédente, nous venons d'examiner la fréquence des marqueurs relatifs et du marquage partiel dans les séries longues. Nous avons également indiqué que les séries courtes sont signalées d'une façon plus explicite que les séries plus longues. Vu nos résultats de la section 7.1 sur la fréquence du marquage exacte dans les séries explicitement signalées, cela indique que le marquage des segments plutôt courts est également plutôt exacte. Cette section traite spécifiquement du marquage de ces séries plutôt courtes et exactes et les raisons derrière la façon dont elles sont signalées. Nous commencerons par la comparaison des séries homogènes (HOMO, HOMO+HOMO) et des séries non marquées ou composées de séries non marquées (HOMO+NM, MIX+NM, NM+NM, NMet, NM). Pourquoi les items des unes sont-ils explicitement signalés quand ceux des autres restent non marqués? Dans un second temps, nous examinerons le rôle des amorces et des clôtures dans les séries non marquées et l'usage des conjonctions de coordination dans celles-ci. Ces trois composants, ont-ils un effet sur la longueur de ces séries ?

Parmi les catégories de série présentées dans la Figure 19, les séries simples non marquées et leurs variantes enchaînées paraissent très courtes. Dans les séries enchaînées, la longueur inférieure des HOMO+NM, MIX+NM et NM+NM par rapport aux HOMO+HOMO est statistiquement significative⁶⁶. Dans les simples, bien que la différence

⁶⁶ Avec le Kruskal Wallis, la valeur $p = 0,0269$, $H(X_2) = 4,90$, $N = 65$.

ne paraisse pas dans la figure⁶⁷, les items des séries HOMO s'avèrent également plus longs que ceux des NM⁶⁸

La faible longueur des NM, HOMO+NM et MIX+NM peut également expliquer leur non marquage complet. En effet, un marquage explicite des items pourrait avoir un effet répétitif, le nombre d'items étant déjà annoncé dans la proximité, soit par une amorce ou une clôture, soit par l'autre partie de la structure enchaînée. Ces résultats soutiennent encore le point de vue exprimé de la maxime de Grice dans le marquage des NM et des séries enchaînées qu'elles composent ; le marquage explicite pourrait engendrer un risque de répétition car les items sont déjà facilement discernables et leur présence déjà indiquée dans la proximité.

L'Exemple 45 et l'Exemple 46 ci-dessous illustrent le lien entre la faible longueur, le non marquage des items et le risque de répétition. Le premier exemple est une série enchaînée se composant de deux structures inférieures, la première NM et la seconde introduite par des constructions Num+N se référant aux items de la première série. Quoique la structure couvre 134 mots et soit déjà plus longue que la majorité des séries de cette catégorie, l'ensemble de la première série, l'amorce et les deux items ne forment qu'une seule phrase. Par conséquent, le marquage explicite des items ou une clôture se référant à eux n'est pas nécessaire pour la lecture de la série mais pourrait plutôt avoir un effet répétitif.

Exemple 45

- 0 En guise de conclusion, on peut noter que le paysage tel qu'il se décline dans les Landes, en particulier au XIXe siècle, doit s'appréhender de **deux façons concomitantes**,
- 1 sous un **angle** purement stratégique (A quoi sert le paysage ? comment et qui se l'approprie ?)
- 2 **et** sous un **angle** perceptif faisant intervenir les ordres du symbolique et de l'affectif (comment est-il vécu, perçu, ressenti ?).
- 1 **Le premier angle** ne semble guère poser de problèmes. Il s'agit de découvrir les cas où le paysage devient le produit de stratégies économiques et l'instrument éventuel de mobilisations contestataires

⁶⁷ L'absence de la différence entre les longueurs totales des NM et HOMO dans la Figure 19 s'explique par deux facteurs : l'influence des séries HOMO introduites par des anaphoriques très courtes (cf. section suivante) et le nombre total des composants (=items, amorce, clôture) plus élevé des NM par rapport aux HOMO, qui obscurcit la différence de longueur des items des deux catégories de série. Une fois les anaphoriques exclues, le nombre des composants des NM est 3,6 et celui des HOMO 3,2, la différence étant statistiquement significative avec le test Kruskal Wallis, la valeur $p = 0,0041$, $H(X^2) = 8,25$, $N = 200$. Dans la pratique, ce sont surtout les clôtures qui sont beaucoup plus fréquentes dans les NM que dans les HOMO.

⁶⁸ Les anaphoriques exclues, la différence de longueur entre les items proprement dits des NM (médiane 22,0 et moyenne 45,6 mots) et ceux des HOMO (médiane 38,5 et moyenne 65,9 mots) est statistiquement significative avec le Kruskal Wallis, la valeur $p = 2,773e-07$, $H(X^2) = 26,40$, $N = 451$.

- 2 **Le second angle** peut paraître plus ardu. Insister sur la perception, le ressenti (qui, en passant, en phénoménologie, ne recouvrent pas de semblables significations), le vécu, conduit nécessairement à s'intéresser à des représentations, des émotions.

Lafargue

L'Exemple 46 présente une série composée de deux items non marqués, d'une amorce exacte et d'une clôture. Comme pour l'Exemple 45 ci-dessus (en fait tiré du même article), les items sont relativement courts bien que la série soit déjà plus longue (115 mots) que la majorité des séries non marquées. Le marquage explicite des items serait tout de même répétitif.

Exemple 46

- 0 Il faut donc considérer ce paysage forestier **sous deux angles** nettement indissociables :
- 1 **un angle purement stratégique** (comment cet espace a-t-il été créé pour répondre à des attentes économiques précises ; comment a-t-il été utilisé à des fins de protestation en retour);
- 2 **un angle perceptif** (comment cette forêt a-t-elle été perçue, vécue, par ceux qui l'habitèrent, ceux qui l'exploitèrent, ceux qui éventuellement n'y eurent aucun attachement direct ; comment devint-elle un objet social, politique, sinon culturel qui en vint à transcender les perceptions originelles).
- Cl. Il s'agit **ainsi** de travailler autant sur des pratiques (l'édification de la forêt elle-même, son exploitation, sa destruction, ses résurgences) que sur des représentations sinon des émotions.

Lafargue

Outre la faible longueur des items, dans les exemples ci-dessus, la répétition de *angle* peut aussi contribuer à l'identification des items et ainsi rendre l'usage des marqueurs explicites répétitif. Dans la partie théorique et lors de la présentation du schéma d'annotation, nous avons déjà discuté de l'importance des indices complémentaires de sériation, tels que la cohésion lexicale, dans l'identification des items non marqués. Selon le modèle d'Ariel (1990, 2001) également, les syntagmes nominaux sont repérables même dans des contextes plus ambigus, le marquage explicite devenant alors inutile. Ainsi, ces séries sont facilement repérables et cohérentes malgré l'absence de marquage.

Bref, la Figure 19 sur les longueurs des catégories de série a déjà présenté les étendues des séries simples NM et NMet. La différence entre celles-ci est également au centre ci-dessous dans la Figure 22 qui prend en compte aussi la présence des amorces et clôtures. Ainsi, la figure permet un examen plus détaillé de l'effet des amorces, des clôtures et des conjonctions de coordination à la longueur de ces séries. Les séries enchaînées sont exclues à cause de leur nature dispersée et parce que les amorces et clôtures y sont difficiles à distinguer les unes des autres (cf. entre autres l'Exemple 41 à la page 157 dans lequel la clôture de la première série est simultanément l'amorce de la seconde).

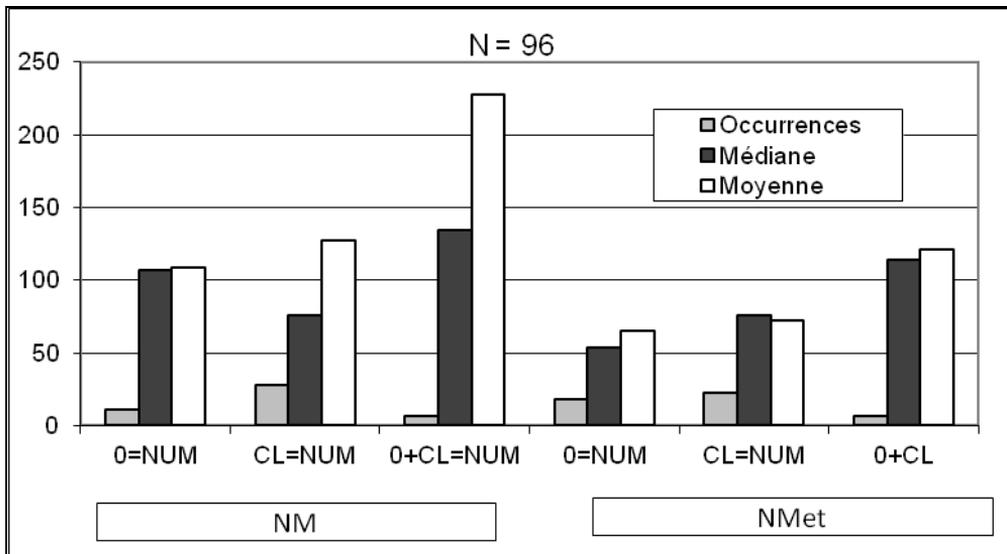


Figure 22 Longueurs des séries NM et NM selon la présence d'une amorce et / ou clôture exacte.

Tout d'abord, l'effet possible des amorces et clôtures à la longueur des séries est intéressant parce que, comme nous l'avons proposé dans les sections précédentes suivant Sinclair (2004 : 88), les amorces guident le lecteur plus explicitement que les clôtures. Ces dernières, en effet, ne peuvent que confirmer l'interprétation fournie par le lecteur pendant la lecture. La Figure 22 ne révèle pas de différence notable entre les séries NM et NMet introduites par une amorce exacte, d'une part, et celle signalées par une clôture exacte, d'autre part⁶⁹. Ainsi, bien que le rôle des amorces dans le guidage du lecteur soit supérieur aux clôtures, cette différence d'importance n'influence pas la longueur de la série.

En revanche, la La Figure 22 confirme une différence de longueur entre les séries NM et NMet. Les conjonctions de coordination sont donc utilisées, notamment, dans les énumérations extrêmement courtes⁷⁰. En général, à l'exception d'un cas, l'item est le dernier de la structure. Il est également intéressant de noter que, contrairement aux autres catégories de série, la moyenne de la longueur des NMet est aussi très basse ; il semblerait ainsi que les conjonctions de coordination ne sont pas utilisées dans des segments plus vastes.

L'Exemple 46 ci-dessus illustre les séries non marquées sans conjonction de coordination. L'Exemple 47 ci-après présente, pour sa part, l'usage des *et* dans la série. Typiques de

⁶⁹ Bien que les séries NM introduites par une clôture exacte semblent plus courtes que celles signalées par une amorce, la différence n'est pas statistiquement significative et ne se répète pas non plus dans les séries NMet. De plus, la différence entre les séries précédées ET suivies d'une amorce et clôture exacte et les autres s'explique par le nombre des composants plus élevé des premières (4,7 pour les NM avec amorce et clôture exacte et de 4,1 pour les mêmes NMet, 3,4 pour les NM avec soit amorce soit clôture exacte et de 3,7 pour les mêmes NMet.).

⁷⁰ La différence de longueur entre NM et NMet est statistiquement significative avec le test Kruskal Wallis, la valeur $p = 0,0101$, $H(X^2) = 6,62$, $N = 96$.

celle-ci, les items ne correspondent qu'aux mots individuels, auxquels la clôture *ces deux notions* se réfère.

Exemple 47

1 La question des rapports entre droit

2 **et** pouvoir est de celles qui se posent toujours sans être résolues jamais

Cl. L'exquise complication qui préside à l'architecture de **ces deux notions**, la dangereuse ambiguïté qui accompagne leur rapprochement et l'ampleur considérable des domaines qui affectent leur mise en œuvre rendent impossible toute approche globale de leurs mécanismes et inopérante toute vision unilatérale de leurs enjeux.

Leveux

7.2.3 MARQUEURS GLOBAUX ET LOCAUX

L'examen des longueurs des séries linéaires montre donc que les marqueurs relatifs, l'usage de marqueurs différents et le non marquage partiel sont typiques de l'organisation des segments textuels longs, tandis que les segments plus courts sont, en général, introduits explicitement par des marqueurs exacts. De plus, dans les segments les plus courts, le non marquage complet de tous les items de la série est fréquent ; en l'espèce, l'organisation est signalée dans la proximité, par une amorce ou une clôture indiquant le nombre d'items ou, dans les séries enchaînées, par l'autre série explicitement introduite. Sans doute, l'absence complète des marqueurs d'item dans ces séries est-elle due à l'effet répétitif potentiel du marquage, l'organisation du segment étant déjà annoncée une fois dans la proximité.

Notre analyse a donc montré que le degré d'exactitude, le marquage partiel et le non marquage varient selon la longueur du segment à organiser. Les résultats présentés dans les travaux antérieurs, en revanche, portent essentiellement sur le rôle des types de marqueurs différents dans le marquage des segments textuels de longueurs différentes. Cette section doit examiner cet aspect dans le cadre des séries linéaires. Les marqueurs examinés sont les anaphoriques, les connecteurs⁷¹ et les constructions numéro+nom, les conjonctions de coordination et les items non marqués ayant déjà fait partie de l'analyse présentée dans la section précédente. Les longueurs des items introduits par des marqueurs différents sont présentées dans la Figure 23 ci-dessous. Les séries méta et section demeurent exclues de l'analyse.

⁷¹ Vu le nombre relativement faible des expressions cadratives, nous commençons par les analyser ensemble avec les connecteurs. Les différences potentielles entre ces deux groupes de marqueurs sont examinées à la suite.

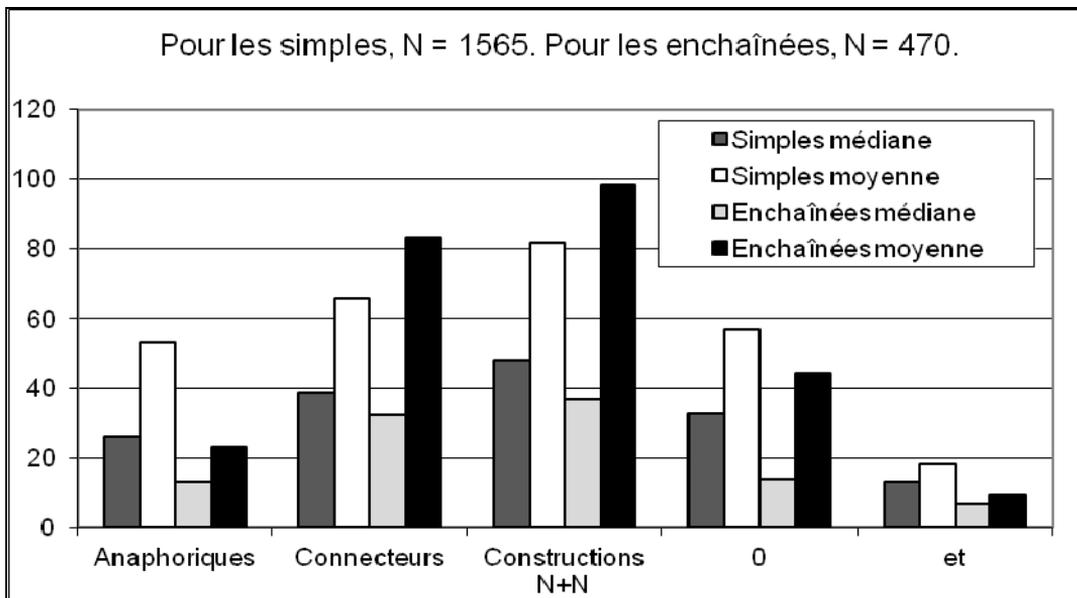


Figure 23 Longueurs des items signalés par des marqueurs différents. Le groupe *connecteurs* inclut également les expressions cadratives.

La Figure 23 montre que les longueurs des items introduits par des marqueurs différents varient incontestablement. Nous retrouvons les mêmes tendances dans les séries simples et les séries enchaînées. Les différences entre les longueurs des items sont statistiquement significatives⁷², à l'exception de quelques groupes de marqueurs. Dans les simples, la différence de longueur entre les items non marqués et ceux signalés par des anaphoriques reste à confirmer. Dans les séries enchaînées, les différences 1. entre les items introduits par des anaphoriques et les items non marqués, et 2. entre les items marqués par des connecteurs et ceux introduits par des constructions numéro + nom ne sont pas significatives. Globalement, les items des séries enchaînées paraissent plus courts que ceux des séries simples. Cette différence est naturelle car la longueur totale des enchaînées n'excède pas notablement celle des séries simples tout en présentant un nombre d'items évidemment supérieur (cf. section 6.2).

⁷² Séries simples: entre les anaphoriques et 0, la valeur $p = 0,0845$, $H(X^2) = 2,97$, $N = 615$;
entre les anaphoriques et *et*, la valeur $p = 1,648e-06$, $H(X^2) = 22,97$, $N = 143$;
entre les 0 et les connecteurs, la valeur $p = 1,253e-05$, $H(X^2) = 19,08$, $N = 1251$;
entre les connecteurs et les constructions Num+N, la valeur $p = 0,0069$, $H(X^2) = 7,29$, $N = 887$.
Séries enchaînées : entre les 0 et les anaphoriques, la valeur $p = 0,4464$, $H(X^2) = 0,58$, $N = 293$;
entre les connecteurs et les constructions Num+N, la valeur $p = 0,7548$, $H(X^2) = 0,10$, $N = 113$;
entre les A et *et*, la valeur $p = 1,586e-05$, $H(X^2) = 18,63$, $N = 153$;
entre les 0 et *et*, la valeur $p = 3,412e-06$, $H(X^2) = 21,57$, $N = 268$. ;
entre les 0 et les connecteurs, la valeur $p = 6,316e-08$, $H(X^2) = 29,26$, $N = 288$.
Toutes les probabilités sont calculées par le test Kruskal Wallis.

Pour les types de marqueurs, les résultats de la Figure 23 valident la plupart des notes des travaux antérieurs. Dans la suite, nous commencerons par discuter les anaphoriques, continuerons par les connecteurs et les expressions cadratives et terminerons par les constructions Num+N.

La Figure 23 indique donc que les anaphoriques sont utilisés dans les segments plutôt courts, conformément aux résultats de Schnedecker (2000 : 22). En discutant des anaphoriques *le premier-le second* dans les cas où l'antécédent est un SN pluriel, elle constate que l'antécédent « ne s'écarte en général jamais de plus de deux propositions des pronoms ». Dans notre corpus, la médiane des items des séries simples introduits par des anaphoriques est de 26 mots. L'Exemple 48 ci-dessous en est une bonne illustration : son premier item comprend 12 mots et le second 37 mots. La source des deux items, un SN au pluriel, se trouve dans la phrase précédant le premier item et est, selon notre schéma, annotée comme une amorce exacte.

Exemple 48

0 Saisi sous cet angle, le rapport des directeurs à l'activité éducative se définirait en fonction de **trois champs d'intervention**.

1 **Le premier** est le champ de l'environnement de l'activité éducative.

2 **Le deuxième** est le champ de l'activité éducative elle-même, lequel comporterait deux zones:

1 une zone qui se veut respectueuse de la responsabilité professionnelle des enseignants

2 **et** une autre qui est celle qui relève de cette responsabilité.

Brassard

Outre l'illustration de la longueur des items introduits par des anaphoriques, l'Exemple 48 présente une curiosité : son amorce indique une structure composée de trois items alors que la série en fait n'en comprend que deux. Ce déséquilibre s'explique par la nature récursive du second item, divisé en deux.

Schnedecker (2000 : 22-23) indique aussi que, si l'antécédent des anaphoriques n'est pas un SN au pluriel mais se compose de plusieurs éléments, il peut être dispersé dans le texte antérieur et ses composants peuvent même se trouver dans des paragraphes distincts. Ce cas correspond, dans notre corpus, aux séries enchaînées dont la première série est NM (ou NMet) et la seconde se compose d'items s'y référant. Dans l'Exemple 49 ci-dessous, notamment, les items non marqués de la première série fournissent les sources des items de la série suivante, signalée par des anaphoriques et s'y référant.

Exemple 49

0 La modélisation de la synchronisation entre syntaxe et sémantique nous a permis de distinguer deux fonctions importantes de la syntaxe :

1 la fonction argumentale

2 **et** la détermination.

Cl. Nous avons suggéré que ces deux fonctions étaient apparues successivement.

1 **La première**, obtenue notamment par des mécanismes de flexion et de position, permet l'expression de la prédication principale, utilisée par le niveau argumentatif

- 2 **La seconde**, obtenue par l'emboîtement des syntagmes et les dépendances à distance, permet l'accumulation de prédications auxiliaires qui servent à contraindre les objets du discours pour l'interlocuteur.

Dessalles

Notre corpus comprend 22 séries enchaînées dont la première série est non marquée et la seconde se compose d'items signalés par des anaphoriques. Ce sont très majoritairement des HOMO+NM ou HOMO+NMet. La longueur totale de ces séries ne diffère pourtant pas de celle des séries simples homogènes introduites par des anaphoriques : la longueur médiane est de 74 mots pour les simples et de 81,5 mots pour les enchaînées⁷³. De plus, la longueur des items signalés par des anaphoriques est similaire dans les séries HOMO+NM et HOMO+NMet et dans les autres séries enchaînées. Ainsi, la nature dispersée des antécédents des anaphoriques, quand il n'est pas un SN, ne se répète pas dans notre corpus dans des séries linéaires. L'Exemple 49 ci-dessus illustre bien ces séries : les antécédents y sont énumérés séparément mais restent toujours proches les uns des autres. Bien que la série soit, avec ses 86 mots, déjà plus longue que la majorité des séries composées de NM(et) et de HOMO introduites par des anaphoriques, elle demeure pourtant plutôt courte et les antécédents, soit la série NMet, sont facilement identifiables.

Après les anaphoriques, les connecteurs sont le deuxième type de marqueur dont le fonctionnement global et local est discuté dans les travaux antérieurs. Nous avons déjà, en introduisant cette section, cité Ho Dac (2007 : 227) qui note que les connecteurs sont en général utilisés dans des contextes intraparagraphiques. Nous avons également évoqué Jackiewicz et Minel (2003) qui proposent que « la portée d'un MIL n'exécède pas en général deux paragraphes ». Goutsos (1996), en revanche, attribue des fonctions différentes à ces marqueurs : certains, tels que *now* et *then* et les corrélatifs anglais des MIL, seraient utilisés seulement pour l'encadrement. D'autres, tels que les marqueurs additifs et adversatifs, peuvent signaler la continuité.

En tout, comme indiqué dans la Figure 23, les connecteurs (expressions cadratives incluses) signalent dans notre corpus des items de 39 (médiane des séries simples) et 32,5 (médiane des séries enchaînées) mots. Notre corpus semble donc valider le fonctionnement plutôt restreint de ces marqueurs noté dans les travaux antérieurs. Or, nous devons relever que les connecteurs peuvent également signaler des segments plus étendus, comme le montre la moyenne relativement élevée des longueurs des items qu'ils marquent. Ce n'est pourtant pas leur fonctionnement le plus typique.

Selon le modèle de Goutsos, les connecteurs peuvent adopter des fonctions différentes dans la segmentation du texte. Pareillement, la théorie de Charolles (1997, présentée en section 3.4.4) distingue les MIL, ouvrant un *espace de discours*, des expressions cadratives proprement dites, ouvrant un *univers de discours*. Afin d'examiner si ces fonctions

⁷³ Selon le test Kruskal Wallis, il est très improbable qu'il y ait une différence entre ces deux groupes, la valeur $p = 0,935$, $H(X^2) = 0,01$, $N = 40$.

différentes sont discernables dans les longueurs des items introduits par ces marqueurs, une présentation plus détaillée des longueurs des items signalés par ceux-ci suit en Figure 24. De même que notre schéma d'annotation (cf. section 5.2.2), cette figure distingue les connecteurs proprement dits des adverbiaux détachés à l'initial de la phrase, c'est-à-dire des expressions cadratives. Les longueurs des items introduits par des connecteurs adversatifs sont aussi présentées. Comme le nombre d'occurrences des items signalés par des connecteurs est relativement faible dans les séries enchaînées, nous nous focalisons sur les séries simples.

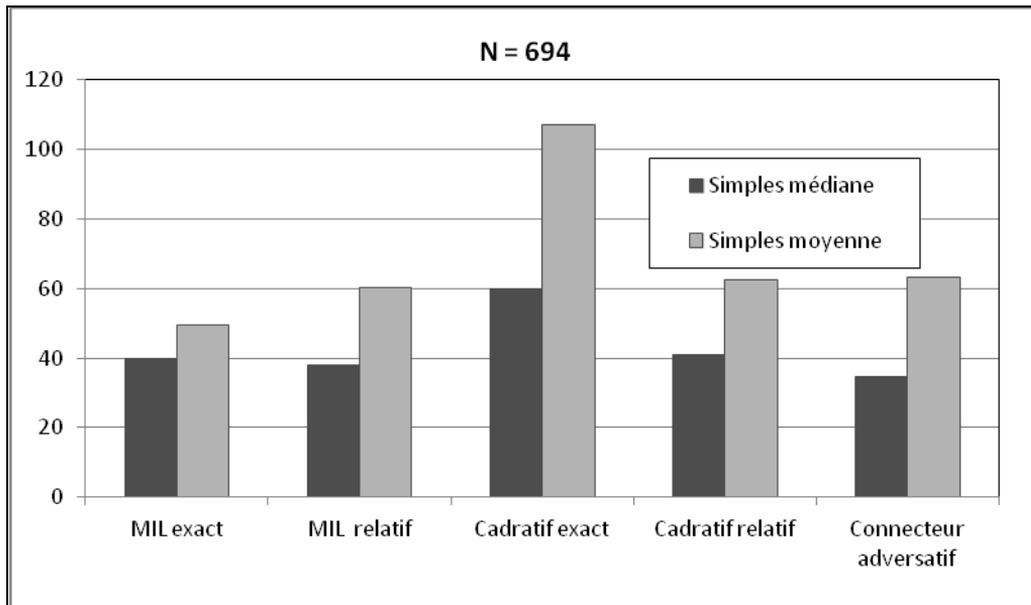


Figure 24 Longueurs des items signalés par des marqueurs différents.

La Figure 24 indique que les longueurs des items signalés par des connecteurs et des expressions cadratives différents ne varient pas d'une façon importante. Additionnellement, les connecteurs adversatifs présentent une longueur très similaire aux MIL et il n'y a pas non plus de différence entre les MIL exacts et relatifs. Toutefois, il semblerait que les items introduits par les expressions cadratives soient un peu plus longs que ceux introduits par des MIL⁷⁴. Cela pourrait être le signe que la portée cadrative de ces marqueurs est plus longue que celle des MIL. De plus, la moyenne plus élevée des cadratives exactes peut aussi indiquer que ces marqueurs sont fréquemment utilisés dans des séries plus longues.

Le fonctionnement des expressions cadratives est illustré par l'Exemple 50 ci-dessous. Il comprend deux items de 108 et 69 mots, donc déjà plus longs que la majorité des items

⁷⁴ Comparées entre les MIL exacts + relatifs et toutes les expressions cadratives, la différence semblerait statistiquement significative avec le test Kruskal Wallis, la valeur $p = 0,0303$, $H(X^2) = 4,69$, $N = 613$.

introduits par ces marqueurs. Nous pouvons cependant noter que, malgré la longueur relative des items, l'analyse de la portée cadrative possible des marqueurs s'avère ardue. En effet, les items ne couvrent qu'une seule phrase chacun et sont toujours courts si comparés aux longueurs des paragraphes ou sections entiers. Ainsi, il est difficile de comparer nos résultats sur leur longueur par exemple à ceux de Ho Dac (2007 : 283), selon lesquels la portée cadrative des adverbiaux spatiaux et temporels est surtout liée à la position initiale de section ou de paragraphe.

Exemple 50

- 0 Pour illustrer cette dynamique des signes, on s'inspirera **d'exemples cités** par Hanson (1958) et Morand (2004).
 - 1 **Dans le premier**, Kepler dispose d'une multitude de faits bruts sur les mouvements de la lune qu'il tente d'interpréter comme des mouvements circulaires autour de la terre ; mais il a la surprise de constater que la lune est toujours en retard sur les prévisions basées sur cet interprétant métaphorique, à moins d'y ajouter des épicycles à divers moments de son parcours ; insatisfait de cette interprétation (cette insatisfaction constituant elle-même un interprétant plus intime), Kepler se réfère alors à l'image de l'ellipse (nouvel interprétant métaphorique) qui aboutit à des prévisions conformes aux faits bruts, tout en soulignant l'effet conjoint de l'attraction du soleil et de la terre.
 - 2 **Dans le second**, il est question d'un parapentiste qui se lance dans le vide en espérant atterrir à tel endroit : cet objectif lui sert d'interprétant potentiel de tous les signes qu'il traite durant son vol : des sensations qu'il éprouve intérieurement aux connaissances techniques dont il dispose sur le vol à voile, en passant par les faits bruts sur la direction du vent et la conformation du terrain.
- Cl. Kepler et le parapentiste sont **donc** des simulateurs d'action qui doivent confronter en permanence leurs prévisions [...].

Demailly

Bref, les résultats sur le fonctionnement des connecteurs et des cadratifs différents ne sont pas très informatifs. Afin de mieux comprendre la façon dont ces marqueurs organisent des segments textuels courts et longs, nous les avons encore examinés plus en détail, en comparant des marqueurs individuels. Lors d'annotation, nous avons ajouté manuellement le marqueur exact utilisé pour 681 des 716 items introduits par un connecteur dans les séries simples. Bien que cette annotation ne couvre pas tout le corpus, à notre avis, elle peut tout de même offrir une approche alternative à la comparaison des connecteurs différents. La Figure 25 ci-dessous présente les longueurs des items introduits par les connecteurs les plus fréquents. Les adversatifs sont exclus de cette figure. En effet, ils sont décrits *supra* dans la Figure 24. Afin de faciliter l'analyse, nous avons combiné certains types de marqueurs de longueurs similaires⁷⁵.

⁷⁵ D'UNE PART inclut les occurrences de *d'une part*, *d'autre part*, *d'un côté*, *d'un autre côté*, *d'autre côté*, *de l'autre côté*, et *de l'autre*. PREMIEREMENT inclut les occurrences de *premièrement* et *dans un premier temps*. AUSSI inclut les occurrences de *aussi*, *également*, *encore*, *par ailleurs*, *de plus* et *de même*. ENFIN inclut *enfin* et *finalement*.

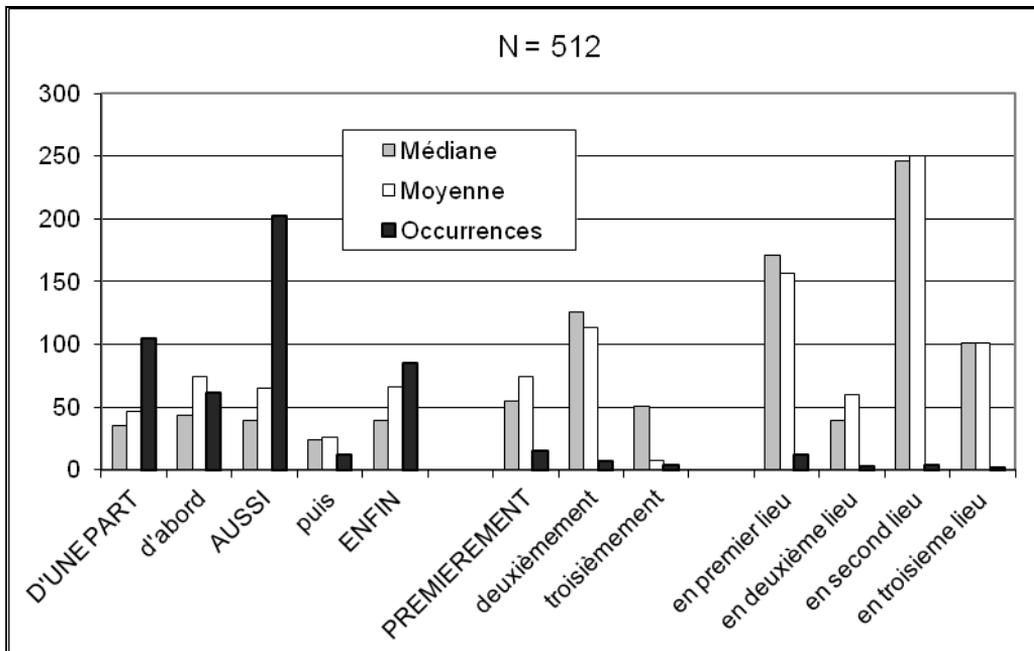


Figure 25 Longueurs des items introduits par les connecteurs les plus fréquents.

Surtout, la Figure 25 prouve que les connecteurs différents sont utilisés pour signaler des items de longueurs différentes. Tout d'abord, les items introduits par des MIL à l'origine spatiale *en x lieu* sont très longs. Bien que ces marqueurs soient rares dans notre corpus, leur longueur par rapport aux items introduits par les autres marqueurs rapportés dans la figure peut être statistiquement prouvée⁷⁶. Les MIL en *ment*, en revanche, semblent signaler des items dont la longueur est entre ceux introduits par des MIL *en x lieu* et ceux marqués par les autres connecteurs⁷⁷. Enfin, les autres connecteurs, c'est-à-dire les variantes de *d'une part*, *aussi* et *enfin* ainsi que *d'abord* et *puis*, sont les plus courts, selon cette figure.

Outre les différences entre ces trois groupes de marqueurs, nous pouvons également discerner quelque variation dans les fonctionnements des connecteurs du « premier groupe » qui ne se terminent ni en *lieu*, ni en *temps*. Surtout, les marqueurs *D'UNE PART* et *puis* semblent plus courts que les autres⁷⁸. L'Exemple 51 suivant illustre la faible

⁷⁶ Avec le Kruskal Wallis, la différence entre les items introduits par *en x lieu* et les autres connecteurs présentés dans la figure est significative, la valeur $p = 1,548e-05$, $H(X^2) = 18,68$, $N = 512$.

⁷⁷ Avec le test Kruskal Wallis, la différence entre les longueurs des items introduits par *en x lieu* et ceux signalés par des marqueurs en *-ment* semble statistiquement significative, la valeur $p = 0,0409$, $H(X^2) = 4,18$, $N = 47$.

De même, aussi la différence entre les marqueurs en *-ment* et les autres connecteurs est significative, la valeur $p = 0,0008$, $H(X^2) = 11,30$, $N = 491$.

⁷⁸ La différence de longueur entre les items introduits par *D'UNE PART* et les autres connecteurs du premier groupe (*puis* exclus) semblerait statistiquement significative avec le Kruskal Wallis, la valeur $p = 0,0532$, $H(X^2) = 3,74$, $N = 453$.

longueur des items signalés par *d'une part* et ses variantes. Les items de cette série couvrent 31 et 43 mots et ainsi illustrent bien la longueur générale des items introduits par ces marqueurs, leur médiane étant de 35 mots.

Exemple 51

- 0 Cette phrase est ambiguë à plusieurs égards.
- 1 **D'une part**, il existe une ambiguïté entre la lecture déontique (permission) et épistémique (possibilité). Pour la suite de notre propos, on se focalisera uniquement sur la première de ces lectures.
- 2 **D'autre part**, cette phrase est encore ambiguë de la manière suivante : ou bien, le quantificateur a le prédicat modal dans sa portée (lecture de *re*), ou bien c'est le prédicat modal qui a la portée la plus large (lecture de *dicto*).

Busquets

L'Exemple 52 ci-dessous, en revanche, présente le fonctionnement de *puis* dans une série comprenant trois items, tous introduits par des MIL. L'item commençant par *puis* couvre 23 mots, soit la médiane des items signalés par ce marqueur. Les autres items sont également plutôt courts, le premier englobant 14 mots et le troisième 23.

Exemple 52

- 0 Dans cette perspective, nous procéderons à une déconstruction de cette campagne de pétitions en examinant **les différentes phases** par lesquelles passent les textes de protestation.
- 1 On suivra **d'abord** l'élaboration et les conditions matérielles de collecte des pétitions.
- 2 **Puis**, on s'attachera à cerner les principales caractéristiques de cette campagne en étudiant les singularités de la parole et de la géographie pétitionnaire.
- 3 **Enfin**, en suivant les pétitions du village au Parlement, nous questionnerons les usages pluriels de ces textes de protestation dans l'arène politique.

Jarrige

Bien que le connecteur *puis* ait une origine temporelle et puisse indiquer la succession chronologique des items (cf. Riegel *et al.* 1994 : 618), il fonctionne, dans l'Exemple 52 ci-dessus, dans une série métadiscursive présentant les étapes de l'article de recherche en question. De fait, la majorité des connecteurs structurant des items courts présentés la Figure 25, c'est-à-dire au moins *d'abord*, *puis* et *enfin*, sont à l'origine temporels (cf. Riegel *et al.* 1994 : 618). L'analyse statistique montre pourtant que, globalement, les items des séries temporelles simples présentent une longueur similaire à celle des autres items et qu'il n'y a pas non plus de différence entre les longueurs totales des séries temporelles simples et des autres⁷⁹.

Entre les items signalés par *puis* et les autres connecteurs du premier groupe, la différence est incontestable avec la valeur $p = 0,0074$ avec le Kruskal Wallis, $H(X_2) = 7,16$, $N = 465$.

⁷⁹ Avec le Kruskal Wallis, la valeur $p = 0,5119$ entre les items des séries temporels et des autres, séries sections exclues. $H(X_2) = 0,43$, $N = 1565$.

Entre les longueurs totales des séries temporelles simples et les autres, la valeur $p = 0,2496$, $H(X_2) = 1,33$, $N = 595$.

En définitive, bien que la temporalité ne motive pas la faible longueur des items introduits par les connecteurs « du premier groupe », la distance entre ceux-ci, ceux en *-ment* et ceux en *x lieu*, semble réelle. Ainsi, il semblerait que ce serait plutôt l'origine sémantique du connecteur que son type exact qui influence la longueur du segment à organiser. Cette question mérite pourtant encore un examen plus détaillé que peut être mené dans cette étude. Comme l'a noté Schnedecker (2001 : 262), les travaux sur les MIL ne sont pas très informatifs quant à la différence entre les marqueurs d'origines différentes.

Finalement, les *constructions numéro+nom* sont le dernier type de marqueur à examiner dans cette section. Comparées aux connecteurs, expressions cadratives et anaphoriques que nous venons d'examiner, ces indices ne forment pas de type de marqueur établi et examiné dans maintes études. Nous les avons cependant classifiés comme un groupe distinct puisqu'ils se distinguent clairement des trois autres en étant sémantiquement indépendants sans être des adverbiaux.

Au début de cette section, la Figure 23 montre que les constructions numéro+nom servent à signaler les items les plus longs des marqueurs examinés. Dans la section 6.3 du début de l'analyse, nous avons également déjà indiqué qu'elles sont plus fréquentes dans les séries simples que dans les enchaînées.

Les constructions Num+N ne formant pas de type de marqueur établi, elles sont relativement variées. Les marqueurs exacts sont très souvent, comme leur nom indique, de forme *le -ième NOM* et les marqueurs relatifs de forme *un autre NOM*. L'adjectif *suivant* est également répandu. Les deux exemples ci-dessous illustrent l'usage de ces marqueurs. Le premier se compose de trois items et d'une clôture pour couvrir en tout 751 mots. Chaque item énumère un récit différent, en commençant par son quasi-synonyme. Les premier et troisième items correspondent aux constructions Num+N fonctionnant comme sujets grammaticaux : *certains fabliaux décrivent* et *une autre version du même conte*. Le deuxième item, en revanche, est introduit par l'expression cadrative relative *dans un autre conte*. La série est encore terminée par la reprise *les scènes de ce genre* qui peut être analysée comme une anaphore résomptive.

Exemple 53

- 1 Les contes cherchent à distraire, non à apitoyer. Pourtant **certains fabliaux** décrivent de façon réaliste les effets de la fatigue. Deux jeunes gens ont fait un si long trajet en cherchant du travail que le soir ils tiennent à peine debout. Dans la ferme où ils sont hébergés, le plus jeune se jette sur la bouillie qu'on leur offre. L'aîné en revanche est si las qu'il ne peut avaler que " La moelle de deux tiges de choux [...]. ". Au milieu de la nuit, il se réveille tenaillé par la faim. [...]
- 2 **Dans un autre conte**, un vilain et sa femme discutent au lit à propos d'avoine et de grain. Nous n'avons pas de deniers, dit l'homme. C'est fort ennuyeux, car nous devons à notre voisin. Il serait grand temps de le rembourser. Faisons demain battre l'avoine qui est dans l'auvent et nous la vendrons. Mais, objecte la femme, notre grenier est encore plein d'avoine en grains ; on en peut tirer bien assez d'argent. Il en reste bien trois ou quatre muids. Quel besoin de faire battre à présent ? [...]

3 **Une autre version du même conte** met en scène un couple de vilains au petit matin. La femme s'éveille et pousse son mari. Il est grand temps d'aller au moulin, dit-elle, car nous n'avons plus que deux pains. Impossible, dit le vilain, depuis trois jours je suis malade. Réveillez donc Martin, ce colporteur qui chaque mois couche chez nous trois ou quatre fois. Promettez-lui une bonne tourte de pain. [...]

Cl. **Les scènes de ce genre** sont fréquentes. Comme on le voit, les poètes du XIII^e siècle n'hésitent pas à donner la parole à des vilains. [...]

Lorcin

Les items de l'Exemple 53 énumèrent donc certains types de contes et la construction Num+N relative, introduisant le troisième item de l'Exemple 53, indique que le conte présenté par l'item doit être ajouté à l'énumération. Le texte ne précise pourtant pas d'ordre entre les contes différents et l'on peut se demander quelle est son importance dans cette série. Dans l'Exemple 54 ci-dessous, en revanche, l'ordre entre les items est un critère d'organisation important. Cette série se compose d'items décrivant des analyses différentes menées dans le cadre d'une étude en éducation afin d'examiner des correspondances possibles entre les qualités personnelles des étudiants médiateurs et leur succès dans cette tâche. Chaque item correspond à un test effectué pour examiner une qualité des élèves. Le test et la qualité forment donc les deux principes d'énumération de la série, soit les points communs à tous les items. Le quatrième item révèle que les items sont très probablement menés l'un après l'autre ; il s'agit donc d'une série temporelle.

Exemple 54

- 1 Pour évaluer **si le sexe des médiateurs** pouvait avoir une influence sur leur développement aux niveaux comportemental, socioaffectif et sociocognitif, **des tests non paramétriques (U de Mann-Whitney)** ont été effectués sur la différence entre le prétest et le post-test pour chacune des variables étudiées. Bien que les résultats n'indiquent aucun résultat significatif pour la variable « sexe », d'autres analyses ont permis de constater que les garçons médiateurs avaient été consultés aussi souvent que les filles médiatrices, toutes proportions gardées.
- 2 **Les tests de corrélation** (r de Pearson) ne révèlent aucun lien significatif **entre l'âge des médiateurs** et les résultats obtenus entre le prétest et le post-test pour les mêmes variables étudiées.
- 3 **La troisième question à l'étude** visait à vérifier **si le nombre de médiations effectuées** par les médiateurs pouvait avoir des effets sur les aspects comportemental, socioaffectif et sociocognitif. **Les tests de corrélation** (r de Pearson) présentés dans le Tableau 1 indiquent une augmentation significative du niveau d'anxiété chez les médiateurs qui ont effectué un plus grand nombre de médiations (r de Pearson = 0,655, $p < 0,01$)
- 4 **La dernière série d'analyses** a été effectuée à l'aide des données tirées des fiches statistiques complétées par les médiateurs après chaque médiation. **Des tests de khi-carré** ont révélé un lien significatif entre **le niveau de satisfaction personnelle exprimé** par les médiateurs et l'issue positive ou négative de leurs médiations. [...]

Beaumont

Les items explicitement signalés dans l'Exemple 54 sont le troisième et le quatrième, tous deux introduits par des constructions Num+N exactes. Les items non marqués répètent de

façon intéressante les deux principes d'énumération, sans être syntaxiquement ou thématiquement parallèles. Le premier item commence par un subordonné exprimant la qualité personnelle énumérée, *le sexe des médiateurs*, et le test appliqué est présenté dans l'objet de la phrase principale, *des tests non paramétriques ont été effectués*. Dans le deuxième item, le test appliqué est le sujet de la première phrase et le second principe d'énumération offert dans le complément d'objet indirect *ne révèlent aucun lien significatif entre l'âge des médiateurs...* En ce qui concerne les items explicitement signalés, ils adoptent encore une autre forme et, outre la répétition des deux principes *test* et *qualité personnelle* présents également dans les items précédents, ces items énoncent encore un autre principe d'énumération. Celui-ci est exprimé par des noms différents qui, pourtant, peuvent être interprétés comme sémantiquement similaires : *la troisième question* et *la dernière série d'analyses*. Bien que ces items fassent augmenter le nombre des principes d'énumération utilisés dans la série ils contribuent, à notre avis, tout de même au parallélisme des items et rendent l'énumération plus explicite. Cet effet est surtout créé par la structure syntaxique et sémantique similaire de ces deux items. Ainsi, ils commencent par les constructions Num+N dans lesquelles les noms correspondent à ces quasi-synonymes.

Dans l'Exemple 54, les constructions Num+N exactes indiquent que les tests doivent être ajoutés dans l'énumération des analyses menées en précisant également l'ordre entre eux. Or, se pose une question complexe : comment ces constructions devraient-elles être définies et, surtout, comment diffèrent-elles des autres types de marqueurs que nous avons examinés ? La réponse à cette interrogation pourrait également expliquer la raison de leur usage dans les séries longues. Bien qu'un examen détaillé de ce propos ne soit pas possible dans cette étude, il mérite discussion.

Une caractéristique typique des constructions Num+N doit être leur variété. Une partie d'entre elles est référentielle et utilisée ensemble avec des anaphoriques, comme nous l'avons illustré en section 6.3.1, entre autres, à l'aide de l'Exemple 34 en discutant des séries MIXAL dont un item est signalé par un anaphorique et l'autre par une construction Num+N. D'autres, comme celles utilisées dans les deux exemples précédents, sont en revanche mieux définies comme revêtant une fonction additive, également susceptible de préciser une place exacte dans l'énumération. Ce statut sémantiquement indépendant distingue ces marqueurs des anaphoriques. En revanche, ils diffèrent des adverbiaux par leur capacité de fonctionner comme sujet ou objet de la phrase dans le noyau syntaxique.

Selon le modèle de Goutsos (1996), les connecteurs exacts signalent l'introduction d'une nouvelle zone de continuation tandis que les connecteurs additifs, de même que les pronoms personnels, n'indiquent que l'ajout du segment dans la zone de continuation déjà ouverte. Notre hypothèse est que les constructions Num+N seraient similaires aux pronoms personnels et aux connecteurs additifs et élaboreraient la zone déjà ouverte en y ajoutant une nouvelle dimension. Par exemple selon notre interprétation, le dernier item de l'Exemple 54 n'ouvre pas de nouvelle zone mais indique que *la dernière série d'analyse*

fait partie de l'ensemble de tests présenté. Ce défaut de capacité d'introduction d'une nouvelle zone de continuation est sans doute lié à la fonction grammaticale de ces constructions ; comme l'indiquent, notamment, les travaux de Charolles (1997) et Virtanen (1992), les adverbiaux détachés au début de la phrase, connecteurs inclus, ont un rôle important dans la structuration du texte et du discours. Les marqueurs correspondant au sujet de la phrase, comme les anaphoriques, affichent, en revanche, une tendance à jouer un rôle moins central, comme l'indiquent leur fonctionnement comme indice de continuation dans le modèle de Goutsos, d'une part, et la faible longueur des items introduits par ces marqueurs dont nous avons attesté dans cette section, d'autre part.

Or, nous avons déjà noté dans cette section que les différences de fonctionnement proposées par le modèle de Goutsos ne sont pas toutes discernables dans la longueur des items introduits par ces marqueurs : il n'y a pas, notamment, de différence notable entre les items introduits par les MIL exacts et ceux signalés par les connecteurs additifs, bien que les premiers soient censés introduire une nouvelle zone de continuation et les seconds non. D'après notre réflexion, la capacité ou l'incapacité à ouvrir une nouvelle zone de continuation peut pourtant refléter une certaine force chez le marqueur : tandis que les uns ont la « force » d'ouvrir de nouvelles séquences, d'autres, plus « discrets », en sont dépourvus.

Si nous acceptons l'hypothèse de la nature plus « discrète » des constructions Num+N, leur fréquence dans les séries longues peut être expliquée par nos résultats antérieurs sur l'introduction des séries plus longues ; comme nous l'avons noté à plusieurs reprises, selon nos analyses, le marquage des éléments plus longs semble plus implicite et plus imprécis que celui des éléments courts. Se pourrait-il donc que les constructions Num+N, plus « discrètes », soient mieux adaptées à cette tendance générale de marquage des séries longues alors que les adverbiaux et les connecteurs soient excessivement « forts » ou explicites pour elles ? Il est évident que cette étude ne peut que relever l'éventualité de ce lien. Nous pouvons uniquement confirmer la fréquence de tous ces indices, soit du marquage partiel, de l'usage des marqueurs relatifs et des constructions Num+N, dans les séries linéaires plutôt longues.

7.2.4 SERIES SECTIONS ET SERIES META – SEGMENTATION DU TEXTE PAR LES TITRES

Cette dernière sous-section sur la variation du marquage des séries linéaires de longueurs différentes se concentre sur les séries longues composées de sections titrées, exclues des analyses des sous-sections précédentes. Dans les séries section, tous les items correspondent aux sections titrées de l'article. Les séries méta se composent (au moins) de deux sections inférieures dont la première, c'est-à-dire la série *metadiscursive*, présente de manière cataphorique les items de la série suivante, correspondant aux sections titrées.

Dans la partie théorique, nous avons déjà présenté le MAT, le modèle d'architecture textuelle (Virbel 1989 ; Luc et Virbel 2001 ; Luc *et al.* 2002) qui tient compte dans

l'analyse de la structure visuelle du texte. Souvent, ces indices visuels correspondent aux puces et tirets utilisés, entre autres, dans les énumérations. Or, la segmentation du texte en sections et sous-sections titrées fait aussi partie de cette structure. Les titres et leurs fonctions discursives ont également fait l'objet de nombre d'autres travaux (notamment Sullet-Nylander 1998 ; Virbel 2002 ; Ho Dac *et al.* 2004 ; Jacques et Rebeyrolle 2006 ; Rebeyrolle *et al.* 2009). Ces travaux identifient deux fonctions principales des titres : ils segmentent le texte en unités hiérarchiques et participent à l'organisation thématique du texte. Dans les séries linéaires, c'est surtout la segmentation du texte par les titres qui importe ; ce sont elles, outre les marqueurs explicites, qui délimitent les items. Ainsi, les titres soutiennent le fonctionnement des marqueurs de sériation signalant le début des items et l'ordre entre eux et fonctionnent ainsi comme indices complémentaires (cf. 5.2).

Dans notre corpus, nous avons repéré 23 séries section et 31 séries méta. Les séries méta étant plus nombreuses, nous estimons que l'introduction cataphorique des sections par une série *métadiscursive*, combinée à un marquage possible des sections, est plus fréquente que le marquage unique des items au début de chaque section.

La longueur importante est un trait typique de ces deux catégories de série. Les longueurs sont présentées dans la Figure 26 et la Figure 27 ci-dessous. Nous notons que les séries section sont plus courtes que les méta : tandis que la majorité des premières n'excède pas 2 000 mots, les secondes couvrent fréquemment plus de 4 000 mots. C'est là l'indication que l'introduction des items par une série métadiscursive est plus fréquente dans les structures très longues. Dans les séries plus courtes, en revanche, l'introduction des sections-items par une amorce et / ou par le marquage des items semblerait suffire. De même que la fréquence du non marquage complet dans les séries courtes annoncées par une amorce ou une clôture exacte, la rareté des séries méta courtes peut être expliquée par la maxime de quantité de Grice (1975) : l'introduction des items à venir par une série métadiscursive risque d'être excessive et répétitive si cette série et les items sont proches les uns des autres. Enfin, il faut noter que, dans notre corpus, les titres ne segmentent toujours les sections en unités hiérarchiques, ni n'indiquent explicitement l'ordre entre eux. En effet, une partie des articles est uniquement composée de sections de même niveau sans sous-sections inférieures. De plus, tous les titres n'y sont pas nécessairement numérotés.

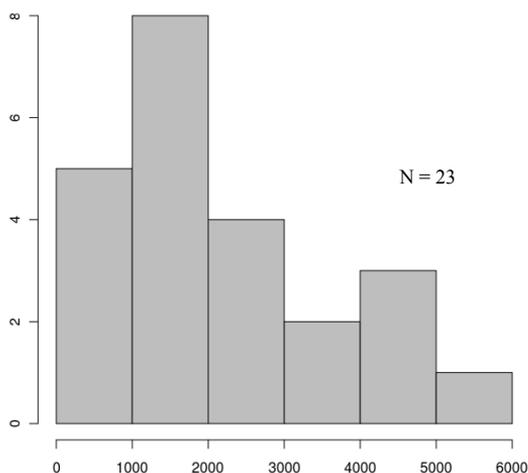


Figure 26 Longueurs des séries section.

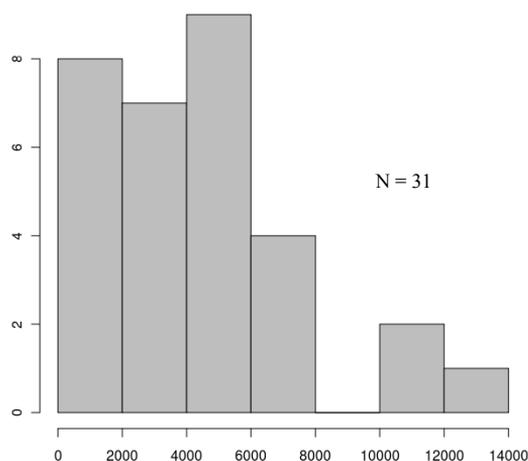


Figure 27 Longueurs des séries méta.

En ce qui concerne le marquage des séries section et méta, les distributions des catégories de série sont illustrées dans la Figure 28 et la Figure 29 ci-dessous. Pour les séries section, les catégories semblent suivre les tendances rapportées pour les autres séries simples longues. Ainsi, comme dans celles-ci, les séries partiellement non marquées sont fréquentes et les HOMO, MIXAL et NM plutôt rares, les NMet étant complètement absentes. La seule différence notable entre les séries section et les autres séries simples longues semblerait être la rareté des MIXREST, fréquemment longues parmi ces dernières mais pratiquement absentes des séries section. D'une part, cela pourrait être une résultante de la présence des titres. Nous pouvons émettre l'hypothèse que le marquage explicite de tous les items ne serait pas nécessaire du fait de leur délimitation par les titres. D'autre part, le non marquage partiel étant aussi fréquent dans les autres séries longues, cette façon d'introduire l'organisation textuelle peut simplement être typique de ce niveau du texte.

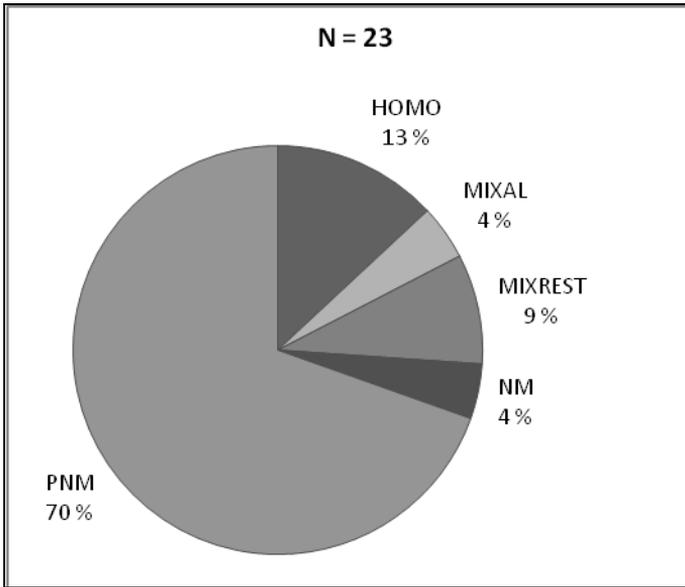


Figure 28 Catégories des séries section.

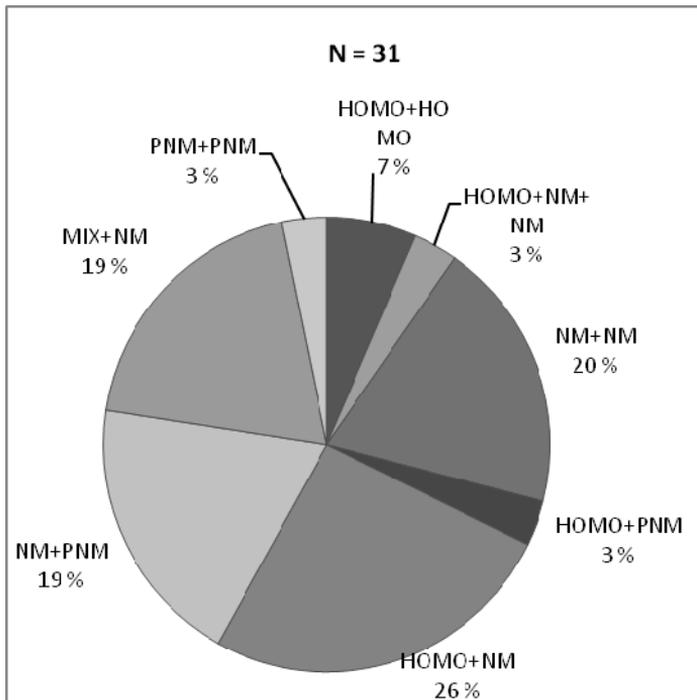


Figure 29 Catégories des séries méta. Les NM et NMet sont toujours combinées sous la catégorie des NM.

Quant au marquage des séries méta, ces structures font très majoritairement partie des MIX+NM, NM+PNM, NM+NM et HOMO+NM. Comme seules les NM+PNM sont plutôt longues dans les séries typiques sans sections titrées, ce sont surtout les catégories NM+NM et HOMO+NM qui semblent plus fréquentes dans les séries méta. Une analyse manuelle montre que la série métadiscursive est la plus fréquemment explicitement signalée. C'est donc surtout le marquage explicite des items-section qui semble évité. D'une part, il est possible que le marquage des items des deux séries risque d'avoir un effet répétitif malgré la longueur évidente de ces structures. Ce peut être le cas surtout dans le marquage du premier item de la seconde série. D'autre part, il est possible que les marqueurs explicites devant les items extrêmement longs soient difficiles à mémoriser pour le lecteur et peuvent, par conséquent, être délaissés.

L'Exemple 55 ci-dessous illustre une série méta très typique, composée d'une série métadiscursive homogène signalée par des MIL et d'une série section non marquée. Typiquement pour les méta, les items explicitement signalés de la première série présentent les thématiques des items-section à venir.

Exemple 55

0 Afin de répondre à ces questions, une démarche en quatre temps s'avère nécessaire.

- 1 **Premièrement**, nous présenterons une brève généalogie de l'idée d'égalité en éducation et de la contribution de la sociologie à l'analyse de son incarnation dans le réel, [...].
 - 2 Il s'agit **deuxièmement** de transposer cette réflexion sociologique et politique dans le contexte québécois en rappelant des éléments pertinents de l'histoire récente des politiques éducatives, dont certaines privilégient l'introduction d'une certaine différenciation dans la dispensation des services et ont une incidence sur l'idée d'égalité,
 - 3 et, **troisièmement**, d'établir à l'aide de données empiriques et d'indicateurs quelques constats sur l'évolution de l'enseignement public et privé, toujours en lien avec l'idée d'égalité.
 4. Ceci nous permettra, **quatrièmement**, de formuler une hypothèse constituant une réponse à la question que nous posions d'entrée de jeu : [...]
- Cl. **En conclusion**, nous situerons notre réponse dans le contexte de tendances lourdes des rapports sociaux, de l'importance des classes moyennes dans les sociétés modernes avancées, et de leurs inquiétudes exacerbées par la mondialisation économique.

1 1. La critique de l'égalité formelle par la sociologie

L'école du début du XXe siècle, autant en Europe qu'en Amérique n'incarnait pas le rôle de redistribution des classes sociales ni non plus celui de former le capital humain assurant le développement des forces économiques de la nation. [...]

2 2. Bref rappel historique : une double logique de décentralisation et de concurrence au Québec

Il importe maintenant de revenir plus spécifiquement à la question posée en titre. Qu'en est-il de la vocation d'égalité de l'école publique généraliste au regard des récentes transformations du système d'éducation québécois? [...]

3. Des constats clairs

Les constats qui suivent documentent à la fois l'importance de la concurrence du secteur privé et la réponse des commissions scolaires. [...]

4. 4. Les effets : hiérarchisation des établissements et chute de la diplomation ?

Tant les écoles publiques que les écoles privées se différencient donc de plus en plus dans leur offre de services éducatifs.

Cl. CONCLUSION : UNE DEMANDE DES CLASSES MOYENNES INQUIÊTES

A l'orée du 21^e siècle, parmi les tendances lourdes des sociétés occidentales, il y a certainement l'érosion de la classe ouvrière et la domination politique de la classe moyenne et des fractions qui la composent. [...]

Lessard

Pour les types de marqueurs, lors de l'analyse de la longueur des séries typiques non composées de sections titrées, nous avons montré que les connecteurs et les constructions Num+N introduisent des items plus longs que les anaphoriques et les conjonctions de coordination. Cette tendance est aussi observée dans les marqueurs des séries section et méta. Comme l'illustre la Figure 30 ci-dessous, les anaphoriques et les conjonctions de coordination sont rares dans les séries section et méta, les dernières sont même absentes des séries section. Les connecteurs (expressions cadratives incluses), les constructions Num+N et les items non marqués sont les plus fréquents. Surtout comparées à la fréquence totale dans toutes les séries typiques, les constructions Num+N affichent une fréquence d'utilisation élevée⁸⁰.

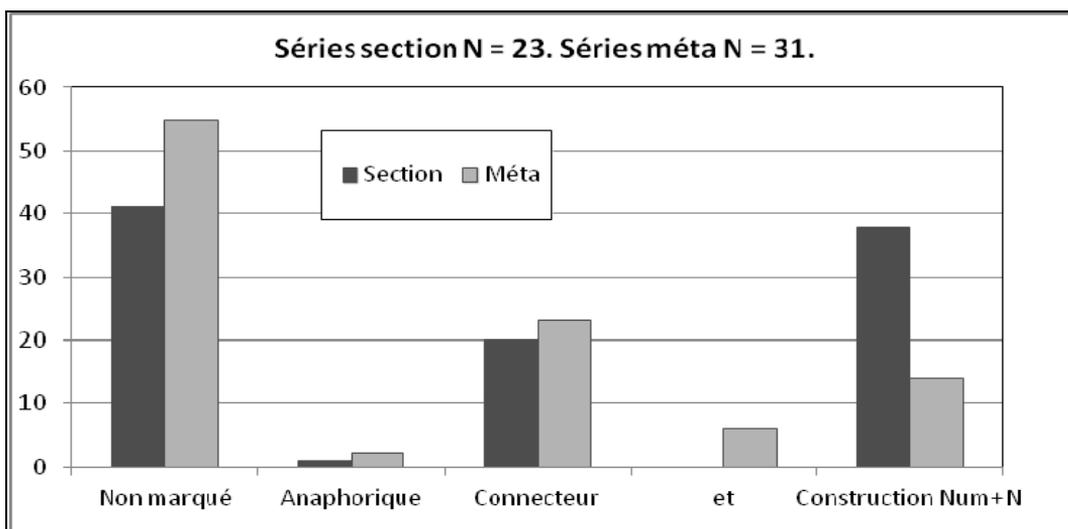


Figure 30 Fréquences des marqueurs des séries section et méta (en pourcentage du total de la catégorie).

⁸⁰ Dans les séries simples typiques, 11 % des items sont signalés par des constructions Num+N alors que c'est le cas de 38 % des items des séries section. De même, dans les séries enchaînées, ces marqueurs sont utilisés dans 14 % des items des séries méta et seulement dans 6 % des autres séries enchaînées.

Comparés aux séries typiques sans titres, les types de marqueurs des séries section et méta comportent pourtant aussi quelques différences. La fréquence des items non marqués s'explique par les séries PNM dans les séries section et par les séries non marquées dans les séries méta. Notons que la Figure 30 inclut également les marqueurs utilisés dans les parties métadiscursives des séries méta. Cela explique, par exemple, la présence de *et* qui n'est jamais utilisé dans les items extrêmement longs. Enfin, comparés aux autres séries enchaînées longues, les connecteurs sont aussi relativement fréquents dans les méta. La série métadiscursive étant généralement la partie explicitement introduite dans celles-ci, il semble que les connecteurs soient souvent utilisés dans cette fonction pour informer le lecteur de la structure du texte à venir.

En définitive, bien que les séries typiques et les structures composés de sections titrées manifestent les mêmes tendances de marquage de l'organisation textuelle, il semblerait donc que l'introduction des items par des titres affecte quelque peu les façons dont les séries section et méta sont introduites. Pour les séries section, le marquage partiel paraît encore plus fréquent que dans les autres séries simples longues. Dans les méta aussi, les séries complètement non marquées sont plus courantes. En l'occurrence, c'est surtout la série composée de sections titrées qui reste non marquée, tandis que la série métadiscursive est souvent explicitement introduite. Ainsi, les titres semblent contribuer au marquage des séries linéaires de sorte que l'usage des autres marqueurs se fait plus rare.

L'Exemple 56 suivant présente une série section avec une construction Num+N. De même que la majorité de ces séries, elle est partiellement non marquée. La structure commence par une amorce, en fait répétée à deux reprises, d'abord dans la section précédente puis dans l'introduction de la section dont les items-sous-sections font partie. La première partie de l'amorce est vague, indiquant que la série à venir se compose *d'une batterie de tests*, et la seconde partie exacte, définissant le nombre de tests discutés. Les items correspondent aux tests comme indiqué dans les amorces. Seul le dernier est explicitement marqué, par la construction N+N *le dernier test*. La série se termine en une conclusion placée dans la section du dernier item, se référant à tous les tests énumérés : *les résultats des tests ci-dessus...*

Exemple 56

- 0 Dans la section suivante, nous proposons **une nouvelle batterie de tests**, plus élaborés cette fois, visant à donner quelques raisons suffisantes pour traiter les exemples ci-dessus comme des cas d'ellipses.

3.3. Seconds diagnostics : VPE vs anaphore nulle

Déterminer si on est en présence d'une ellipse, dans le sens d'un matériel syntaxique/sémantique qui a été supprimée, ou d'une anaphore (nulle) n'est pas chose facile. Néanmoins, il semble exister au moins **trois tests permettant de mettre au jour** cette distinction. De manière générale, **ces tests** tendent tous à souligner la présence d'un constituant syntaxique dans la proposition cible. [...]

- 1 3.3.1. Recyclage d'n antécédent
 Une différence fondamentale entre ellipse et anaphore (nulle) concerne la capacité de « recycler » un matériel absent dans l'antécédent. Comme le signalent Grinder & Postal (1971), seule l'ellipse verbale possède cette propriété [...]
- 2 3.3.2. Les constructions ACD
 Comme nous l'avons déjà souligné plus haut à § 2 (cf. Abeillé, 1991), l'ellipse modale partage avec la VPE anglaise, le fait d'être licite dans les constructions relatives ; en particulier, dans les constructions où le SV élidé est contenu dans le SV antécédent. [...]
- 3 3.3.3. Le Pseudo-gapping
Le dernier test est un peu moins direct, puisqu'il se fonde sur un exemple de Pseudo-Gapping : [...]
 Cl. **Les résultats des tests ci-dessus** sont résumés dans le tableau suivant [...]
 Busquets

7.2.5 BILAN : MARQUAGE DES SERIES LINEAIRES DE LONGUEURS DIFFERENTES

Nos résultats indiquent donc un lien entre la longueur de la série et la façon dont elle est signalée. Dans les séries longues, le marquage partiel et l'usage des marqueurs relatifs sont fréquents. Pour les séries plus courtes, le marquage se fait généralement plus explicite et plus exact. De plus, dans les séries très courtes dont le nombre d'items est annoncé dans la proximité, soit dans l'amorce, la clôture ou l'autre partie de la série enchaînée, le non marquage complet est fréquent. Il est sans doute dû au risque de répétition excessive du marquage. Dans ces séries, les items sont souvent bien délimités et faciles à repérer, ce qui peut également contribuer à leur non marquage.

Outre le marquage partiel et l'usage des marqueurs relatifs, les types de marqueurs proposent un deuxième facteur dans l'introduction des séries de longueurs différentes. De fait, ils ne sont pas utilisés pour signaler des items de la même longueur. Selon notre analyse, ce sont les anaphoriques qui sont utilisées pour marquer des items plus courts, tandis que les connecteurs et les expressions cadratives introduisent des segments plus longs. Les constructions Num+N, enfin, sont les plus fréquentes dans les séries les plus longues. Or, les types de connecteurs comportent également certaines différences additionnelles et, surtout, l'origine sémantique semble influencer le contexte d'utilisation de ces marqueurs : les marqueurs en *-ment* signalent des items plus longs, ceux en *en x lieu* de longueur intermédiaire et les autres connecteurs, souvent d'origine temporel, plus courts.

Pour les expressions anaphoriques et les connecteurs, nos résultats sur les marqueurs individuels confirment ceux de travaux antérieurs (sur les anaphoriques : Ariel 1990 ; Schnedecker 2000 ; sur les connecteurs : Jackiewicz et Minel 2003 ; Ho Dac 2007). La longueur des items signalés par des constructions Num+N, en revanche, est plus inattendue. Nous l'avons expliquée par la nature plus discrète de ces indices et par le fait

que, selon nos résultats, le marquage de l'organisation textuelle est plus implicite et imprécis dans les segments plus longs. Ainsi, les constructions Num+N seraient les mieux adaptées à ce niveau de texte. Certes, nous devons souligner que dans l'analyse de cette section, nous nous sommes concentrée sur les usages les plus fréquents des séries linéaires et leurs marquages. Par conséquent, nos résultats ne sont pas très informatifs en regard de l'univers des possibles, ce qui complique la comparaison de nos résultats avec une partie des travaux antérieurs. Nos résultats, par exemple, ne se combinent pas facilement avec l'hypothèse de Charolles (1997) sur la capacité d'extinction des expressions cadratives au-delà de la phrase dont elles font syntaxiquement partie. Bien que nos résultats indiquent que ce fonctionnement n'est pas fréquent, nous ne pouvons pas en conclure que ce soit impossible.

Finalement, la dernière partie de cette section a été consacrée aux séries composées de sections titrées, aux séries section et méta. Les résultats de cette analyse ont indiqué que les séries métadiscursives énumérant les thématiques des sections à venir sont surtout utilisées pour annoncer des sections relativement longues ; dans les plus courtes, le marquage simple des items semble suffire. De même, dans les méta, les parties composées de sections sont souvent non marquées ; la répétition du marquage avant la série et au début de chaque item est rare. Les résultats indiquent également que les titres semblent contribuer à l'introduction des séries linéaires de manière à diminuer le besoin de marquage explicite. En effet, les séries mixtes autrement fréquentes dans les séries longues sont presque absentes des séries section, en grande majorité partiellement non marquées.

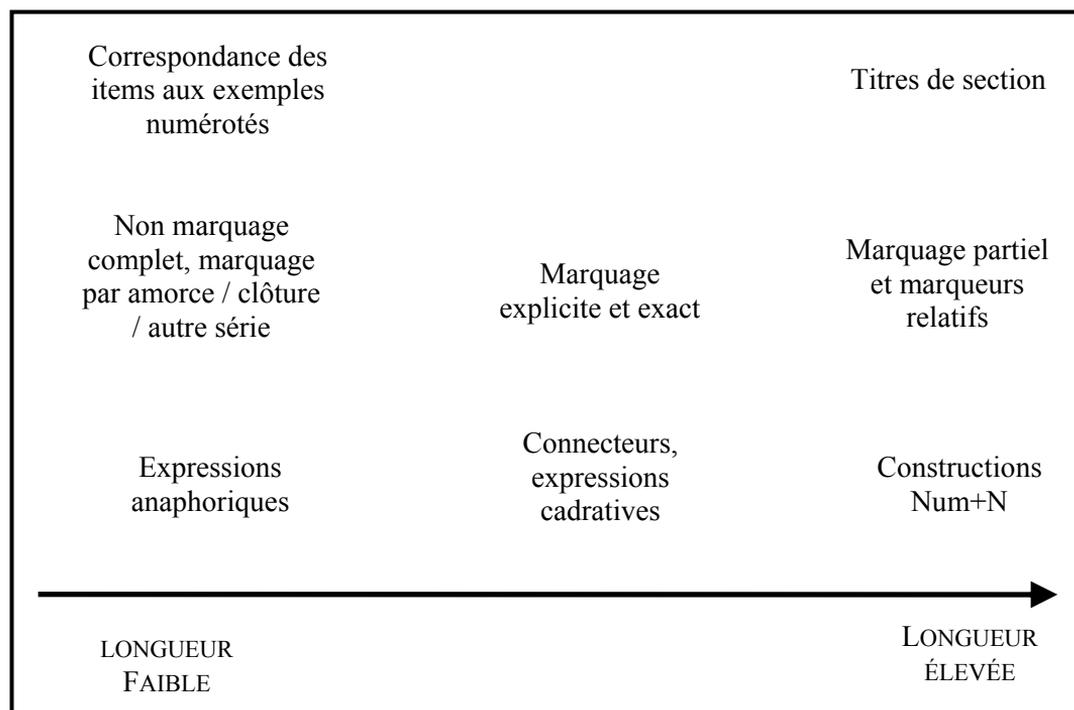


Figure 31 Variation du marquage des séries linéaires selon la longueur de la série.

Pour terminer, la Figure 31 résume les types de marqueurs et le degré d'exactitude typiques ainsi que les indices complémentaires fréquemment utilisés dans les séries linéaires courtes et longues. Cette figure présente bien les nombreux aspects du marquage d'organisation textuelle sur lesquels la longueur du segment à organiser a un effet.

7.3 IDENTIFICATION DES ITEMS NON MARQUÉS ET INDICES COMPLÉMENTAIRES

Dans les sections précédentes, nous avons étudié plusieurs facteurs sous-jacents du marquage et du non marquage de l'organisation textuelle. Entre autres, nous avons expliqué le non marquage et le défaut de marquage exact par la stratégie textuelle adoptée, par le marquage explicite de l'organisation ailleurs dans le texte et par les indices complémentaires contribuant au guidage du lecteur de sorte que les marqueurs explicites deviennent inutiles.

Jusqu'ici, notre examen du non marquage s'est pourtant concentré sur les analyses quantitatives. Les aspects nécessitant une étude manuelle, tels que la cohésion lexicale et le rôle du parallélisme dans les items non marqués, n'ont pas ainsi été étudiés en détail. C'est notre propos dans cette section. Nous privilégions les items non marqués et leur identification en examinant des indices facilitant leur interprétation comme partie de la série linéaire. Ces traits, c'est-à-dire les indices complémentaires de sériation, peuvent être variés ; leur caractéristique commune est qu'ils guident le lecteur dans l'identification de l'item sans indiquer l'addition ou l'ordre. De même que les amorces et clôtures dans les séries courtes, ces indices sont également un facteur potentiel du non marquage ; si leur présence seule suffit pour identifier l'item, les marqueurs explicites peuvent devenir inutiles.

Nous commencerons par présenter les groupes d'indices complémentaires qui, selon notre analyse, semblent centraux pour les items non marqués du corpus : les différentes formes de répétition lexicale, la contiguïté sémantique des items ainsi qu'un groupe d'indices variés comme que certains traits syntaxiques, la correspondance des items aux exemples, etc. Dans un second temps, nous examinerons la fréquence de ces indices dans un extrait de notre corpus. Ce faisant, nous discutons également de l'effet de ces indices sur l'absence de marqueurs explicites.

Notre analyse de cette section nécessite pourtant quelques précisions. Tout d'abord, le premier point concerne le but de notre analyse. La question de la cohérence et de la cohésion dans la création du texte a été discutée dans une multitude de travaux depuis des années (cf. articles dans Sözer 1984 ; Enkvist 1985 ; Tomlin 1987 ; Bublitz, et al. 1999). En analysant le manque de marquage explicite des séries linéaires, cette section participe à cette discussion. Eu égard à l'importance et à la complexité de ces phénomènes, il est cependant clair que leur analyse détaillée et complète est impossible dans le cadre de notre

travail. Pour cette raison, nous soulignons que notre but se limite simplement à relever les indices principaux qui, selon notre analyse, semblent centraux dans l'identification des items des séries linéaires. En général, ces indices peuvent être définis comme des aspects de cohésion. Un examen complet de tous les indices de cohésion et du rôle de la cohérence dans l'identification des items non marqués devrait pourtant faire l'objet d'un travail ultérieur.

Une deuxième précision concerne le parallélisme des items. Lors de la discussion sur le parallélisme, dans la partie théorique, nous avons noté que les séries linéaires sont souvent définies comme composées d'items équivalents ou parallèles, soit de nature similaire et introduits par le même type de marqueurs. Présentant les résultats de l'analyse quantitative dans les sections précédentes, nous avons également noté que nos méthodes ne permettent pas une analyse quantitative de cette propriété. Cette section offre encore un autre aspect du fonctionnement de cette propriété dans les séries linéaires, le parallélisme des items pouvant être lié à un nombre d'indices différents contribuant également à l'identification des items non marqués. Il convient pourtant de noter que, de même que les précédentes, cette section ne vise pas à classer les séries comme parallèles ou non. Le parallélisme est plutôt envisagé comme une caractéristique possible des items non marqués susceptible de fortement contribuer à leur identification.

Enfin, comme cette section nécessite une analyse manuelle, nous avons privilégié une partie restreinte du corpus autorisant un examen plus détaillé des items non marqués. Pour les séries enchaînées, ce sous-corpus est très similaire à celui analysé dans les sections précédentes ; il comprend les catégories de séries les plus fréquentes, c'est-à-dire les HOMO+NM, les MIX+NM, les NM+PNM et les NM+NM. Le sous-corpus comprend toutes les occurrences de ces catégories, à l'exception des séries à deux temps et des séries composées plus de deux séries inférieures. Comme dans les sections précédentes, nous rassemblons toujours sous le titre NM les séries complètement non marquées et celles composées d'items non marqués dont le dernier est signalé par une conjonction de coordination (les NMet). En ce qui concerne les séries simples, nous avons analysé toutes les séries NM et NMet ainsi que 74 séries partiellement non marquées typiques et six correspondant aux sections titrées. Ainsi, bien que plus restreinte que celle de l'ensemble du corpus, cette sélection offre une perspective relativement ample du non marquage dans ces structures. Les chiffres d'occurrences des séries dans chaque catégorie du sous-corpus sont présentés dans le Tableau 6.

Catégorie de série	Nombre de séries
Séries simples	
NM	97
PNM	80
Séries enchaînées	
HOMO+NM	41
MIX+NM	13
NM+PNM	20
NM+NM	22

Tableau 6 Sous-corpus examiné dans cette section

7.3.1 REPETITION LEXICALE

La répétition lexicale est l'un des indices complémentaires de sériation déjà défini dans le schéma d'annotation (cf. 5.2.2). Outre la répétition exacte du même mot, elle est annotée pour les items reliés par *la répétition lexicale complexe* (Hoey 1991) ; la reprise peut s'établir entre deux mots tant que le même morphème de base y est répété. Par exemple, seraient ainsi annotés les items présentant les mots *étude* et *étudier*.

Déjà, l'annotation de tout le corpus nous a révélé que, malgré la création de cohésion, la simple répétition de mots individuels ne permet nécessairement d'identifier des items, ni ne rend la structure parallèle. Surtout, dans les séries longues, les mots isolés n'y suffisent pas. Ainsi, dans l'Exemple 57, les mots *vente* et *vêtement* sont répétés dans tous les items dont le premier est non marqué et ces deux mots correspondent à la thématique principale de la série. Néanmoins, nous estimons que cette seule répétition n'indique pas l'appartenance des items à la série et ne permet pas leur identification. Plutôt, celle-ci est assurée par leur correspondance au principe d'énumération implicite, soit aux différents types de ventes de vêtements. Au lieu de mots répétés, l'expression naît de l'énumération des organisateurs et lieux de ces ventes : *grossistes, ventes des vêtements mis en nantissement, les Domaines, l'administration des Douanes, le commerce en neuf*.

Exemple 57

- 1 Pour le **vêtement** civil, les grossistes trouvent dans **les ventes** judiciaires et volontaires opérées suite à des liquidations, des faillites ou des décès le moyen de se procurer des lots conséquents. **Ces ventes**, faites aux hôtels ou chez les particuliers, annoncées par voie d'affiche et par les Petites Affiches, s'opèrent quotidiennement et les grossistes y sont les principaux acheteurs.
- 2 Viennent **ensuite**, par leur importance, **les ventes des vêtements** mis en nantissement au Mont-de-Piété. Elles sont essentiellement fréquentées par des marchands, au point que les particuliers ont à se plaindre de leur présence envahissante. [...]
- 3 Les Domaines fournissent **eux aussi** quantité de lots au commerce de l'occasion. À chaque changement de saison, les hôpitaux parisiens organisent **la vente** de quantités énormes de **vêtements** en souffrance. [...]

- 4 L'administration des Douanes ou les hospices organisent, **eux aussi**, régulièrement des **ventes** considérables.
- 5 **Enfin**, dès avant les années 1830, le commerce en neuf et l'industrie de la confection fournissent leur lot de **vêtements** détériorés par les inondations ou les incendies, passés de mode ou comportant des malfaçons. [...]

Charpy

Or, bien que la répétition ne soit pas toujours cruciale pour l'identification des items non marqués, l'analyse manuelle a montré qu'elle peut également l'être, voire créer un certain parallélisme entre les items. Pour ce faire, les mots répétés doivent pourtant jouer le même rôle syntaxique dans les items. De plus, « l'effet répétitif » semble plus fort si l'élément répété est placé à l'initiale de la phrase ou si un syntagme complet est répété. L'Exemple 58 ci-dessous décrit deux séquences d'enseignement dans deux écoles. Les deux items, non marqués, commencent par des expressions cadratives de forme *dans l'école* + *NOM*. La série correspond également au modèle de thèmes dérivés de Danes (1974).

Exemple 58

- 0 Nous avons eu la chance d'assister à la même séquence dans les **deux classes** qui constituent les bornes de notre échantillon sur le plan du recrutement social. [...]
 - 1 **Dans l'école Jean Jaurès**, l'histoire choisie avait pour titre « Mon ours bleu de toutes les couleurs ». [...]
 - 2 **Dans l'école « Le Prieuré »**, le maître avait choisi un livre intitulé « Où est maman ? » ; cette histoire parlait de trains, de wagons, de rails, de gare..., termes qui ne semblaient pas surprendre les élèves. Il s'agissait de savoir « pourquoi » le chapeau de maman s'était envolé. [...]
- Cl. Le tableau suivant récapitule les différences dans les sollicitations des élèves lors de **ces deux séquences d'enseignement**. [...]

Maso-Taulère

Dans l'Exemple 58 ci-dessus, la répétition des cadratifs au début des items semble décisive pour leur identification. De plus, elle crée une équivalence de forme et de sens dans la série, qui semble résulter en un parallélisme entre les items facilitant encore l'identification des items. Cette forme de répétition correspond également à la façon dont les items sont signalés dans les énumérations pour lesquelles l'ordre n'est pas un critère d'organisation (cf. 4.2.1). Ainsi, la répétition peut être un indice complémentaire très fort et suffire seule pour rendre les items non marqués repérables.

Or, dans l'Exemple 58, nous pourrions dire que ce sont les cadratifs comme marqueurs de cohésion, et non la répétition, qui ont cet effet. L'Exemple 59 ci-dessous montre pourtant que la répétition peut être cruciale même quand les mots répétés ne correspondent pas aux marqueurs préalablement définis, tels que les cadratifs. Dans cet exemple de série enchaînée, le syntagme nominal *le laid* + *ADJECTIF* est répété dans trois des quatre items, la partie adjectivale changeant dans chacun. De plus, la première série inclut une amorce dans laquelle *le laid* est aussi répété. Dans la première série, les items ne correspondant qu'aux syntagmes nominaux *le laid* + *ADJECTIF*. La seconde série, en revanche, commence par une expression anaphorique se référant au premier item de la série

précédente : *le premier*. Le SN *le laid* + *ADJECTIF* n'est répété que dans le second item qui, autrement, est non marqué.

Exemple 59

0 Il est deux sortes de laid :

1 le laid difforme

2 et le laid hirsute

1 **Le premier** est par exemple le nabot des XXIII manières de vilains, ou celui qui marche toujours courbé sous le fardeau. Dans les contes, il est rare. La laideur due à une mauvaise conformation est réservée à quelques marginaux et subordonnés, tel ce bouvier au service d'un riche vilain [...].

2 Plus fréquent est **le laid hirsute**, qui est une apparence, et qui est le fait de paysans aisés comme des pauvres [...].

Lorcin

Dans l'Exemple 59 ci-dessus, la répétition de *le laid* + *ADJECTIF* contribue sans doute à l'identification des items. Les items de la première série suivant directement l'amorce et présentant simplement les thématiques des items, l'importance de cette répétition est sûrement plus essentielle pour l'ultime item de la série. Or, simultanément, tous les éléments répétés contribuent au parallélisme des items et rendent ainsi la structure plus facile à repérer. Il est pourtant à noter que le parallélisme est incomplet ; le premier item de la seconde série le rompt, le SN *le laid* + *ADJECTIF* en étant absent. Cette répétition partielle, qui s'établit entre une partie seulement des items de la série, est très courante dans les NM+PNM, ainsi l'exemple ci-dessous, dont le syntagme répété est absent dans les items explicitement signalés par un marqueur d'addition ou d'ordre. De même, la répétition d'un même syntagme dans une partie des items est fréquente dans les séries simples PNM.

Bref, la répétition lexicale peut donc revêtir un rôle important pour le marquage des séries linéaires. Or, d'autres formes de cohésion lexicale peuvent aussi être utiles à l'identification des items. Surtout dans les séries méta, en lieu et place d'une répétition exacte, les items de la seconde série incluent souvent une paraphrase⁸¹ de la thématique exposée dans la première série métadiscursive. Dans l'Exemple 60, notamment, les items de la première série décrivent ce qui sera discuté dans les sections de la seconde série.

Exemple 60

0 Dans cet article, nous présentons une étude de faisabilité sur l'annotation des fonctions des passages entre guillemets, réalisée dans le cadre d'un projet de plan-pluriannuel formation (PPF) sur les marqueurs linguistiques de la subjectivité et de la polyphonie.

1 **Dans un premier temps**, en nous appuyant sur la littérature, nous récapitulons **les principales valeurs sémantiques véhiculées par les guillemets**, en faisant l'hypothèse de configurations relativement stables de leurs emplois.

2 **Puis**, nous rapportons **une méthode d'interannotation**, qui permet de tester la faisabilité du projet en cernant les variations d'interprétation et d'ajuster ainsi les

⁸¹ Ici, nous utilisons la notion de *paraphrase* dans son sens usuel, *pour dire en d'autres mots*.

consignes. Nous détaillons alors les indices formels sur lesquels l'annotateur peut en partie s'appuyer et nous affinons la description des valeurs des passages à annoter

3 Nous montrons **enfin la manière dont le codage peut rendre compte** de la plurivocité des guillemets comme propriété interprétative de ces signes

1 1. Les difficultés d'interprétation des guillemets

Pour paraître accessoires, les guillemets n'en sont pas moins nécessaires à la construction du sens. Ordinairement employés pour indiquer une citation, **ils ne se limitent pas à cet usage mais ont ceci de particulier qu'ils bloquent** l'interprétation littérale de l'élément qu'ils entourent. [...]

2 2. Une méthode d'interannotation comme étape de validation du projet

Annoter les PEG dans un corpus de textes variés présuppose un modèle stable. Une phase intermédiaire d'annotation doit permettre de déterminer si un modèle de description linguistique des guillemets est envisageable et c'est dans cette perspective que nous proposons une méthode d'annotation interannoteurs. [...]

3 3. L'élaboration d'une procédure d'annotation

Pour faciliter la cohérence de l'annotation, il est souhaitable de recourir le plus possible à des indices formels, bien qu'à eux seuls ces éléments soient rarement suffisants pour déterminer la fonction du PEG, en tout cas pour certaines d'entre elles. [...]

Rinck

Évidemment, une partie des items de l'Exemple 60 inclut également la répétition de quelques items lexicaux ; les mots *guillemets*, *méthode* et *annotation*, par exemple, sont répétés dans les items des deux séries. A notre avis, elle demeure pourtant secondaire comparée à l'importance des paraphrases. Ainsi, le premier item de la première série indique que le premier item-section discutera les *principales valeurs sémantiques [...] des guillemets*, ce qui correspond à l'énoncé du début de cette section, commençant une discussion sur ces valeurs : *ils ne se limitent pas à cet usage...* Le lien entre les deuxièmes items est encore plus clair grâce à la répétition du SN *méthode d'interannotation* dans les deux items. Or, dans ces items aussi, le premier item propose que la *méthode d'interannotation [...] permet de tester la faisabilité du projet*, ce qui correspond au titre *une méthode d'interannotation comme étape de validation du projet*. Le dernier item de la première série, enfin, indique comme sujet de discussion le codage des valeurs des guillemets, ce qui paraît déjà dans le titre de l'item-section *élaboration d'une procédure d'annotation* et encore dans le texte même.

7.3.2 CORRESPONDANCE DES ITEMS AU PRINCIPE D'ENUMERATION ET CONTIGUÏTE SEMANTIQUE

Dans les sections précédentes, nous avons plusieurs fois noté que les séries linéaires sont souvent définies comme étant composées d'items similaires ou équivalents (cf. 4.2.4). Outre la répétition de mots ou de syntagmes dans les items, cette similarité peut être sémantique, provoquée par l'appartenance des items « à la même catégorie », et explicitée

par une amorce ou une clôture avec une hypéronyme décrivant le principe d'énumération de la série. Cette annonce facilite également l'interprétation des items non marqués puisqu'elle indique explicitement au lecteur la nature des items qu'il doit trouver dans le texte (cf. Tadros 1994). Par exemple, dans l'Exemple 61 ci-dessous, l'amorce *trois de ces contributions majeures* explicite le fait que les items énumérés correspondent aux travaux scientifiques.

Exemple 61

- 0 Pour notre part, nous appelons l'attention préalable des lecteurs à **trois de ces contributions majeures, présentées ici** dans un ordre régressif.
 - 1 **L'ouvrage d'Étienne Karabétian (2000)**, au-delà même de la précieuse remise en perspective historique qui le motive, et grâce à elle, réduit le sentiment décourageant de l'hétéroclite au profit de la légitimité d'un champ humaniste aussi prodigieusement profus que fécond ; il rend ainsi à la fois plus prudents les enthousiasmes et plus enthousiastes les adhésions critiques.
 - 2 **L'entreprise de reconception de Jean-Michel Adam (1997)**, permet aux linguistes du discours de revenir pour ainsi dire « collectivement », comme courant, vers le champ conflictuel de la stylistique ; tâche nécessaire, tant ce champ, contrairement aux prédictions de Michel Arrivé (1969), reste et/ou est redevenu ardent, comme le présent volume en témoigne aussi.
 - 3 Quant à **l'essai de Laurent Jenny (1990)**, sa place dans cette triade paraîtra parfois plus surprenante ; il n'y est guère question de style et pas du tout de stylistique ; pourtant, c'est un éclairage essentiel, venu de la philosophie « genevoise » du langage, sur une des manières (celle qui me meut) dont on peut traverser et baliser le territoire des stylistiques.
- Cl. L'intérêt de **ces trois références**, outre leur grande valeur particulière, est de renvoyer toutes trois, et chacune sur un mode différent, à une lecture critique de Bally (1909 notamment).

Viprey

La contiguïté sémantique des items pourrait également être interprétée comme un type de répétition ; déjà Halliday & Hasan (1976 : 274 et passim) comptent ce lien comme un type de *réitération* capable de lier aussi bien des synonymes que des répétitions exactes de mots ou des hypéronymes et hyponymes. Toutes ces formes de répétition font donc partie de la cohésion lexicale. Nous devons pourtant noter que, par exemple, Halliday et Hasan (*ibid.*) examinent la cohésion lexicale surtout entre des mots individuels, tandis que la contiguïté sémantique des items repose sur des syntagmes, des phrases voire des segments textuels plus vastes. Pour cette raison, au lieu de la répétition, il est préférable de fonder l'analyse de cette similarité sur les notions de principe d'énumération et de contiguïté sémantique, utilisées dans des travaux sur les séries linéaires (cf. section 4.2.1).

Dans l'Exemple 61 ci-dessus, l'amorce et la clôture contribuent effectivement à la lecture de la série mais elle serait sans doute aussi cohérente sans elles ; les items sont également entendus comme similaires grâce à l'énumération des noms d'auteurs et des années de parution de leur ouvrage dans chaque item. Ainsi, l'annonce explicite du principe d'énumération ne s'avère pas aussi importante que dans les séries dont les items ne sont

pas aussi facilement identifiés. Dans l'exemple suivant, notamment, la présence de la série et des items apparaît uniquement dans la clôture. Malgré la présence de la paire *analyser – analyse* dans les items, seule la clôture, se référant aux items et en annonçant le principe d'énumération, rend la structure explicite. En son absence, toute la structure serait difficile à interpréter comme une série linéaire.

Exemple 62

0 Or que faire d'exemples comme (9) ?

1 On pourrait essayer **d'analyser** les « pourtant » qu'ils contiennent, en les ramenant à des « donc » qui seraient concédés.

2 Pour ma part, je ne ferai pas cette **analyse** car les enchaînements en « pourtant » me semblent réaliser des structures tout à fait parallèles à celles des enchaînements en « donc ».

Cl. **Dans les deux cas**, il s'agit [...].

Carel

Dans l'exemple, la clôture *dans les deux cas* indique donc au lecteur que les actions possibles décrites dans les phrases précédentes, soit l'analyse des connecteurs *pourtant* et l'abandon de cette analyse, forment une série linéaire. Bien que l'hypéronyme pourvu par la clôture, *cas*, soit très imprécis, il semble tout de même réunir les items en leur proposant un concept plus général et, ainsi, faciliter leur identification. Or, bien qu'elle contribue à l'identification des items de la série et y joue même un rôle crucial, la clôture ne semble pas rendre les items plus parallèles. Ainsi, l'annonce explicite de la contiguïté sémantique des items ne semble pas pouvoir adopter cette fonction.

Outre les amorces et les clôtures, le principe d'énumération peut également être annoncé par les constructions numéro+nom avec un hypéronyme décrivant les items de la série. L'identification des items non marqués s'en trouve facilitée car, en introduisant explicitement un item de la série, ces marqueurs procurent également une définition vague de tous les autres items, explicitement signalés ou pas. Ce fonctionnement est le mieux illustré par l'Exemple 63, composé de deux items dont le premier est non marqué et le second introduit par la construction Num+N *la seconde grande recommandation*. Jusqu'au début du deuxième item et son annonce du principe d'énumération *recommandation*, la présence d'une série linéaire est, selon notre analyse, difficile à interpréter.

Exemple 63

1 Dans l'ouvrage *Préparer à l'histoire au cycle 2* (Loison, 2003, 176), nous avons attiré l'attention sur le fait que pour favoriser les questionnements comparatifs, **il est nécessaire que les élèves observent et interrogent « des ressources patrimoniales en nette rupture avec leur milieu »**. Heimberg (2002, 85-87), après avoir transposé à l'histoire scolaire l'affirmation provocatrice de Robert Ferras d'un enseignement de la géographie de la maternelle à l'université, ne dit pas autre chose lorsqu'il affirme que **l'intérêt majeur d'un travail sur le terrain est de plonger les élèves « dans un univers mental différent – relatif à un autre contexte ou à une autre époque que la leur – et de faire en sorte qu'ils s'en rendent compte pour en tirer du sens »**. [...]

2 **La seconde grande recommandation portera sur l'explicitation du capital historique.** En son temps, Johnson (1979, 97) relevait que certains enfants ont le privilège de vivre, dans leurs familles, des expériences les mettant au contact des traces laissées par le passé et de se construire ainsi un capital historique, un stock d'images mentales ou de représentations, celui que nous qualifions d'implicite²³. Johnson insistait alors sur le fait que « la majorité des enfants n'ont pas cette chance » et [...].

Loison

Le premier item de l'Exemple 63 décrit deux activités proposées aux élèves, présentées dans deux ouvrages. Toutefois, seule la construction Num+N du second item indique au lecteur que le texte précédent doit être interprété comme une recommandation et que la structure forme une série linéaire, *recommandation* étant le principe d'énumération ; en son absence ou combinée à un autre type de marqueur, l'item ne serait pas nécessairement interprété de la même manière. Ainsi, le marqueur explicite contribue à l'identification de l'item non marqué et guide l'interprétation de toute la série.

Enfin, l'annonce du principe d'énumération par l'amorce, la clôture ou un marqueur d'item peut donc faciliter la lecture de la série et contribuer aussi à l'identification des items non marqués. De plus, dans certaines séries, le principe d'énumération et l'appartenance des items à la même catégorie sémantique semblent importants même en l'absence d'annonce explicite. En général, les items non marqués de ces séries incluent peu d'autres indices complémentaires permettant leur identification. De là une importance accrue de leur contiguïté sémantique. Ainsi, dans la série de l'Exemple 64 ci-dessous, les items semblent être le mieux reconnus grâce à leur correspondance avec le principe d'énumération, soit avec les propriétés des gares comme lieux complexes dont le propriétaire est difficile à déterminer. Évidemment, dans cette série aussi, quelques répétitions relient des mots individuels. Néanmoins, comme dans les exemples précédents, nous ne pensons pas que ce soit le trait principal autorisant l'identification des items non marqués. Notre analyse penche plutôt pour leur appartenance à la même catégorie sémantique.

Exemple 64

- 1 En effet, le sol sur lequel furent **construites** les **gares** reste la propriété de **l'État** ou de la Ville de Paris et ce sol est simplement concédé à une **compagnie** ferroviaire, constituée la plupart du temps en société anonyme.
 - 2 La **compagnie construit** les bâtiments, parfois avec l'aide subsidiaire de **l'État**, ce qui fait qu'elle se sent largement propriétaire de « sa » **gare**.
 - 3 Enfin, **le lieu** est ouvert largement au public des voyageurs qui l'arpentent, l'habitent (même transitoirement) et l'investissent de nombreuses manières.
- Cl. Quel statut peut donc légalement définir la **gare** du XIXe siècle ?

Sauget

En définitive, en l'absence d'annonce explicite, l'identification du principe d'énumération des items repose plutôt sur l'interprétation du lecteur, soit la cohérence, que sur des indices linguistiques. De même dans l'Exemple 64 ci-dessus. L'annonce explicite du principe d'énumération, en revanche, se rapproche davantage de la cohésion, bien qu'elle comporte

aussi certains aspects de la cohérence ; les items doivent toujours être entendus comme similaires selon le principe annoncé, grâce à l'interprétation du lecteur en l'absence de signalisation par d'autres indices linguistiques complémentaires.

7.3.3 AUTRES INDICES COMPLEMENTAIRES

Dans les deux sections précédentes, nous avons présenté quelques modes de fonctionnement des différents types de répétition et de la correspondance des items au principe d'énumération en tant qu'indices complémentaires, pour ainsi contribuer à l'identification des items des séries linéaires. Dans cette section, nous étudierons encore quelques autres indices susceptibles d'adopter cette même fonction : la correspondance des items aux exemples cités par l'auteur et les marqueurs non annotés comme des indices de sériation explicites. De plus, nous montrerons que certains traits syntaxiques peuvent aussi faciliter l'identification des items non marqués.

Tout d'abord, nous avons déjà illustré comment les items des séries linéaires peuvent correspondre aux exemples de l'article, cités et numérotés par l'auteur (cf. entre autres l'Exemple 39 à la page 155). Dans ces cas, plutôt fréquents dans le corpus, les items sont évidemment facilement repérables et bien délimités malgré le non marquage ; les limites des exemples indiquent aussi celles des items. De plus, un autre type d'indice complémentaire est fourni par un groupe de marqueurs. Il s'agit souvent de connecteurs ou d'expressions cadratives mais qui ne signalent pas l'addition ou l'ordre et ne sont, par conséquent, pas analysés comme marqueurs explicites. Par exemple, dans les séries à deux temps (cf. 6.1), les connecteurs thématiques, tels que *quant à* ou *en ce qui concerne*, introduisent l'item bien qu'il soit annoté comme non marqué. Les indices temporels et spatiaux peuvent également adopter cette fonction. L'Exemple 58 a déjà illustré la façon dont les cadratifs spatiaux peuvent contribuer à l'identification des items non marqués. L'exemple suivant, en revanche, présente le fonctionnement d'un indice temporel dans la série. L'exemple, énumérant des envois des missionnaires de Paris aux départements au XVIIIème siècle, est organisé autour d'un axe temporel et ses items suivent l'ordre temporel des envois. Ceux explicitement signalés, dont les longueurs varient de quelques mots à une phrase complexe, sont en général introduits par des connecteurs temporels, indiquant que l'envoi présenté a eu lieu après celui de l'item précédent. D'une façon analogue, l'item non marqué, selon notre schéma d'annotation, commence par une expression temporelle *En février-mars*. Ainsi, c'est grâce à cet indice que l'item non marqué est identifié.

Exemple 65

- 1 C'est tout d' **abord, en janvier 1795**, Duport qui est envoyé en Seine-Inférieure afin de tout à la fois surveiller l'axe fluvial fondamental pour la circulation des grains importés au Havre à destination de Paris, et assurer le propre ravitaillement de ce département.
- 2 **Puis, quelques jours plus tard**, une mission collective apparentée à celles de l'été 1793 (pour la forme tout au moins) est décidée : Froger-Plisson, Jacomin, Lequinio et Soulignac se voient confier les fameux « départements voisins de Paris ».

- 3 Cette mission collective, toujours dans la tradition de l'été 1793, est **ensuite** complétée par les envois de Bernier
 - 4 **puis** de Hourier.
 - 5 **En février-mars**, Laurent et Fleury viennent compléter les équipes missionnaires,
 - 6 **puis** Castilhon est envoyé au Havre tandis que Blaux part pour la Somme.
- Cl. **En deux mois environ**, ce sont **donc** onze missionnaires qui quittent Paris pour les départements sensibles qui entourent la capitale, là où trois seulement avait reçu ce type de mission spécifique dans le second semestre de 1794, preuve d'un évident changement de cap.

Biard

Outre les expressions temporelles et spatiales et les connecteurs thématiques déjà présentés, d'autres types de marqueurs peuvent aussi contribuer à l'identification des items non marqués. Dans l'Exemple 62, notamment, *pour ma part* indique clairement le début de l'item. Plutôt qu'un groupe défini, les marqueurs pouvant adopter ce fonctionnement sont donc très divers et dépendants du contexte. En pratique, pourtant, ce sont le plus souvent des indices temporels fonctionnant dans des séries temporelles ainsi que des connecteurs thématiques.

Certains traits syntaxiques, le troisième type d'indice complémentaire discuté dans cette section, sont encore plus difficiles à définir d'une façon holistique que ceux que nous venons de décrire. Surtout dans les séries non marquées et NMet, l'énumération des items courts mais non marqués dans une seule phrase facilite leur identification, en particulier quand ils sont directement précédés ou suivis d'une amorce ou d'une clôture et quand leurs fonctions syntaxiques sont similaires. D'une certaine manière, la similarité des fonctions syntaxiques des items les rend parallèles tout comme la similarité à partir d'autres critères. Dans l'Exemple 66, les deux items correspondent aux syntagmes verbaux coordonnées. Ils font ainsi partie de la même phrase. De plus, ils sont directement suivis d'une clôture s'y référant. Grâce à cette structuration, les items sont bien délimités. Selon notre analyse, ils sont plus faciles à identifier pour le lecteur que s'ils étaient placés dans des phrases séparées et distinctes l'une de l'autre et si leurs structures syntaxiques différaient de manière importante.

Exemple 66

- 1 D'une certaine manière, elle propose plutôt de les poursuivre **en englobant** une périodisation plus longue (1900-1960)
 - 2 **et en abordant** la question de l'identité étudiante selon une perspective genrée.
- Cl. **Ces deux choix** se répondent : une identité, qu'elle soit étudiante ou autre, n'est jamais définitive.

Fish

Évidemment, les traits syntaxiques seuls ne suffisent pas pour indiquer au lecteur les items non marqués ; dans l'Exemple 66, entre autres, l'anaphore résomptive de la clôture y contribue également. Or, comparé aux cas dans lesquels les items ne sont pas énumérés à l'intérieur d'une phrase, nous pouvons avancer que leur emplacement syntaxique facilite le repérage des items aussi.

7.3.4 FREQUENCES DES INDICES COMPLEMENTAIRES DANS LE CORPUS

Les sections précédentes ont présenté les indices complémentaires principaux relevés par notre analyse manuelle. Cette section se propose comme but d'en examiner les fréquences dans les différentes catégories de séries linéaires. Cette analyse permet également d'évaluer l'effet possible des indices complémentaires sur le non marquage. Le non marquage pourrait-il, dans certains cas, être provoqué par les indices complémentaires suffisants pour guider le lecteur ? En étudiant les fréquences des indices complémentaires, nous ne cherchons pas à classer les séries du corpus par catégories : un tâche impossible car une série peut simultanément inclure plusieurs traits contribuant à l'identification de ses items. Pour cette raison, nous examinons simplement la présence des indices différents dans chaque série, un reflet plus fidèle de la façon dont les items peuvent être identifiés par le lecteur (pour d'autres méthodes possibles, cf. Taboada 2009).

Nous commencerons par étudier la fréquence des différentes formes de répétition lexicale dans les catégories de série. Dans un deuxième temps, nous nous focaliserons sur la contiguïté sémantique des items. Finalement, nous discuterons du rôle des traits syntaxiques, connecteurs et exemples numérotés pour le guidage du lecteur dans les catégories de séries linéaires examinées. Cette section privilégie les séries typiques non composées de sections titrées. Les séries section et méta, dans lesquelles les titres ont un rôle important dans le marquage des items, seront analysées séparément à la section suivante.

Avant de présenter les résultats, une remarque sur la nature des indices complémentaires s'impose. L'analyse des sections précédentes a déjà montré que tous les indices repérés ne guident pas le lecteur d'une manière aussi efficace ; tandis que certains, tels que la répétition proprement dite, peuvent seuls indiquer au lecteur la présence de l'item, d'autres se limitent à faciliter son identification. Tous les indices ne sont donc pas aussi efficaces comme marqueurs complémentaires, un élément à prendre en compte pour leur analyse⁸².

Nous avons décidé de diviser les indices analysés en deux groupes. Ainsi, en calculant des fréquences d'indices complémentaires, nous disposons pour chaque série identifiée d'au moins un indice complémentaire *primaire* qui pourrait déjà seul suffire comme indice d'item pour le lecteur. Ces indices correspondent à la *répétition proprement dite*, à la *paraphrase*, à la *contiguïté sémantique* annoncée par l'amorce, la clôture ou une construction Num+N, et à la *correspondance des items aux exemples numérotés dans l'article*. Plusieurs de ces indices peuvent également être présents simultanément dans une seule série et, sans doute, quelques uns peuvent rendre le non marquage inutile, au moins dans certains contextes. En revanche, si aucun de ces indices n'est présent dans la série,

⁸² Le statut différent de certains indices de cohésion a déjà été noté également par Morris & Hirst (2004) pour certains indices de cohésion lexicale et par Piérard et Bestgen (2006) pour les marqueurs de segmentation, dont les plus forts, selon les auteurs, sont les adverbiaux temporels. Ces résultats ne sont pourtant pas facilement comparables avec notre étude car ces indices sont très dépendants du contexte.

elle se voit attribuée l'indice *contiguïté sémantique implicite*. Il signifie que les items sont sémantiquement similaires et correspondent à un principe d'énumération qui n'est pas déclaré explicitement. De plus, comme il est utilisé uniquement en l'absence d'autre indice primaire, il sous-entend que l'identification des items de la série repose sur des indices très implicites et plutôt sur l'interprétation du lecteur que sur des aspects linguistiques.

Le premier groupe d'indices dont nous analysons les fréquences est les différentes formes de *répétition*. En les examinant, nous avons distingué plusieurs types, suivant l'analyse présentée en section 7.3.1. Ainsi, *la répétition proprement dite* correspond à la reprise de mots ou de syntagmes selon le même rôle syntaxique dans tous les items de la série, comme celle de l'Exemple 58. Cette répétition crée toujours un parallélisme entre les items. *La répétition des éléments différents (Rep. dif.)* établit le même type de lien mais seulement entre une partie des items. Par exemple, le syntagme répété peut être absent de l'item explicitement signalé par l'anaphorique *le premier*, comme c'est le cas dans l'Exemple 59. *La répétition partielle (Rep.par.)* correspond uniquement à la répétition de mots individuels dans les items, insuffisante seule pour leur identification (cf. l'Exemple 57). *La paraphrase (Para.)*, pour sa part, est souvent employée dans les séries métadiscursives afin de présenter les items-sections suivants et est illustrée dans l'Exemple 60. *Nom (Nom)*, enfin, indique la répétition d'un nom propre dans chaque item. Les noms se distinguant des autres SN répétés⁸³, nous les avons analysés comme un trait à part des autres types de répétition. La répétition des noms propres est illustrée ci-dessous dans l'Exemple 67. Chaque item y énumère un nom propre (et une année), sans doute le trait le plus important pour leur identification dans cette série. La répétition de noms propres peut donc suffire à elle seule à l'identification des items. De plus, la structure de la série est encore clarifiée par la clôture qui se réfère aux items tout en indiquant leur nombre et leur classe sémantique : *les trois approches...*

Exemple 67

- 1 **H&S (1976, 1984)** considèrent les exemples tels que (9) comme un cas d'Anaphore de Complément Nul (dorénavant NCA pour Nul Complement Anaphora).
- 2 La théorie proposée par **Chao (1988)** établit, elle **aussi**, une distinction binaire, entre deux classes d'anaphores. [...]
- 3 Plus récemment, **Gardent (1996)** propose **une autre classification** basée sur les traits distributionnels des anaphores, sur les propriétés de l'antécédent (c.-à-d., la possibilité de faire référence à un antécédent linguistique [L], pragmatique [P], voire les deux), et enfin sur le rôle restrictif ou non de la syntaxe [\pm S] vis-à-vis de l'interprétation de l'anaphore.

Cl. Le tableau ci-dessous résume **les trois approches discutées** plus haut.

Busquets

⁸³ Cf., notamment, le schéma d'Ariel (1990), dans lequel les noms propres sont décrits comme facilement repérables même dans des contextes ambigus.

Les fréquences des types de répétition dans le sous-corpus examiné⁸⁴ sont illustrées dans la Figure 32 à l'exception des séries section et méta. La figure présente les pourcentages des séries comportant le trait en question.

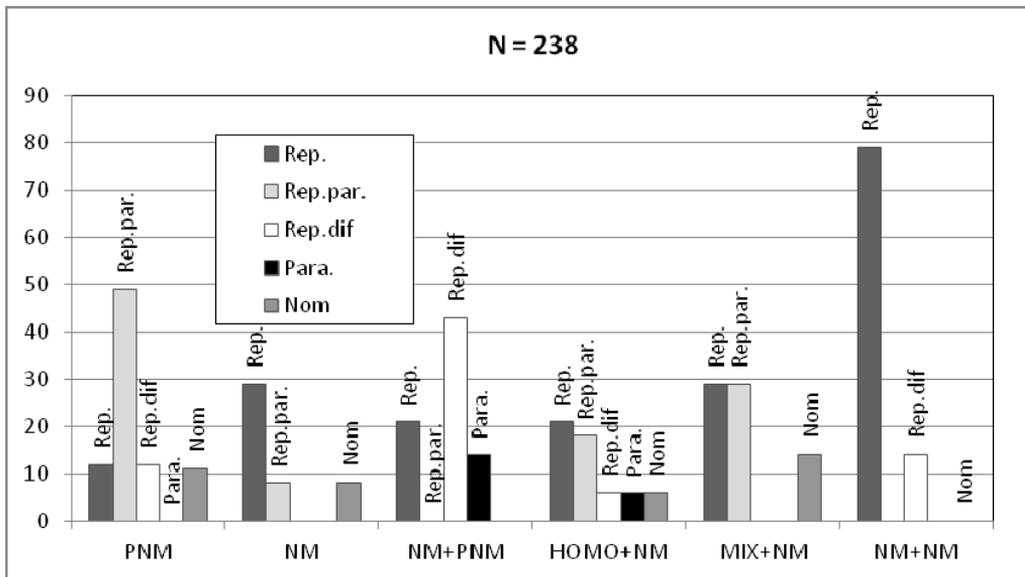


Figure 32 Types de répétition dans les séries (en pourcentage des séries dans lesquelles ce type de répétition est présent). Les NM et sont toujours unies sous le titre NM.

La Figure 32 montre que les types de répétition contribuent différemment à l'identification des items des catégories différentes. Les séries simples partiellement non marquées (PNM) se caractérisent par la fréquence de la répétition partielle, soit la répétition de mots isolés, tandis que la répétition proprement dite y est rare ; comme indice complémentaire, la répétition n'est donc pas très importante dans cette catégorie de série. Dans les séries enchaînées composées uniquement d'items non marqués (NM+NM), la répétition a, en revanche, un effet très important car elle est présente dans la grande majorité des séries. Dans les autres séries, sa fréquence est relativement similaire, de même que celle de la répétition partielle : les deux sont présentes dans 20 à 30 % de ces structures.

La répétition des éléments différents, pour sa part, sépare aussi les catégories. Bien que plutôt rare voire absente dans les autres, elle relie plus de 40 % des séries enchaînées composées de séries non marquées et de séries partiellement non marquées (NM+PNM). Cette fréquence s'explique par le fait que le ou les items explicitement signalés n'incluent pas de syntagme répété comme les autres items de la structure mais sont introduits, par exemple, par l'anaphorique *le premier* (voir l'Exemple 59). La répétition des noms propres ou des paraphrases, enfin, est relativement rare dans toutes les catégories. Leur inclusion comme indices distincts dans l'analyse tend donc plus à expliciter leur présence dans le

⁸⁴ Cf. l'introduction de cette section.

corpus qu'à présenter des indices complémentaires de sériation plus communément utilisés.

Les fréquences de la répétition proprement dite, de la répétition des éléments différents et de la répétition partielle sont, surtout, très instructives sur la façon dont les items non marqués sont identifiés dans les différentes catégories de série. Comme nous l'avons noté, la répétition proprement dite est un indice complémentaire très fort qui peut, seul, rendre les items facilement repérables et parallèles. Ces séries correspondent en général aux énumérations étudiées dans des travaux pour lesquels l'ordre entre les items n'est pas un critère d'organisation important mais qui ne semblent pas nécessairement incomplètes. La répétition partielle, soit la répétition de mots individuels, en revanche, ne suffit pas pour identifier les items non marqués, ni ne contribue à leur parallélisme. La répétition des éléments différents, enfin, est un indice très important pour les items dans lesquels elle est présente. C'est toutefois également le signe d'un certain non parallélisme des items. En effet, elle n'apparaît que dans une partie d'entre eux et ne facilite donc pas l'identification des items dont elle est absente. Ainsi, suivant la Figure 32, le parallélisme des séries créé par la répétition lexicale est le plus fréquent dans les séries enchaînées composées uniquement d'items non marqués (NM+NM) et plus rare dans les séries PNM. Dans les autres catégories, la proportion des séries parallèles est relativement similaire : de 20 à 30 %. Nous devons pourtant noter que la répétition lexicale n'est pas la seule source du parallélisme. La similarité formelle (cf. la suite) des items, entre autres, peut y contribuer.

L'Exemple 68 ci-dessous, entre autres, illustre la répétition proprement dite et le parallélisme des items dans ces séries. Cet exemple est composé de deux séries, dont la première est NMet et la seconde complètement non marquée.

Exemple 68

- 0 Le terme de « pratique » relève de **deux thèmes**,
- 1 du *pragma*
- 2 **et de la praxis**, soulignant l'origine de la pensée humaine et la nature du lien qu'elle entretient avec l'action, entre théorie et pratique.
- 1 La **praxis** s'entend comme lien étroit (immanence) du monde l'action et de la pensée, dans le sens du réalisme de la raison. La praxis suppose une réalisation par l'action, et non une source dans l'action. La praxis hégélienne assemble l'historicité et l'histoire dans la réalisation de la Raison, la réalité étant le développement. [...]
- 2 Le **pragmatisme** relève **pour sa part** d'une enquête sur l'origine de nos pensées dans l'action. Il souligne le rôle créateur de l'action, pour elle-même, ses déterminations et ses valeurs. Les pragmatiques, ou pragmatisme d'expériences, relèvent de la réalisation accompagnée de savoirs par l'action et de la place de formes d'actions dans le savoir. [...]

Morandi

À notre avis, le parallélisme fort des items de l'Exemple 68 relève surtout de la répétition du *pragma* (ou du *pragmatisme*) et de la *praxis* dans les items et du fait que ces noms y sont utilisés dans les mêmes fonctions syntaxiques. De plus, la série inclut plusieurs autres indices complémentaires qui y contribuent aussi. La similarité des fonctions syntaxiques

des items crée un certain parallélisme entre eux et ainsi facilite leur identification, surtout quand ils font partie de la même phrase. L'identification des items est encore facilitée par l'amorce spécifiant le principe d'énumération, la conjonction de coordination *et* dans la première série et le marqueur *pour sa part* dans le dernier item. En fait, cette série ressemble aux séries à deux temps examinées par Porhiel (2007), discutées dans 6.1 ; la première partie présente les items encore thématiquement nouveaux pour le lecteur alors que les items de la deuxième série poursuivent leur discussion en présumant qu'ils sont déjà connus pour le lecteur. La série peut correspondre également au modèle de *thème éclaté* de Danes (1974 : 120).

Dans les sections précédentes, nous avons déjà émis l'hypothèse que la faible longueur et la présence d'une amorce, ou d'une autre série explicitement signalée, pourraient rendre inutile le marquage explicite des items qui pourrait même avoir un effet répétitif. Dans l'Exemple 68 ci-dessus, les indices complémentaires forts ont ce même effet ; ils autorisent l'identification des items et, étant donnée la faible longueur de la série, rendent le marquage explicite excessif. Par conséquent, dans ce contexte et des contextes similaires, les indices complémentaires peuvent être décrits comme des facteurs expliquant le non marquage. Or, il est à noter que cet effet est typique uniquement pour les séries courtes, ainsi les non marquées et leurs variantes enchaînées, dans lesquelles le marquage explicite est toujours dans la proximité.

Ensuite, le deuxième groupe d'indices complémentaires contribuant à l'identification des items non marqués présenté est la contiguïté sémantique. En général, elle est signalée comme un hypéronyme exprimé par l'amorce ou la clôture. Dans l'Exemple 68, notamment, l'amorce *deux thèmes* indique que les items énumérés correspondent à des thèmes.

Ci-dessous, la Figure 33 présente les fréquences des types de contiguïté sémantique analysés. Le plus souvent présent, le *PE*, correspond à la définition du principe d'énumération explicite par une amorce, une clôture ou une construction Num+N (cf. l'Exemple 69). Le *PE implicite*, en revanche, indique qu'il n'est pas explicitement annoncé. Comme nous l'avons précisé en introduction à cette section, il est pourtant défini uniquement pour les séries dans lesquelles il est crucial pour l'identification des items, soit les séries n'incluant pas d'autres indices complémentaires primaires susceptibles de garantir l'identification des items (l'Exemple 64). La fréquence de cet indice dans une catégorie de série indique donc que leur marquage est généralement très implicite et inclut peu de marqueurs linguistiques.

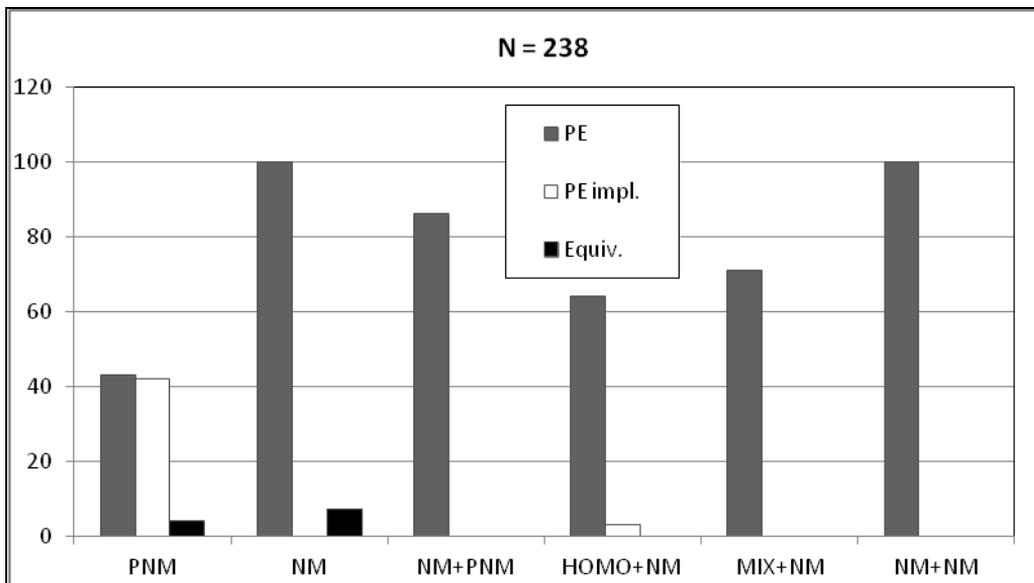


Figure 33 Types de contiguïté sémantique des items (en pourcentage des séries dans lesquelles il est présent).

Enfin, le dernier type d'indice présent dans la Figure 33, *Equiv.*, correspond à l'équivalence structurelle des items. Cet indice, relativement rare dans le corpus, ressemble à la répétition lexicale mais relie des structures énonciatives et non des mots. Par exemple, dans l'Exemple 69 ci-dessous, les items correspondent aux questions, ce qui les rend facilement repérables et toute la structure relativement parallèle. Aussi, l'annonce explicite du principe d'énumération par la clôture contribue à l'identification des items. De la sorte, dans cet exemple également, les items sont si facilement repérables que leur marquage explicite pourrait être inutile.

Exemple 69

- 1 Le maître demande « Combien de bosses le chameau ? »
- 2 « Et qui est-ce qui n'en a qu'une ? »

Cl. Les élèves répondent **aux deux questions.**

Maso-Taulère

Les fréquences des indices *PE*, *PE implicite* et *Equiv.* dans les catégories de série présentées dans la Figure 33 illustrent bien la nature de ces séries. Dans les séries non marquées et les NM+NM, dont tous les items sont non marqués, le principe d'énumération est par défaut indiqué par une amorce ou une clôture. Pour cette raison, l'indice *PE* est présent dans 100 % de ces séries. A l'exception des séries simples partiellement non marquées, cet indice est également très commun dans les autres catégories de série enchaînées. Cette fréquence correspond à nos constatations répétées des analyses antérieures : les séries enchaînées sont relativement explicitement signalées, même en l'absence de marquage.

Les PNM simples constituent une catégorie de série unique dont le principe d'énumération n'est pas explicitement signalé dans leur majorité. En effet, seules 43 % des séries contiennent l'annonce explicite du PE alors que 42 % de ces structures sont analysées comme PE implicite. Ce type d'indice étant noté uniquement pour les séries sans aucun autre type d'indice complémentaire primaire, sa fréquence illustre inévitablement la nature du non marquage de ces séries. Ainsi, elles incluent relativement peu de répétition de syntagmes entiers susceptibles de contribuer à l'identification de leurs items et de rendre la structure parallèle. L'annonce explicite du principe d'énumération ou de la catégorie sémantique des items s'y montre également plus rare que dans les autres séries. L'identification des items de ces séries repose donc plutôt sur l'interprétation par le lecteur de la similarité des items que sur des indices linguistiques.

Ci-dessous, nous illustrons deux types de séries partiellement non marquées, analysées comme PE implicite. Le premier, l'Exemple 70, se compose de trois items dont le dernier est explicitement signalé par le connecteur *enfin*. La série discute de la nature de deux groupes de conseillers, ceux des parlements et ceux du Grand Conseil, et énumère des atouts des conseillers des parlements face à ceux du Grand Conseil. Malgré la répétition des mots *conseiller*, *parlement* et *Grand Conseil* dans une partie d'items, la répétition lexicale ne semble pas contribuer à leur identification. De plus, les items ne sont pas non plus structurellement similaires. De même, leur principe d'énumération n'est pas explicitement annoncée bien que les syntagmes *leur nombre*, *la nomination des conseillers* et *la promotion des conseillers* puissent inclure un faible degré de similarité. Plutôt, l'identification des items semble se baser sur l'organisation thématique de la série et sur la correspondance des items aux sous-thèmes dérivés de l'hyperthème annoncé avant la série.

Exemple 70

... Les conseillers des parlements l'emportent également en dignité sur leurs homologues du Grand Conseil.

- 1 Leur **nombre** est plus élevé et dans presque toutes les causes, la pluralité des personnes est préférée au petit nombre, comme en attestent dix raisons juridiques citées comme autant d'exemples.
- 2 La **nomination** des **conseillers** est entourée d'une plus grande solennité. L'auteur du *Tractatus celebris* en veut pour preuve la procédure de nomination par le roi des conseillers des parlements parmi trois candidats proposés par les cours, système déjà prescrit par les ordonnances anciennes et récemment rappelé par Louis XII dans celle de 1499, comme le fait remarquer Montaigne. À l'inverse, nulle solennité n'est prescrite pour la promotion des **conseillers du Grand Conseil**, simplement pourvus de leur office par lettres.
- 3 **Enfin**, la **promotion** des **conseillers** s'effectue dans un sens favorable aux **parlementaires**. Car les **conseillers du Grand Conseil**, relève Montaigne, sont parfois envoyés par le roi siéger dans les cours de parlement (ad eorum importunam suggestionem!), et non le contraire.

Arabeyre

L'Exemple 71 illustre un autre type de séries PNM que nous dénommons *additives*. De même que toutes les séries linéaires, ces structures énumèrent des éléments souvent similaires, d'une certaine manière. De plus, la composition de la série de deux items, le premier non marqué, et d'une clôture, signalée par un connecteur conclusif ou consécutif, est caractéristique de celles-ci. Ainsi, le deuxième item de ces séries présente souvent un ajout au premier item et la clôture une conclusion ou conséquence de ceux-ci. La série suivante discute du transfert des stratégies cognitives chez les jeunes enfants. Le premier item affirme que l'apprentissage pour la première fois de l'utilisation d'une stratégie est toujours exigeant. Le second item lui apporte un ajout en indiquant que c'est aussi le cas de l'usage de cette stratégie pour une autre tâche. La clôture conclut ces deux items affirmant que, pour les motifs invoqués, le transfert n'est pas « une solution miraculeuse ».

Exemple 71

...Contrairement à ce que l'on pourrait croire, le transfert n'est pas particulièrement économique, principalement du fait de la nécessité de décontextualiser puis recontextualiser la connaissance qui fait l'objet du transfert (Cox, 1997 ; Halpern, 1998 ; Mendelsohn, 1990). Le prévoir dès le début peut certes représenter une relative économie, mais celle-ci réside uniquement dans le fait que le sujet est prévenu dès le départ que le transfert de la stratégie est possible.

1 L'apprentissage à utiliser la stratégie dans la première tâche n'en nécessitera pas moins d'efforts et n'en sera pas moins complexe.

2 **De même**, la réadaptation de la stratégie à la tâche de transfert demandera de réels efforts sans garantie de réussite immédiate.

Cl. Le transfert ne doit **donc** pas être présenté comme une solution « miraculeuse » permettant de s'économiser : c'est tentant car c'est un argument auquel les enfants seraient sensibles, mais c'est tout simplement faux.

Clerc

Comparées aux autres séries PNM dont l'identification se fonde sur le PE implicite, le marquage des séries additives revêt certains traits particuliers. Le non marquage du premier item et le marquage explicite du second, surtout, ont comme conséquence que le dernier est interprété comme un ajout au premier. Cette nature additive concerne également l'identification de l'item non marqué. De fait, dans un tel cas, il est facilement repérable grâce à sa position initiale dans la liste et au marquage explicite de l'item suivant. Par conséquent, bien que sa présence ne soit pas autrement signalée par des indices complémentaires, le lecteur peut la déduire relativement aisément, pour le moins en lisant le marquage explicite du second item. Cette organisation facilitant le repérage de l'item peut également expliquer le non marquage.

Enfin, le dernier groupe d'indices dont nous n'avons pas encore examiné les fréquences inclut des types variés : certains traits syntaxiques pouvant contribuer à l'identification des items, la correspondance des items aux exemples et les marqueurs, tels que les expressions temporelles, ayant la capacité d'introduire les items sans signaler l'addition ou l'ordre. Les fréquences sont présentées ci-dessous dans la Figure 34.

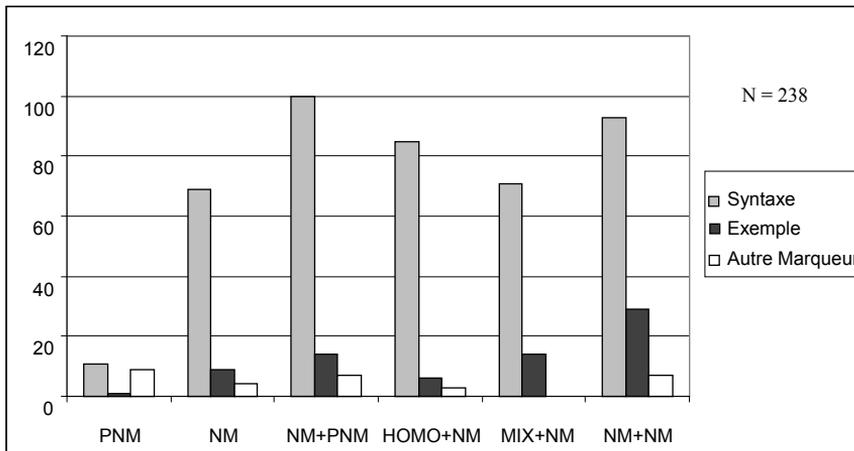


Figure 34 Fréquences des traits syntaxiques, de la correspondance des items aux exemples et des autres marqueurs en pourcentage du total.

Selon la Figure 34, les traits syntaxiques sont un indice complémentaire très fréquent dans la plupart des catégories, présents même dans toutes les NM+PNM. Cette fréquence s'explique par la correspondance de l'autre partie des séries enchaînées avec une seule phrase, dans laquelle les items sont des syntagmes dont l'un peut même être signalé par une conjonction de coordination. Cet indice est donc fréquent dans toutes les séries courtes et non marquées. Or, bien que plus rares, les traits syntaxiques peuvent également contribuer à l'identification des items des séries dont un ou plusieurs items sont explicitement signalés. Dans la série suivante, par exemple, les deux premiers items de la série font partie de la même phrase, commençant par *après avoir*, et le dernier item, formant sa propre phrase, est explicitement introduit par *pour finir*. Le second item est en principe non marqué bien que, dans la pratique, il soit automatiquement repéré grâce à la structure syntaxique qui exprime déjà l'ordre entre les items.

Exemple 72

- 1 **Après** avoir précisé ce que recouvre [le terme] *thématique*,
- 2 cette section examine les propriétés des phrases thématiques vs. celles des phrases athématiques.
- 3 **Pour finir**, nous verrons que l'utilisation des mêmes critères pour caractériser les non-phrases thématiques (vs. les non-phrases athématiques ; section 1.3.) s'avèrera insuffisante.

Porhiel

Contrairement aux aspects syntaxiques, les deux autres traits illustrés dans la Figure 34 sont plutôt rares dans toutes les catégories. Toutefois, les séries dans lesquelles les items correspondent aux exemples forment 21 occurrences dans le corpus. Bien que très minoritaires, la présence seule de ces séries est sans doute caractéristique des articles de recherche. De plus, la correspondance des items aux exemples, le cas échéant, est un indice très fort des items qui peut même motiver le non marquage.

Enfin, le dernier trait illustré dans la Figure 34 rassemble les marqueurs ne signalant pas l'addition ou l'ordre mais pourtant susceptibles de contribuer à l'identification des items. Souvent des expressions spatiales ou temporelles voire des connecteurs thématiques, ces indices sont assez efficaces comme marqueurs d'items non marqués. Leur capacité est pourtant affaiblie par la façon dont ils sont utilisés. En effet, ils fonctionnent très rarement dans l'introduction de tous les items de la série. Pour cette raison, ils créent très rarement un parallélisme entre les items et sont, en général, insuffisants seuls pour signaler tous les items non marqués de la série. Par exemple, l'introduction d'un item de la série par une expression cadrative temporelle telle que *dans les années 1920* ne suffit pas seule pour rendre toute la série facilement repérable ou parallèle si les autres items ne sont pas signalés de façon similaire.

7.3.5 INDICES COMPLEMENTAIRES ET TITRES

Les titres sont des indices complémentaires de sériation excellents ; bien qu'ils ne marquent pas l'addition ou l'ordre des items dans la série linéaire, ils les introduisent et délimitent et ainsi signalent au lecteur leur présence (cf. 7.2.4). Par conséquent, les items non marqués des séries section et méta sont par défaut signalés par des moyens linguistiques et plutôt facilement repérables. De plus, comparées aux séries typiques, les items de ces séries ont la particularité d'être très longs, ce qui influence le fonctionnement des autres indices complémentaires aussi.

Les fréquences des indices complémentaires dans les séries section et méta sont présentées ci-dessous dans la Figure 35. Il est à noter que le nombre d'occurrences dans une partie des catégories est relativement limité ; par exemple les séries sections non marquées n'ont qu'une seule occurrence (pour les nombres de séries section et méta dans les catégories différentes, cf. section 7.2.4). Pour cette raison, la variation des indices dans les catégories présentée n'est pas quantitativement fiable. Un deuxième aspect à noter est l'absence de certains types d'indices, due à la nature des séries section et méta. La correspondance des items aux exemples et l'équivalence formelle des items ne sont pas présentes dans ces séries du corpus. La présence des traits syntaxiques contribuant à l'identification des items, en revanche, s'explique par les séries métadiscursives faisant partie des structures enchaînées *méta*. Ces séries présentent les sections-items de la série à suivre en formant elle-même une série inférieure. Par conséquent, elles peuvent être relativement courtes et les aspects syntaxiques y peuvent contribuer à l'identification des items.

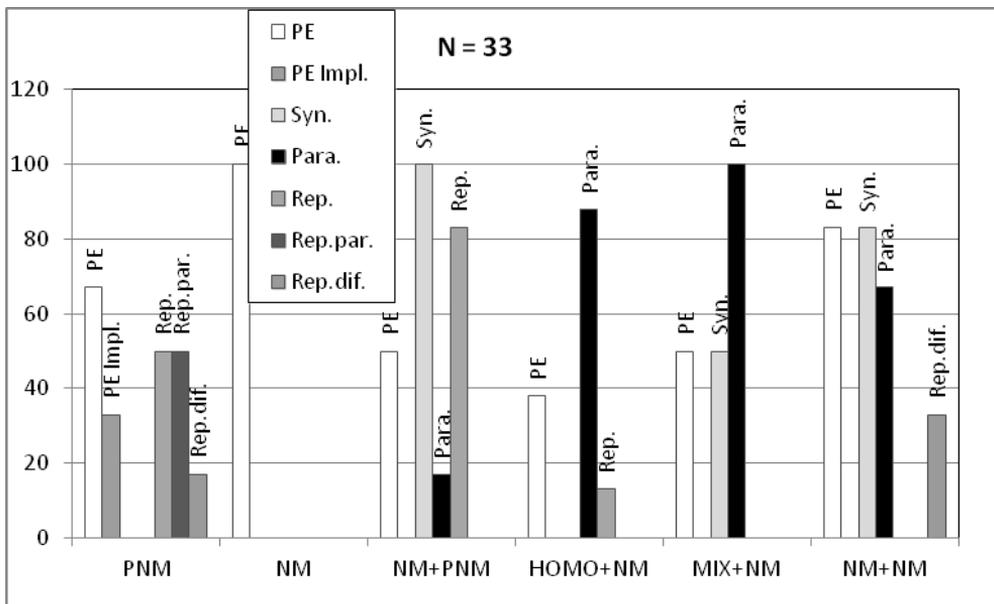


Figure 35 Indices complémentaires dans les séries section et méta.

Selon la Figure 35 tout d'abord, comparées aux séries typiques sans sections titrées, *la répétition proprement dite* est plutôt rare dans les séries méta ; elle est relativement utilisée uniquement dans les séries NM+PNM. *La paraphrase*, en revanche, est extrêmement utilisée. La fréquence de cette dernière cause également l'absence des types de répétition, c'est-à-dire de *la répétition partielle* et de *la répétition des éléments différents*, car ces deux traits ne sont pas analysés pour les séries avec paraphrase. La rareté de la répétition et la fréquence de la paraphrase dans les séries méta s'expliquent sans doute par des raisons stylistiques et par le fonctionnement des séries métadiscursives dans l'introduction des sections à venir. Pour ces structures, la paraphrase est sans doute plus convenable que la répétition exacte de mêmes syntagmes, qui risquerait souvent d'être moins informative et plus maladroite (cf. entre autres l'Exemple 60 à la page 196).

Contrairement aux séries méta, dans les séries section PNM, composées uniquement de sections titrées, la répétition lexicale est selon la Figure 35 présente. Dans ces séries, comme l'illustre l'Exemple 73 ci-dessous, la répétition ne risque pas d'être maladroite mais plutôt semble simplement relier les items et même créer un certain parallélisme entre eux. Ainsi, la répétition contribue au marquage de la série bien que les items y soient repérables sans celle-ci. Dans cet exemple, le syntagme répété est annoncé déjà dans l'amorce qui correspond à un titre supérieur indiquant que les sections suivantes discutent les multiples dimensions du travail (d'enseignants) : *un travail aux multiples dimensions*. Le syntagme est répété dans tous les titres des items-sections et au début de chaque section, quoique l'adjectif y lié varie dans chaque titre en fonction de la dimension discutée.

Exemple 73

0 Un travail aux **multiples dimensions**.

1 **Dimension cognitive**

Le travail des enseignants du CFER étudié comporte **une dimension cognitive**. Toutefois, la compréhension de la matière enseignée est somme toute secondaire pour ces enseignants, en raison de la nature même des apprentissages visés. [...]

2 **Dimension affective**

La dimension affective constitue, pour les enseignants du CFER rencontrés, une facette particulièrement importante de leur travail, davantage d'ailleurs que s'ils enseignaient au secteur régulier, ou même auprès d'élèves en adaptation scolaire qui fréquentent une école plus traditionnelle. [...]

3 **Dimension interactive**

La troisième dimension du travail enseignant que nous avons présentée dans notre cadre d'analyse est de l'ordre de l'interaction, tant avec les élèves qu'avec les collègues. [...]

Martineau

Enfin, outre les différents types de répétition, un indice complémentaire majeure dans les séries section et méta est la contiguïté sémantique des items annoncée par l'amorce, clôture ou une construction numéro+nom. Comme l'illustre la Figure 35, cet indice est présent dans une grande partie des séries dans toutes les catégories, entre autres dans l'Exemple 73 ci-dessus. Le *PE implicite*, en revanche, n'est utilisé que pour les séries simples PNM. Dans cette perspective, les séries section PNM ressemblent leurs variantes typiques qui ne sont pas composées de sections entières.

7.3.6 BILAN : INDICES COMPLEMENTAIRES DE SERIATION, FACTEURS DE NON MARQUAGE ?

Le but de cette section a été d'examiner la façon dont les items non marqués sont identifiés dans les séries linéaires de notre corpus. En particulier, nous avons privilégié la nature et la fréquence des indices complémentaires de sériation, soit des traits permettant au lecteur d'identifier l'item non signalé par un marqueur d'addition ou d'ordre. Comment sont-ils ? Pourraient-ils expliquer l'absence de marquage des items ?

Tout d'abord, notre analyse a relevé la présence de deux types d'indices complémentaires contribuant à l'identification des items non marqués : ceux qui suffisent seuls pour signaler au lecteur la présence de l'item et ceux qui peuvent seulement faciliter une identification effectuée principalement grâce à d'autres indices. Par exemple, la répétition d'un syntagme lexical à l'initiale de tous les items de la série semble suffire pour indiquer au lecteur la présence des items, tandis que leur appartenance à une même phrase n'y suffit pas. La répétition de mots individuels, qui est souvent examinée dans des travaux sur la cohésion lexicale (cf. entre autres Hoey 1991), nous a surtout paru insuffisante pour indiquer au lecteur la présence des items. De plus, notre analyse a souligné la coexistence de plusieurs

indices qui tous guident le lecteur au fil d'une seule série. La présence d'un type d'indice ne rend donc pas les autres inutiles. Entre autres, dans les séries section, la répétition lexicale au début de chaque item rend la série encore plus facilement repérable bien que les items soient déjà délimités par des titres.

En ce qui concerne la nature des indices relevés par notre analyse, la plupart d'eux sont cohésifs, ainsi la répétition lexicale ou l'annonce explicite du principe d'énumération. De fait, la cohésion et les indices linguistiques de surface semblent centraux dans l'identification des items non marqués. Les items que nous avons classifiés, lors de l'annotation, comme non marqués sont en fait souvent identifiables grâce aux indices linguistiques. Ce résultat soutient également la remarque de Taboada (2009) sur le fait que les relations de cohérence sont souvent signalées dans le texte même si aucun marqueur explicite n'est utilisé. De même, le résultat indique que le modèle de Goutsos (1996), tenant uniquement compte des indices explicites de l'organisation textuelle, est tout à fait applicable bien qu'il n'analyse pas les liens non marqués. Cela dit, nous devons admettre que l'interprétation du lecteur joue aussi un rôle lorsque les indices linguistiques sont présents. En effet, le lecteur doit toujours les interpréter correctement afin qu'ils lui facilitent la lecture. Notamment, l'annonce explicite du principe d'énumération par l'amorce, la clôture ou une construction Num+N requiert encore que le lecteur identifie les items correspondants à ce principe. Ce processus peut à nouveau reposer sur des connaissances extralinguistiques du lecteur.

Quant aux fréquences des indices complémentaires des différentes catégories de séries, certaines catégories semblent se distinguer des autres. En particulier, la présence de la répétition proprement dite, de la paraphrase et de l'annonce explicite du principe d'énumération varient. La répétition semble plutôt caractéristique des séries typiques qui ne sont pas composées de sections titrées, alors que la paraphrase est plus utilisée dans les séries section et méta. Cette différence s'explique sans doute surtout par les séries métadiscursives. De fait, dans les séries méta, elles présentent les thématiques des items-sections à venir ; une répétition risquerait d'être maladroite dans cette fonction. Dans les séries typiques, en revanche, la répétition est un signal fréquent ; dans d'autres travaux sur les énumérations dont l'ordre n'est pas un critère organisation important, elle est même considérée comme un indice primaire de l'énumération (cf. Luc 2001 ; Ho Dac *et al.* 2009). La fréquence de la répétition comme indice complémentaire varie pourtant également au sein des séries typiques. Elle est la plus fréquente dans les séries enchaînées composées d'items non marqués (NM+NM) et la plus rare dans les séries simples partiellement non marquées. Au sein des autres catégories, sa présence est du même ordre.

Outre les formes différentes de répétition, les séries typiques se distinguent par la présence de l'annonce explicite du principe d'énumération. Bien que sa fréquence soit relativement similaire dans la grande majorité des séries typiques non composées de sections titrées, elle s'avère plutôt rare dans les séries PNM. Les autres indices complémentaires primaires pouvant garantir l'identification des items non marqués s'y montrent également rares ;

comme nous venons de le constater, entre autres, la répétition est peu utilisée dans cette catégorie. Par conséquent, pour ces séries, l'identification des items semble souvent reposer sur la contiguïté sémantique des items bien qu'elle ne soit pas explicitement annoncée. Ainsi, le manque d'indices linguistiques de surface et l'importance de l'interprétation du lecteur sont plus essentiels dans ces séries PNM que dans les autres catégories analysées.

En définitive, les résultats indiqués dans cette section correspondent et confirment plusieurs de nos remarques et hypothèses relatives aux différences entre les catégories de série présentées dans les sections précédentes. Tout d'abord, les séries enchaînées se montrent relativement facilement repérables même non marquées ; les indices complémentaires y sont plutôt fréquents. Ce résultat confirme nos démonstrations des sections précédentes sur la nature explicite et exacte de ces structures. L'examen du non marquage des séries partiellement non marquées, en revanche, indique que même les indices complémentaires s'y font plutôt rares. Outre la rareté de leur combinaison à une amorce ou une clôture et l'inclusion de peu de marqueurs exacts, elles incluent, en outre, peu d'indices complémentaires. Tous nos résultats indiquent ainsi que les séries partiellement non marquées sont relativement implicitement signalées et que l'interprétation du lecteur y revêt un rôle plus important que dans des séries dont tous les items sont explicitement marqués.

Enfin, cette section devait aussi examiner les indices complémentaires comme facteurs d'explication du non marquage des items. La réponse à cette question est à la fois positive et négative. Sans doute, les indices complémentaires primaires peuvent-ils contribuer au non marquage, notamment dans les séries courtes dont le nombre d'items est signalé dans la proximité, soit les séries simples non marquées et les séries enchaînées. Selon notre analyse, ces indices sont surtout la répétition proprement dite, l'équivalence structurelle des items et la correspondance des items aux exemples numérotés. Pour les séries additives PNM également, leur organisation interne peut être à l'origine du non marquage.

Dans les autres séries, en revanche, les indices complémentaires nous semblent insuffisants pour expliquer le non marquage. En outre, selon notre analyse, ils se révèlent même plutôt rares dans bon nombre de séries. Ainsi, les marqueurs complémentaires sont loin d'expliquer l'intégralité du non marquage. De même que nos résultats des sections précédentes, notre analyse propose simplement que certaines séries ne recherchent pas à un guidage aussi explicite que d'autres.

7.4 BILAN : FACTEURS DU MARQUAGE ET DU NON MARQUAGE DES SÉRIES LINÉAIRES

Le but de cette section a été d'examiner les facteurs textuels qui pourraient expliquer la façon dont les séries linéaires sont signalées dans le texte. Surtout, notre étude avance donc que le marquage de l'organisation textuelle est le résultat de plusieurs facteurs. Plus

particulièrement, les résultats indiquent que le marquage et le non marquage de l'organisation textuelle ainsi que des séries linéaires varient selon le contexte et que toutes les séries ne tendent pas à guider pareillement le lecteur, qu'elles ne réalisent pas les mêmes stratégies textuelles.

Tout d'abord, nos résultats sur le degré d'exactitude du marquage énoncent que certaines structures recherchent simplement une organisation moins explicite et moins exacte que d'autres. Les séries avec des traits associés aux non parallélisme, notamment le non marquage et l'usage de plusieurs types de marqueurs, sont plutôt rarement combinées aux amorces et clôtures exactes. Les marqueurs d'item exacts s'y font aussi plus rares. Dans les séries explicitement introduites par le même type de marqueur, en revanche, la structure de la série est souvent répétée par l'amorce ou la clôture et les marqueurs d'items exacts sont fréquents.

La longueur est un deuxième facteur du marquage des séries linéaires : le marquage des séries longues est plus souvent moins exact et moins explicite que celui des séries courtes. Les types de marqueurs remplissent aussi une fonction d'introduction des items de longueurs différentes : anaphoriques pour les plus courts, connecteurs pour ceux de longueur intermédiaire et constructions numéro+nom pour les plus longs. Selon notre analyse, le non marquage complet des items d'une série, en revanche, est souvent dû à la longueur limitée de la structure et au marquage des items dans la proximité, soit dans l'amorce ou la clôture, soit dans l'autre partie de la série enchaînée. Dans celles-ci, les items semblent demeurer non marqués afin d'éviter la répétition excessive des marqueurs.

Enfin, les indices complémentaires de sériation sont le dernier facteur dont nous avons analysé l'effet sur le marquage. Ces indices peuvent contribuer à l'identification des items mais ne sont pas compris comme marqueurs explicites devant, selon notre schéma, signaler l'addition ou l'ordre. Surtout, notre analyse indique que plusieurs types d'indices guident le lecteur dans la lecture des items non marqués. Certains, tels que la répétition de mots ou de syntagmes nominaux dans tous les items, facilitent la lecture de façon importante. D'autres, tels que la répétition de mots isolés dans les items, se bornent à soutenir le travail du lecteur. Certains types d'indice peuvent également être à l'origine de l'absence des marqueurs explicites ; ensemble avec la faible longueur, la présentation des items, par exemple par une amorce, les indices complémentaires forts, comme la correspondance des items aux exemples numérotés voire leur similarité structurelle ou lexicale peuvent rendre leur marquage explicite inutile. Ce n'est pourtant toujours pas le cas. En effet, de même que l'usage des marqueurs exacts et relatifs, la présence des indices complémentaires varie selon la façon dont la série est autrement signalée. Dans les séries courtes dont le nombre d'items est indiqué dans la proximité, les indices forts sont plutôt fréquents et les séries facilement repérables. Dans les séries plus longues aux items non marqués et explicitement signalés, les indices complémentaires sont en revanche plutôt rares.

En définitive, les raisons sous-jacentes de nos résultats sont une question incontournable. Pourquoi le marquage est-il plus imprécis et moins explicite dans les passages longs que dans les segments plus courts ? Contrairement au non marquage des séries les plus courtes, qui s'explique par la maxime de Grice (1975), ce résultat est plus difficile à justifier. Nous pouvons toujours spéculer sur des aspects cognitifs derrière l'interprétation des marqueurs des passages longs : les marqueurs exacts pourraient-ils être difficiles à mémoriser dans les segments longs et, de ce fait, être délaissés ? Intuitivement, nous penchons cependant pour l'alternative contraire : le marquage explicite et exact serait nécessaire afin de rappeler au lecteur une organisation autrement difficile à mémoriser. La remarque de Jackiewicz (2005 : 98) sur le marquage plus explicite des séries plus longues tendrait d'ailleurs à conforter cette hypothèse. Par conséquent, nos résultats étant plutôt en opposition avec cette supposition, notre travail montre au moins la nécessité d'études quantitatives sur l'usage de la langue. Pour clarifier les raisons sous-jacentes de ce phénomène, des études plus approfondies semblent cependant s'imposer – le nôtre ne saurait que confirmer son existence.

8. SERIES LINEAIRES ET VARIATION INTERDISCIPLINAIRE

Dans les sections précédentes, nous avons examiné les séries linéaires, leur marquage et les facteurs qui les influencent. Cette section envisage ces structures selon une autre perspective en comparant leur usage dans les différents domaines scientifiques. Nous avons déjà, en section 2.2.3, décrit les trois disciplines dont les articles de notre corpus sont issus : la linguistique, l'éducation et l'histoire. Cette description fonde également leur comparaison car leurs caractéristiques peuvent expliquer leurs différences et similarités en regard de l'usage des séries linéaires.

La linguistique et l'éducation se ressemblent à plusieurs titres : toutes deux se trouvent à l'intersection des sciences humaines et sociales, dont les résultats peuvent reposer sur des faits mesurables ainsi que sur le raisonnement du chercheur. Comme la nature des résultats influence, selon un nombre d'études (Hyland 2005b : 187-188, cf. aussi Hyland 2000 : 33 ; Dahl 2005 : 295 ; Fløttum *et al.* 2006 : 20-21), plusieurs caractéristiques des articles, cette similarité peut proposer que la linguistique et l'éducation se ressemblent. La similarité peut, par exemple, concerner la nature argumentative des articles et la présence d'indices de certitude et d'incertitude de l'auteur. De même, les conventions d'écriture et « les règles de la communauté scientifique » sont plutôt libres dans les deux domaines. Cela peut transparaître dans la variation des articles ainsi que dans l'emploi du métadiscours ; la structure des articles n'étant pas figée dans ces domaines, le lecteur doit être guidé plus explicitement à travers le texte.

Science humaine plus typique, l'histoire diffère de la linguistique et de l'éducation ; ses résultats sont rarement mesurables et le rôle de l'argumentation y est très important (voir par exemple Hunston 1993 ; Martin 2003 ; Bondi 2009a). Les conventions d'écriture libres sont également typiques de ce domaine. Or, l'histoire comporte également des particularités la distinguant des autres sciences. La façon d'argumenter de ce domaine, notamment, est différente car elle ne recherche pas la généralisation des résultats mais simplement l'explication de « ce qui s'est réellement passé » (cf. section 2.2.3.5, Carr 1961). Cet objectif transparaît pareillement dans la nature des articles que Bondi (2009a : 98, aussi 2009b), par exemple, décrit comme narratifs. Cette narrativité est particulièrement remarquée dans les ouvertures des articles ; Bondi (2009a) et Bondi et Silver (2004) notent que les sources des voix y sont dramatisés à l'aide du discours rapporté. Nous estimons que le marquage de l'organisation textuelle et l'usage des séries linéaires, plus rares dans cette discipline selon notre hypothèse, peuvent aussi être influencés.

À la suite, nous chercherons à examiner la variation de l'usage des séries linéaires dans les trois disciplines étudiées. Nous commencerons par les fréquences de toutes les séries ainsi que par celles des types de série particulières, soit entre autres les séries temporelles

s'organisant selon un axe chronologique. Dans un deuxième temps, nous comparerons les longueurs des séries et les présences des séries s'étendant sur plusieurs sections titrées, puis nous continuons par la variation du marquage des séries dans les disciplines étudiées. Enfin, la dernière partie de cette section sera consacrée à un autre type de comparaison ; elle compare les articles de notre corpus parus dans des revues canadiennes et ceux parus en France.

8.1 FREQUENCES

Le premier aspect évident de comparaison entre les disciplines est la fréquence des séries linéaires, présentée dans la Figure 36 ci-dessous. Comme nous le notons, l'histoire se distingue clairement de la linguistique et de l'éducation avec sa fréquence de série plus basse. La linguistique et l'éducation, en revanche, sont très similaires. Ces résultats confirment notre hypothèse introductive ; les séries linéaires sont plus fréquentes en linguistique et éducation qu'en histoire. La linguistique et l'éducation, en revanche, sont très similaires.

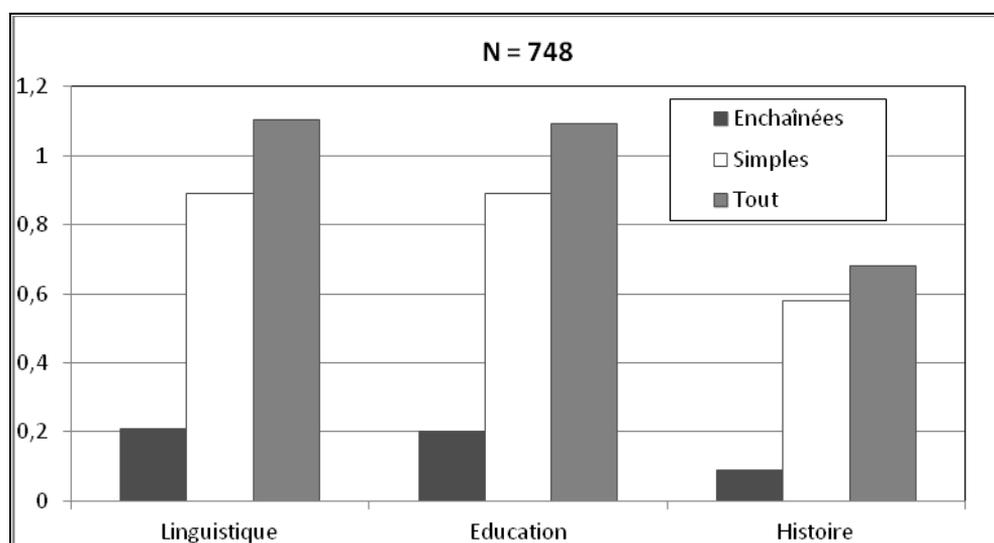


Figure 36 Fréquences des séries dans les sous-corpus, en moyenne du nombre de séries / 1 000 mots.

Or, bien que les séries soient moins fréquentes en histoire, elles revêtent tout de même un rôle important dans ces articles. En moyenne, un article d'histoire inclut toujours sept séries linéaires. Par ailleurs, la proportion des séries enchaînées est similaire en histoire et dans les deux autres disciplines étudiées⁸⁵. Ainsi, bien que le style narratif des articles d'histoire soit souvent souligné, le marquage de l'organisation textuelle dans ce domaine

⁸⁵ La différence de proportion des séries enchaînées en histoire ainsi qu'en linguistique et en éducation ne semble pas statistiquement significative avec le test X2, la valeur $p = 0,1423$, $X^2 = 2,15$, $N = 748$.

ne doit pas être sous-estimé. L'Exemple 74 ci-dessous illustre l'usage des séries linéaires en histoire. Décrivant le ravitaillement de Paris en 1793, il se compose de deux items et d'une amorce répétant le nombre d'items.

Exemple 74

- 0 En revanche, au cours de l'automne 1793, se produit progressivement une quasi disparition des missions spécifiques destinées au ravitaillement de Paris, ce **pour deux raisons**.
 - 1 **D'une part**, un décret du 1er brumaire an II (22 octobre 1793) crée une Commission des subsistances, chargée initialement de veiller à l'exécution du Maximum des prix des grains voté le 11 septembre 1793, renforcé par le Maximum général des prix et salaires le 29 septembre. Celle-ci élargit très vite ses attributions et veille non seulement à l'approvisionnement de Paris, mais aussi à celui des armées, à celui des départements déficitaires, aux greniers d'abondance, aux importations de subsistances, etc.
 - 2 **D'autre part**, les représentants du peuple envoyés dans les départements pour des missions à «généralistes» se substituent alors souvent à leurs prédécesseurs pour veiller aux approvisionnements, force étant en effet de constater que la plupart d'entre eux passent une partie importante de leur temps de mission à s'occuper des problèmes du ravitaillement, y compris, le cas échéant, le ravitaillement de Paris.
- Cl. Il faut **donc** attendre les lendemains de Thermidor pour voir réapparaître les missions spécifiquement réservées à cet approvisionnement de la capitale, mais avec cette fois des missionnaires dont les orientations politiques sont bien différentes.

Biard

Dans l'Exemple 74, les items correspondent au principe d'énumération annoncé par l'amorce, c'est-à-dire *la raison*. Ils sont également explicitement marqués par des connecteurs. Il s'agit donc d'une série assez typique.

Outre les simples fréquences de toutes les séries, les proportions des séries avec une étiquette particulière reflètent aussi l'usage de ces structures dans les différents domaines. La Figure 37 ci-dessous présente les fréquences de ces structures, présentées en détail à la section 6.1. Les séries *temporelles* se composent d'items correspondant aux événements du monde extérieur décrits par le texte. Dans les *séries argumentatives*, les items correspondent à une thèse originale, à l'antithèse et à la conclusion, ou nouvelle thèse, selon le modèle de Toulmin (1958). Les séries *accumulatives* incluent un nombre important d'items, certains correspondant au principe d'énumération de la structure et d'autres présentant par exemple contre arguments ou concessions. Les séries *à deux temps*, enfin, comprennent deux séries inférieures dont la première présente les items auxquels ceux de la seconde série se réfèrent par des marqueurs thématiques, tels que *en ce qui concerne* et *quant à*. Les étiquettes *section* et *méta*, absentes de la figure, sont discutées lors de la comparaison des longueurs des séries en section suivante.

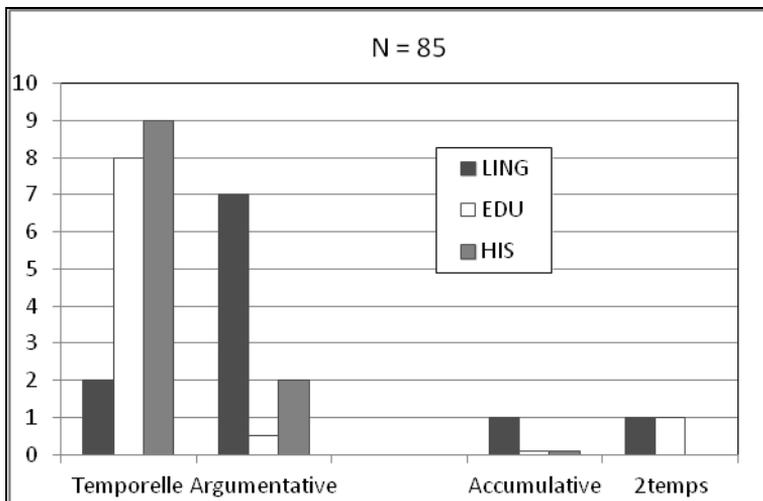


Figure 37 Fréquences des séries avec étiquettes particulières (en pourcentage du total).

La Figure 37 montre que ce sont surtout les proportions des séries temporelles et argumentatives qui varient entre les domaines. Les séries temporelles sont moins fréquentes en linguistique qu'en éducation et en histoire⁸⁶. La fréquence de ces séries en histoire est conforme à nos attentes car les articles de ce domaine s'efforcent de décrire et à interpréter des événements du passé. L'importance des notions temporelles dans ces domaines est également notée par Bondi (2009a : 87, voir aussi Martin 2003 : 28). Leur fréquence similaire en histoire et en éducation est en revanche intéressante⁸⁷. Ainsi, quoique les séries temporelles soient typiques des articles en histoire, elles ne sont pas uniques en ce domaine. De plus, il faut noter que les séries temporelles ne forment toujours que 10 % environ des séries en histoire et en éducation alors que les séries typiques sont toujours majoritaires dans les deux domaines. La présence des séries temporelles n'affaiblit donc pas l'importance des séries typiques.

En histoire, le rôle des séries temporelles est plutôt évident : elles décrivent souvent des événements du passé qui forment le centre de l'étude. En éducation, en revanche, leur fonctionnement est moins intuitif. Selon notre examen manuel, elles peuvent caractériser des périodes temporelles du passé mais décrivent plus souvent des étapes de l'analyse effectuées lors de l'étude rapportée. L'Exemple 75 ci-dessous illustre ce fonctionnement.

⁸⁶ Avec le test X2, la différence de fréquence des séries temporelles est statistiquement significative. Entre la linguistique et l'éducation, la valeur $p = 0,0022$, $X^2 = 9,34$, $N = 539$; entre la linguistique et l'histoire, la valeur $p = 0,0006$, $X^2 = 11,79$, $N = 509$.

⁸⁷ Avec le test X2, la proportion des séries temporelles dans les articles en éducation et en histoire ne diffère pas. La valeur $p = 0,793$, $X^2 = 0,07$, $N = 448$.

Exemple 75

- 0 Ces observations se déroulent à l'UFR Sciences du Langage, de l'Homme et de la Société de Besançon **en deux temps** :
- 1 **d'abord** sur l'ensemble du 2^o semestre de l'année universitaire 2005-2006 (c'est-à-dire de mi-janvier à fin avril. Elles s'échelonnent sur 12 séances d'une heure chacune, à raison d'une séance par semaine ;
 - 2 **ensuite** sur le 2^o semestre de l'année universitaire 2006-2007 avec un autre groupe d'étudiants, de la même file et dans les mêmes conditions de temps et de durée.
- Cl. Les données recueillies permettront ainsi une comparaison entre **les deux groupes**.

Parmentelat

Dans l'Exemple 75, la série décrit l'analyse du travail en distinguant deux étapes auxquelles les items se réfèrent dans l'ordre chronologique. La fréquence de ce type de séries en éducation peut refléter certaines caractéristiques de ce domaine. Elle peut, par exemple, indiquer que la présentation détaillée de la méthode du travail et de ses étapes est importante dans ce domaine ou que les méthodes reposent souvent sur des étapes se succédant selon un ordre chronologique. Toutefois, la présentation de la méthode de cette façon semble typique uniquement pour l'éducation.

Outre les séries temporelles, les séries étiquetées argumentatives aussi reflètent l'organisation et surtout l'argumentation typiques des disciplines différentes. Comme l'indique la Figure 37, les séries argumentatives sont clairement les plus fréquentes dans les articles en linguistique, présentent quelques occurrences en histoire et sont pratiquement absentes en éducation⁸⁸. Cette distribution semblerait indiquer que l'organisation typique des séries argumentatives est plus courante en linguistique. Le plus fréquemment, ces séries semblent organiser une réflexion sur les aspects positifs et négatifs, par exemple d'un choix méthodique ou théorique du travail présenté. De plus, les items de ces séries peuvent correspondre aux conclusions ou interprétations alternatives de l'auteur. L'Exemple 76 ci-dessous se compose des aspects positifs et négatifs d'un programme de concordance. Le premier item présente un atout de ce programme et les deux autres des constats opposés à celui-ci, soit des antithèses. Enfin, la clôture conclut que, pour ces raisons, le programme ne serait pas très utile pour notre travail.

Exemple 76

- 1 Le concordancier, employé a minima comme simple outil de recherche dans le texte en version électronique [...], est un précieux outil de repérage de certains marqueurs.
- 2 **Mais** il ne pallie pas le problème de l'équivocité ou de la sous-détermination que nous avons à rencontrer (§ 4.2).
- 3 **Et** il n'est de toutes façons que de peu d'utilité dans certains cas, comme celui de la causalité : sans développer plus avant, signalons simplement que celle-ci, omniprésente

⁸⁸ La différence de fréquence des séries argumentatives est statistiquement significative avec le test X².

Entre la linguistique et l'éducation, $p = 0,0003$, $X^2 = 13,08$, $N = 539$;
entre la linguistique et l'histoire, $p = 0,0342$, $X^2 = 4,49$, $N = 509$.

dans les textes, y prend dans les faits des tours très divers : les exemples [1] et [3], notamment, donnent lieu à une interprétation en termes de causalité.

Cl **En conséquence**, celle-ci ne relève pas d'un simple « repérage », et suppose une interprétation appuyée sur un empan large (palier du paragraphe au moins, ou encore rapport entre titre et texte d'un article).

Lecolle

En définitive, la rareté des séries à deux temps et des séries accumulatives empêche leur analyse statistique. Ainsi, ce sont surtout les fréquences des séries temporelles et argumentatives qui signalent des spécificités de l'organisation textuelle des domaines examinés. La fréquence des séries temporelles, surtout, est très informative ; elle indique que ce type d'organisation est typique de l'histoire et de l'éducation. Simultanément, le fait que la majorité des séries est tout de même typique et ne repose pas sur un axe temporel démontre que les séries linéaires et le marquage de l'organisation textuelle sont caractéristiques de toutes les trois disciplines examinées. Néanmoins, la rareté extrême des séries à deux temps et des accumulatives en histoire peut également être un indice d'une caractéristique de ce domaine bien qu'elle ne puisse pas être prouvée par une analyse quantitative. Nous pouvons toutefois nous demander si leur quasi absence signale un usage moins varié des séries en histoire tandis qu'en linguistique et en éducation la variété des structures différentes serait plus importante. Pareillement, la fréquence des séries argumentatives, surtout en linguistique, soutiendrait cette observation.

8.2 LONGUEURS

Outre la simple fréquence des séries, la longueur même de ces structures est un aspect susceptible de refléter la façon dont le métadiscours et le marquage de l'organisation textuelle sont présents dans les textes des différents domaines. Dans les sections précédentes, nous avons déjà décrit les longueurs des séries du corpus et examiné la variation de leur marquage. La Figure 38 ci-dessous présente les longueurs des séries selon les domaines d'étude. Il est à noter que les séries *section* dont les items correspondent aux sections titrées de l'article et les séries *méta* composées d'une série section et d'une série métadiscursive présentant les items-section à venir sont exclues de cette figure. Elles seront discutées par la suite.

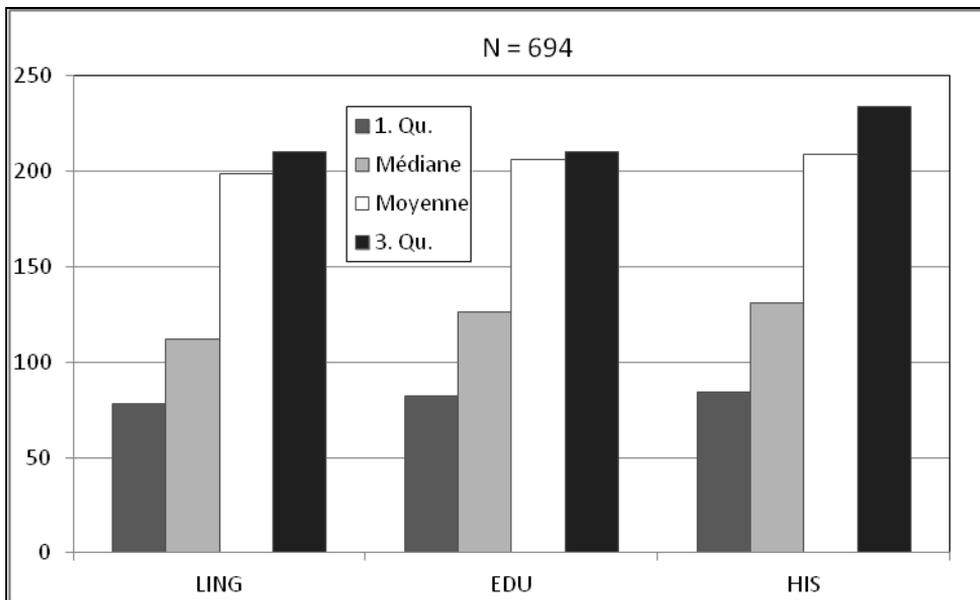


Figure 38 Longueurs des séries des trois disciplines, séries section et méta exclues⁸⁹.

Selon la Figure 38, les longueurs des séries paraissent relativement similaires dans les différents domaines. Bien que la médiane en histoire (131) semble plus élevée que celle en éducation (126,5) et en linguistique (112), la différence ne peut pas être statistiquement prouvée⁹⁰. La longueur des séries typiques ne semble donc pas varier dans les textes différents.

Or, les fréquences des séries section et méta semblent indiquer que la longueur est tout de même une qualité distinctive selon les disciplines étudiées ; comme l'indique la Figure 39 ci-dessous, les séries longues composées de sections entières ne sont pas utilisées de façon similaire dans les différentes disciplines. Comme pour la fréquence des séries typiques, c'est surtout l'histoire qui se distingue des deux autres domaines.

⁸⁹ Le 1^{er} quartile sépare les 25 % inférieurs des données et le 3^e quartile les 75 % inférieurs.

⁹⁰ Entre l'histoire et la linguistique, la valeur $p = 0,2025$ avec Kruskal Wallis ($H(X^2) = 1,62$, $N = 472$; entre la combinaison de la linguistique et de l'éducation d'une part et de l'histoire de l'autre, la valeur $p = 0,3179$, $H(X^2) = 0,99$, $N = 694$.

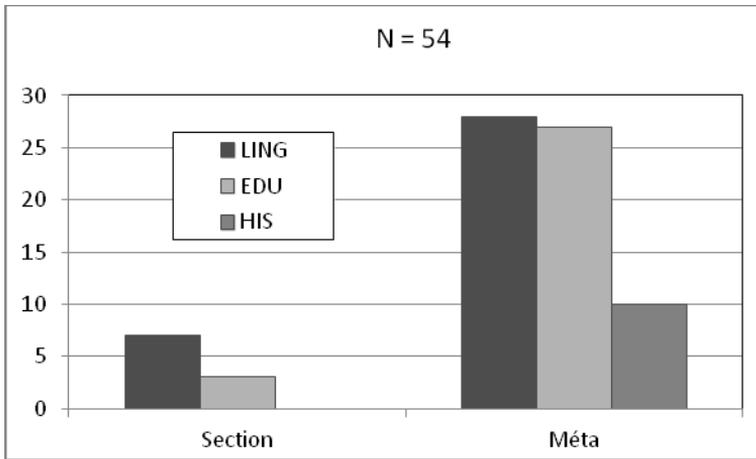


Figure 39 Séries section et méta dans les trois domaines. Les fréquences sont en pourcentages du nombre total des séries simples / enchaînées dans le sous-corpus.

Selon la Figure 39, la grande majorité (18/23) des séries section est utilisée dans les articles en linguistique⁹¹. En histoire, elles ne présentent aucune occurrence. Le marquage explicite de la façon dont les sections sont organisées diffère donc surtout entre la linguistique et l'histoire.

L'absence des séries section en histoire peut être due à deux facteurs ; soit leurs items ne sont pas structurés autour d'un principe d'énumération, soit cette organisation n'est pas suffisamment explicitement signalée pour former une série linéaire. La différence entre les fréquences de ces séries selon les domaines est tout de même claire. L'Exemple 77, tiré du sous-corpus linguistique, présente une série section dont les sections, soit les items, sont bien organisés autour d'un principe d'énumération et pour laquelle l'ordre entre les items est noté. Les sections discutent des *prédications* différentes, chacune traitant un type de prédication spécifique. Chaque item commence également par un indice signalant sa place exacte dans la série.

Exemple 77

1 2.1. Les prédications connectives.

Je commence par l'étude de (1) : (1) *cet homme vertueux me sera cher* qui constituera un exemple de prédication «connective». Plus précisément, je m'intéresserai à l'emploi de [...]

2 2.2. Les prédications centrées sur le groupe sujet.

Deuxième type de construction : les prédications centrées sur le groupe sujet. Il s'agit cette fois d'exemples comme [...]

3 2.3. Les prédications centrées sur le groupe verbal.

⁹¹ La fréquence des séries section en linguistique et en éducation est statistiquement significative avec le test X2, la valeur p = 0,04385, X2 = 4,06, N = 539.

Troisième et dernier type de prédication, celui repéré par le Moyen Age qui y voyait, comme Frege le fera, la construction prédicative de base. Il s'agit de cas où le sens du sujet [...]

Cl. 2.4. Bilan.

En résumé, tous les énoncés expriment selon moi des principes, en «donc» ou en «pourtant», et ce qui distingue les constructions c'est la manière dont ce principe est fourni [...]

Carel

Les fréquences des séries méta présentées dans la Figure 39 proposent aussi que la façon dont les sections sont organisées, ou du moins dont cette organisation est marquée, varie selon les domaines. Comme l'indique la figure, les séries méta, composées d'une série section et d'une série métadiscursive présentant les items-section à venir, sont relativement utilisées en linguistique et en éducation mais plutôt rares en histoire⁹². Ce résultat conforte notre hypothèse de rareté du métadiscours en histoire. L'Exemple 78 ci-dessous illustre la façon explicite dont les séries méta guident le lecteur et discutent l'organisation de l'article. Dans cette série, les deux items de la première série, signalés par des MIL, annoncent les thématiques des sections à venir. Ces thématiques sont également répétées dans les titres des sections respectives.

Exemple 78

- 0 Dans la suite de cet article, nous retiendrons **deux aspects** qui semblent importants pour la suite de la scolarité de ces élèves.
- 1 Nous examinerons **tout d'abord** l'aménagement des classes
- 2 **puis** nous détaillerons deux séances de langage observées dans les deux écoles les plus éloignées sur le plan du recrutement social.

1 L'aménagement de la classe

Comme dans la plupart des petites sections d'école maternelle, ces cinq classes comportaient plusieurs espaces matérialisés destinés à des activités différentes. Outre l'endroit réservé au regroupement des élèves, on y trouvait des «coins» réservés à la peinture, [...]

2 L'enseignement dispensé : exemple de séances de «langage»

Nous avons eu la chance d'assister à la même séquence dans les deux classes qui constituent les bornes de notre échantillon sur le plan du recrutement social ([...]). Il s'agissait de la lecture d'un album, activité classique en maternelle. [...]

Maso-Taulère

Dans l'Exemple 78 précédent, la série méta annonce au lecteur la structuration de l'article de façon très explicite et englobante ; elle présente les thématiques des sections à venir et indique leur ordre de discussion. La nature explicite du guidage transparaît également dans les choix lexicaux ; la série décrit l'action de l'auteur (*nous examinerons – nous*

⁹² La fréquence des séries méta entre l'histoire d'une part et la linguistique et l'éducation de l'autre est statistiquement significative avec le test X², p = 0,0348, X² = 4,45, N = 748.

détaillerons). Les titres des sections répètent encore les thématiques avec les mêmes choix lexicaux. Grâce à ces caractéristiques, de l'ensemble des séries linéaires, les séries méta correspondent peut-être le mieux au métadiscours dans sa définition traditionnelle (cf. section 2.2.2) ; la série a pour but principal la présentation de la structure du texte à venir. L'information fournie par les autres types de série peut être plus implicite. De fait, l'auteur n'y annonce pas nécessairement aussi directement la façon dont il organisera son texte mais guide plutôt le lecteur parmi des segments plus courts. Comme exposé dans les sections précédentes, le guidage fourni par les autres types de séries linéaires peut également être imprécis.

Comme exemple de la nature différente des séries typiques face aux séries méta, prenons le marquage des items relativement courts par *d'une part – d'autre part* ou par *le premier – le second*. Par comparaison avec l'Exemple 78 ci-dessus, il semble fonctionner à un tout autre niveau de guidage. Alors que le marquage des séries linéaires naît plutôt des indices de cohésion, sans doute présents dans pratiquement tous les textes cohérents, le métadiscours est plus restreint, couvre uniquement une partie mineure de l'usage des indices de cohésion et est utilisé seulement dans des textes de certains genres discursifs. Cette différence peut également expliquer la rareté des séries méta et la fréquence relative des autres types de série dans les articles d'histoire. Il se peut que le guidage du lecteur fourni par les autres types de série soit plus implicite et donc plus usité dans tous les textes. Le guidage pourvu par les séries méta serait, en revanche, entendu comme plus explicite, direct et typique uniquement de certains textes.

L'Exemple 78 ci-dessus a illustré la façon explicite et holistique dont les méta guident le lecteur. L'Exemple 79 ci-dessous, en revanche, présente la façon plus implicite dont les autres types de série structurent le texte et participent ainsi à ce même guidage. Composée d'une série NMet et d'une autre PNM, elle inclut des items explicitement signalés et des items non marqués. De plus, elle est précédée d'une amorce exacte et suivie d'une clôture.

Exemple 79

- 0 Ainsi, pour prendre **deux exemples** parmi les plus précis,
1 il existe de fortes différences entre la révolte des métayers du Bas-Adour en 1920
2 **et** les contestations multiples des métayers gemmeurs dans l'entre-deux-guerres.
1 **Les premiers**, exerçant leurs activités dans le sud du département, en Chalosse, protestèrent contre le régime draconien des contrats qui leur étaient imposés.
2 Les gemmeurs, **quant à eux**, tout en contestant également les conditions contractuelles, ne réclamaient pas l'accès à la propriété de la terre, mais demandaient à l'inverse l'obtention du statut de salarié.
Cl. **Cette différenciation** s'explique avant tout par leur relation distincte à l'économie de marché, les gemmeurs ayant dû faire face au passage douloureux d'une exploitation agro-pastorale de la terre à une exploitation agro-sylvicole, la sylviculture étant largement dominante.

Lafargue

La lecture de l'Exemple 79 permet de noter que le guidage fourni par cette série est différent et plus indirect que celui de la série méta dans l'exemple précédent. Nous pouvons peut-être conclure que cette série contribue plus à rendre la représentation mentale de l'auteur (*discours*) à un format textuel physique, lisible pour le lecteur (*texte*)⁹³ que les méta qui approchent le lecteur plus directement.

8.3 MARQUAGES

Dans les deux sections précédentes, nous venons de montrer que la fréquence des séries linéaires varie dans les trois disciplines étudiées bien que des similarités importantes existent aussi. Notamment, les articles en linguistique et en éducation se ressemblent selon plusieurs critères ; les fréquences de toutes les séries sont similaires et les séries méta sont aussi utilisées de manière semblable. L'éducation se caractérise pourtant par la présence de séries temporelles s'organisant sur un axe chronologique tandis que les séries argumentatives, structurant le texte selon une thèse originale, une antithèse et une conclusion, sont typiques de la linguistique. Les séries section composées d'items correspondant aux sections titrées présentent également une fréquence supérieure en linguistique qu'en éducation. Les articles en histoire, en revanche, diffèrent plus clairement des deux autres disciplines. Ils se caractérisent par une fréquence plus faible de tout type de série. De plus, comme en éducation, les séries temporelles s'organisant sur un axe chronologique sont souvent employées bien que les structures soient tout de même majoritairement des séries typiques. Le manque des séries section et méta, s'organisant autour des sections titrées, est également une caractéristique des séries en histoire. Il semble donc que le guidage explicite du lecteur assuré par les méta ainsi que l'organisation des sections autour d'un principe d'énumération sont plus rares dans ce domaine.

Les résultats jusqu'ici présentés sont tous liés à la fréquence des séries linéaires ; soit à la fréquence de tout type de structure, soit à la fréquence des variantes particulières. Or, outre celle-ci, la qualité des séries de notre corpus varie aussi. Cette section traite donc uniquement cet aspect et la variation de la qualité des séries entre les domaines de la linguistique, de l'éducation et de l'histoire.

Dans notre recherche, nous avons déjà examiné au cours des sections précédentes la qualité du marquage des séries linéaires, surtout à partir des catégories de série et des amorces et clôtures possiblement liées à la série. À la suite, nous comparerons leur variation dans les domaines étudiés en commençant par les catégories de série et en continuant par les amorces et clôtures.

⁹³ Dans la partie théorique, nous avons, suivant Cornish (1999 : 32-35, aussi 1990 et 2006) et Péry-Woodley (2000 : 13-14), discuté de la conception du *discours* comme représentation mentale de l'auteur et du *texte* comme trace physique de celui-ci.

8.3.1 CATEGORIES DE SERIES DANS LES DISCIPLINES ETUDIEES

Comme nous l'avons montré, les catégories de séries sont établies selon la présence des différents types de marqueurs et des items non marqués. Parmi les séries simples, les séries homogènes (HOMO) se composent d'items signalés par le même type de marqueurs alors que, dans les séries mixtes (MIX), les types de marqueurs varient et celles partiellement non marquées (PNM) incluent des items non marqués ainsi que d'autres explicitement signalés. Dans les séries non marquées (NM), enfin, tous les items sont non marqués et le nombre d'items annoncé par une amorce. Les NMet sont autrement similaires aux NM-ci mais incluent également un item signalé par une conjonction de coordination. Les séries enchaînées sont classifiées à partir de ces mêmes critères, selon le marquage de leurs séries inférieures. La distribution des catégories de séries entre les trois domaines d'étude est présentée ci-dessous dans la Figure 40. Afin de simplifier l'analyse, les séries NM et NMet sont réunies et, de même que dans les sections précédentes, seules les catégories de séries enchaînées les plus fréquentes sont présentées.

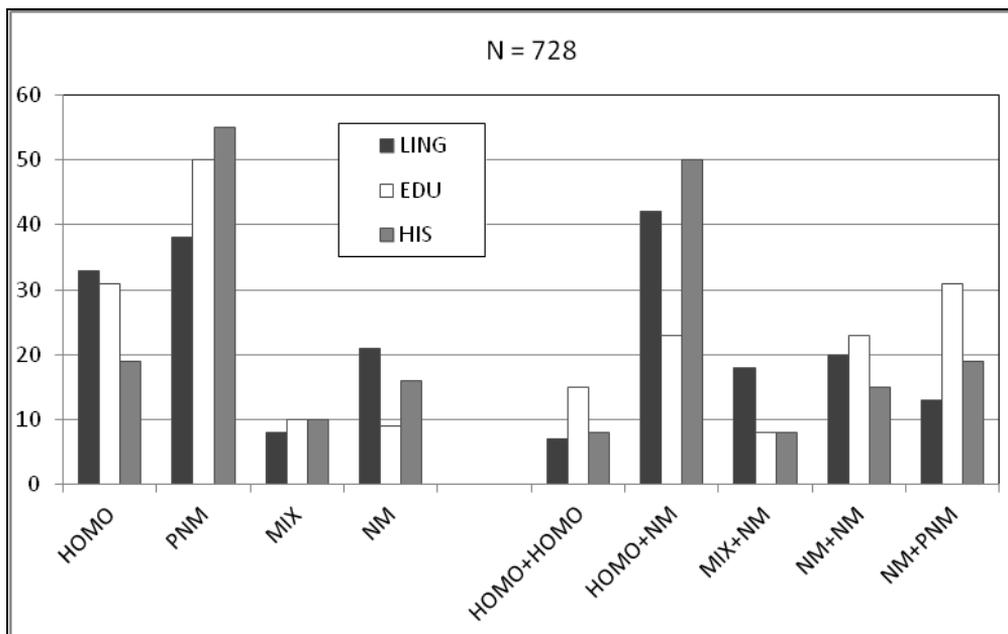


Figure 40 Catégories de séries selon les trois disciplines examinées. Les fréquences sont en pourcentages du total.

A partir de la Figure 40, il semblerait que la présence de presque toutes les catégories de séries varie entre les domaines. Dans la pratique, pourtant, le nombre d'occurrences des séries enchaînées est si bas que les différences sont, dans la plupart des cas, difficiles à prouver. Pour les séries simples, en revanche, les différences sont significatives ; c'est surtout l'histoire qui se distingue des deux autres domaines, les séries homogènes s'y faisant plus rares⁹⁴ et les séries partiellement non marquées plus fréquentes⁹⁵. Les séries

⁹⁴ La différence de fréquence des séries HOMO est statistiquement significative avec le test X2.

partiellement non marquées sont également plus fréquentes en éducation qu'en linguistique⁹⁶. Entre les trois domaines, le marquage des séries paraît ainsi le plus explicite en linguistique et le plus implicite en histoire. L'éducation se situe entre ces deux disciplines.

Ci-dessous, l'Exemple 80 illustre le marquage explicite des séries linéaires à partir d'une série homogène tirée d'un article en histoire ; il s'agit donc d'un mode de marquage plus rare dans ce domaine. L'exemple se compose d'une amorce vague indiquant que la série se compose de *quelques ouvrages* et de deux items explicitement signalés. *Avant tout*, le premier marqueur, surtout n'est pourtant pas des plus typiques bien qu'annoté comme connecteur. Par conséquent, le marquage de la série n'est peut-être pas très prototypique bien qu'il s'agisse d'une structure annotée comme homogène.

Exemple 80

- 0 Je m'appuie, pour ce faire, sur **quelques ouvrages récents** ;
- 1 **avant tout**, les études pionnières d'Yves Bénot, et singulièrement son étude fouillée des positions des rédacteurs de *La Décade* sur le problème de l'esclavage entre 1795 et 1802 dans *La démenche coloniale sous Napoléon*. [...]
- 2 Il faut **encore** citer l'ouvrage de Claude Wanquet, *La France et la première abolition de l'esclavage, 1794 - 1802*. Cet ouvrage étudie en détail l'évolution des colonies orientales ([...]), mais il accorde également une large place aux hésitations et contradictions de la politique officielle de la Première République. [...]

Gainot

L'Exemple 80 illustre le marquage plutôt exact des séries linéaires et de l'organisation textuelle, qui donc est plus typique en linguistique et en éducation qu'en histoire. L'Exemple 81 ci-dessous, en revanche, présente une série PNM qui, selon nos résultats, représente une façon typique d'organisation textuelle en histoire.

Exemple 81

- 1 **Au premier plan des destinations de ces exportations**, et loin devant l'Amérique Latine, viennent les pays industrialisés ou en voie d'industrialisation. Les pays européens, des États Sardes aux villes hanséatiques, et notamment les pays frontaliers, absorbent en effet tout au long du siècle plus de 60 % du volume de vêtements en vieux exportés par les marchands parisiens.
- 2 Les exportations vers le reste de l'Europe et la Russie, vers l'Amérique Latine et la façade atlantique des États-Unis s'élèvent, elles, à 10 %.
- 3 L'Afrique du Nord et le Sénégal apparaissent au début de la Monarchie de Juillet. Infimes en 1839, les quantités expédiées augmentent constamment grâce à l'ouverture

Entre la linguistique et l'histoire, la valeur $p = 0,0025$, $X^2 = 9,12$, $N = 423$;
entre l'éducation et l'histoire, la valeur $p = 0,0083$, $X^2 = 6,96$, $N = 375$.

⁹⁵ La différence de fréquence des séries PNM est statistiquement significative avec le test X^2 . Entre l'histoire et la combinaison de l'éducation et de la linguistique, la valeur $p = 0,0095$, $X^2 = 6,73$, $N = 618$.

⁹⁶ La différence de fréquence des séries PNM est statistiquement significative. Entre la linguistique et l'éducation, avec le test X^2 la valeur $p = 0,0165$, $X^2 = 5,75$, $N = 438$.

de nouveaux marchés, notamment en Algérie qui en 1860 en capte à elle seule près de 5 %.

- 4 **Enfin**, on trouve en dernière position et dès les années 1830 les Indes et 30 ans plus tard la Chine et le Japon.

Charpy

L'Exemple 81 énumère des destinations d'exportations importantes au XIX^{ème} siècle, chaque item correspondant à une destination ou un groupe de destinations. De ces quatre items, seuls le premier et le dernier item sont explicitement signalés par des marqueurs exacts, soulignant leur place dans la liste : *au premier plan* et *enfin, en dernière position*. Les autres items peuvent être identifiés surtout grâce à la répétition des noms de destination.

Les séries homogènes et partiellement non marquées sont celles distinguant le plus clairement les domaines examinés. Or, outre celles-ci, les séries NM semblent aussi varier ; elles sont plus fréquentes en linguistique qu'en histoire et en éducation⁹⁷. Dans les sections précédentes, nous avons décrit les séries NM comme courtes et souvent combinées aux amorces et clôtures. Grâce à ces caractéristiques, le marquage exact des items deviendrait facilement répétitif dans ces séries. Or, elles s'avèrent en général relativement faciles à repérer malgré le non marquage. L'Exemple 82 ci-dessous illustre ces séries. Cet exemple est composé d'une amorce exacte annonçant les items et leur nombre, de deux items et d'une clôture. Parmi les items, le premier est non marqué et le deuxième signalé par la conjonction de coordination *et*.

Exemple 82

0 On distingue donc schématiquement **deux phases** dans une étude linguistique :

1 la phase d'analyse d'un ensemble fini de données

2 **et** la phase de confrontation des résultats de cette analyse, c'est-à-dire des hypothèses avancées, à la réalité.

Cl. **Il s'ensuit que** le recours explicite au corpus peut intervenir dans une phase liminaire de la recherche au moment où l'on tente de cerner les faits pertinents ou en fin de recherche au moment de valider les hypothèses émises.

Dalbera

Malgré le non marquage, les séries NM (et NMet) structurent le texte d'une manière relativement exacte. Dans l'Exemple 82 ci-dessus aussi, l'amorce exacte, la clôture et la faible longueur des items assurent l'identification facile des items et la cohérence du texte. De même, la répétition du mot *phase* dans l'amorce et les deux items contribue à ce processus. Étant donné cette nature explicite, la fréquence plutôt importante des séries NM dans les articles en linguistique semble refléter le même phénomène que la fréquence des

⁹⁷ La différence de fréquence des NM est statistiquement significative avec le X².

Entre la combinaison de l'histoire et de l'éducation et de la linguistique, la valeur $p = 0,0051$, $X^2 = 7,83$, $N = 618$.

La fréquence de ces séries semble plutôt similaire entre l'histoire et l'éducation avec le test X². La valeur $p = 0,0877$, $X^2 = 2,92$, $N = 375$.

séries homogènes et la rareté des séries PNM ; le marquage de l'organisation textuelle est plus exacte et explicite en linguistique que dans les deux autres domaines.

8.3.2 AMORCES ET CLOTURES

Outre les marqueurs d'item, aussi les amorces et les clôtures peuvent refléter la nature exacte et explicite du marquage des séries linéaires. La Figure 41 ci-dessous expose les fréquences des amorces selon les disciplines. L'étiquette $0=0$ indique l'absence de l'amorce, $0=NUM$ la présence d'une amorce exacte, $0=VAG$ une amorce vague (cf. l'Exemple 80), $0=IMP$ une amorce implicite et $0=Q$ une amorce question. Les types sont discutés en détail à la section 4.2.4 et 5.2.2.

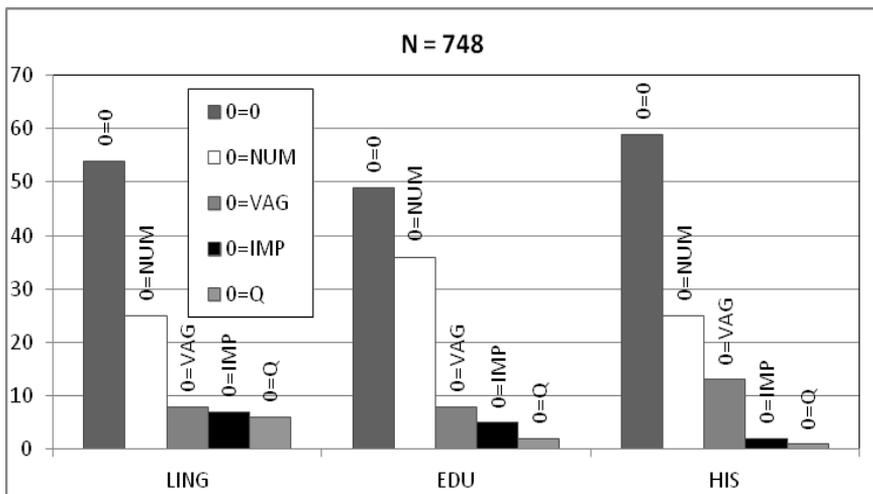


Figure 41 Types d'amorces selon les disciplines dans toutes les séries du corpus en pourcentages du total.

La Figure 41 indique tout d'abord que les séries sans aucune amorce semblent plus fréquentes en histoire (59 %) qu'en linguistique (54 %) et en éducation (49 %). La différence entre les séries en histoire et celles en éducation semblerait également statistiquement significative⁹⁸, bien que celles entre les autres disciplines n'aient pu être prouvées. Pour autant, comme les amorces, le cas échéant, contribuent au marquage des séries linéaires, cette observation conforte nos résultats de la section antérieure sur le marquage moins exact des séries en histoire. De plus, la Figure 41 semble également proposer que les amorces exactes, précisant le nombre d'items de la série, sont plus fréquentes en éducation qu'en linguistique et en histoire. Cette variation disparaît pourtant si les amorces et clôtures exactes sont examinées ensemble⁹⁹.

⁹⁸ Avec le test X2, la valeur $p = 0,0558$, $X2 = 3,66$, $N = 448$.

⁹⁹ Pour les articles en linguistique, 58 % ne sont ni précédées, ni suivies d'une amorce ou clôture exacte alors que le rapport s'élève à 59 % en éducation et à 62 % en histoire.

Les fréquences des clôtures, en revanche, sont présentées dans la Figure 42 ci-dessous. Comme pour la fréquence des amorces, celle des clôtures reflète aussi le marquage moins explicite des articles en histoire. Les séries sans aucun type de clôture sont les plus fréquentes en histoire et les moins utilisées en linguistique¹⁰⁰. En éducation, elles semblent se situer entre ces deux disciplines. Dans la figure, l'étiquette $R=0$ indique l'absence de clôture, $R=NUM$ une clôture exacte, $R=ANA$ une clôture anaphorique, $R=CON$ un connecteur conclusif ou explicatif et $R=ACON$ un connecteur adversatif.

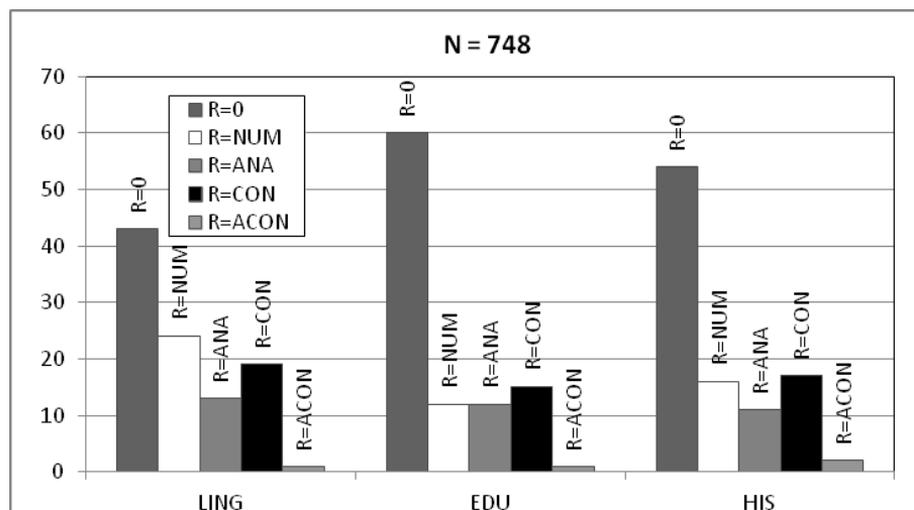


Figure 42 Types de clôtures selon les disciplines dans toutes les séries du corpus en pourcentages du total.

Nos résultats sur les fréquences des amorces et des clôtures soutiennent donc tous deux le marquage moins explicite des séries dans les articles en histoire ; les séries sans aucun type d'amorce ou de clôture sont plus fréquentes en histoire. De plus, l'éducation semble se situer entre elles, similaire à la linguistique au regard des amorces alors que la fréquence des clôtures la relie davantage à l'histoire.

Nous avons démontré maintes fois la façon dont les amorces et clôtures peuvent contribuer au guidage du lecteur dans le texte. L'Exemple 83 est similaire. Énumérant des travaux sur l'apprentissage lors d'échanges sociocognitifs, il se compose de trois items qui tous correspondent à un type de travail : la parution des revues consacrées à la thématique, un colloque et une autre revue scientifique.

¹⁰⁰ Avec le test X², la différence de fréquence des séries sans aucune clôture semble statistiquement significative.

Entre la linguistique et l'histoire, la valeur $p = 0,0178$, $X^2 = 5,61$, $N = 509$;
entre la linguistique et l'éducation, la valeur $p = 0,0002$, $X^2 = 14,41$, $N = 539$.
La différence entre l'histoire et l'éducation n'a pas pu être prouvée.

Exemple 83

- 1 Le numéro du Bulletin de Psychologie « Interaction, acquisition de connaissances et développement » (2003) ainsi que le numéro spécial de l'INETOP « Coopérer et apprendre par le dialogue » (2003) offrent un certain nombre d'exemples de ces recherches visant à mettre en évidence le rôle des échanges dialogiques dans l'appropriation des savoirs scolaires.
 - 2 La tenue récente du colloque « Faut-il parler pour apprendre ? » (IUFM Nord — Pas de Calais, Université Lille 3) qui réunissait des chercheurs de différents horizons montre que cet intérêt est désormais partagé par différentes disciplines, [...].
 - 3 On pourrait **aussi** citer plusieurs articles de la revue de didactique du français *Repères*.
- Cl. **L'ensemble de ces travaux** concourt à analyser comment le sujet apprend au cours d'échanges sociocognitifs et à préciser le rôle des échanges langagiers dans cet apprentissage.

Berzin

Des items de l'Exemple 83, seul le dernier est signalé explicitement, par le connecteur *aussi*. Bien que les items soient compréhensibles sans la clôture, elle nous semble toutefois faciliter la lecture. L'annonce explicite, surtout, du principe d'énumération *ces travaux* aide le lecteur à comprendre la raison de l'énumération de ces items, leur trait commun. De plus, la clôture indique explicitement au lecteur la fin de la série de sorte qu'il n'ait pas à la déduire lui-même.

En définitive, outre les fréquences de toutes les amorces et clôtures, la Figure 41 et la Figure 42 présentent également les fréquences des types d'amorce et clôture plus rares. L'amorce implicite et l'amorce question, principalement, ainsi que les clôtures signalées soit par un connecteur explicatif ou conclusif, soit par un connecteur adversatif, sont toutes relativement peu utilisées. De même que la présence des séries avec des étiquettes particulières, ainsi les séries temporelles et argumentatives, ces types d'amorce et de clôture peuvent pourtant aussi refléter un usage ou marquage différent des séries dans des différents domaines. Pour cette raison, leurs fréquences sont encore répétées ci-dessous dans la Figure 43.

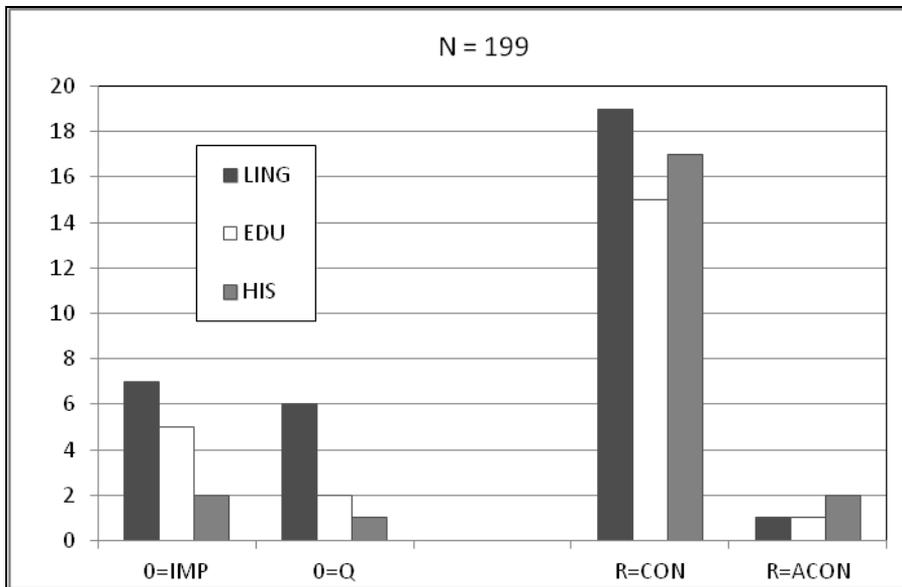


Figure 43 Fréquences des types d’amorce et clôtures rares dans les domaines (en pourcentage du total).

Selon la Figure 43, les fréquences des types d’amorce et des types de clôture présentés semblent varier. Seules les différences entre les amorces sont pourtant significatives. Les amorces implicites indiquant uniquement la thématique discutée dans la série sont plus rares en histoire qu’en linguistique¹⁰¹. L’éducation semblerait encore se situer entre les deux mais les différences n’ont pu être prouvées statistiquement. De plus, les amorces question semblent aussi plus fréquentes en linguistique qu’en histoire et en éducation¹⁰².

La fréquence des deux types d’amorces rares en linguistique reflète surtout l’alternance du marquage selon les différents domaines ; il semblerait donc plus varié en linguistique alors qu’en éducation et histoire seules les façons de marquage plus typiques sont utilisées. Outre sa fréquence supérieure et son caractère plus explicite, le marquage des séries linéaires semble ainsi également plus varié en linguistique. L’Exemple 84 ci-dessous illustre cette variation. Il comprend une série de trois items, tous explicitement signalés par des MIL, et une amorce question à laquelle les items donnent chaque une réponse.

Exemple 84

0 On pourrait se demander pourquoi elles ne sont pas, dès cette époque, intervenues directement dans la question de l’origine des langues.

1 **D’abord**, les créoles sont considérés comme des langues formées ;

¹⁰¹ Avec le test X2, la différence est significative avec la valeur $p = 0,0342$, $X2 = 4,49$, $N = 509$.

¹⁰² Avec le test X2, la différence est statistiquement significative.

Entre la linguistique et l’histoire, la valeur $p = 0,0027$, $X2 = 8,97$, $N = 509$;
entre la linguistique et l’éducation, la valeur $p = 0,0439$, $X2 = 4,06$, $N = 539$.

- 2 **ensuite**, ils invalident les hypothèses biologiques sous la forme racialisiste (une langue = une race) spécifique de l'époque, puisqu'on parle à leur propos de langues hybrides (Adam) ou mixtes (Schuchardt) ;
- 3 **enfin**, ils ne respectent pas le modèle arborescent (1 langue mère, n descendantes par divergence), en mettant en lumière le schéma inverse (n « langues mères», descendante par convergence).

Auroux

8.4 VARIATION INTERCONTINENTALE ?

Dans cette section, nous avons examiné, pour l'heure, la variation de l'usage des séries linéaires dans les trois domaines de notre étude, soit la linguistique, l'éducation et l'histoire. Nous adoptons à présent une autre perspective et menons une brève comparaison sur une autre base : les revues scientifiques françaises et les revues scientifiques canadiennes.

Comme expliqué à la section 5.1.2, l'accessibilité supérieure des revues canadiennes au format électronique nous a menée à inclure dans notre corpus quelques articles parus dans ces revues. Ainsi, notre sous-corpus en éducation intègre 16 articles parus dans des revues françaises plus 14 articles parus au Canada et 24 articles de revues françaises plus 6 de revues canadiennes en histoire. Outre la facilité d'accès, nous avons justifié ce choix par le nombre final relativement faible de ces articles dans notre corpus et par le constat de Fløttum *et al.* (2006 : 10-11, 18), selon qui la similarité des textes écrits dans une langue dépasse les frontières nationales. Ils ajoutent également que la norme française semble valoir pour tous les textes rédigés en français. Dans cette section, nous cherchons à valider ce constat en montrant que l'usage des séries linéaires ne diffère pas d'une manière significative entre les revues françaises et canadiennes, pour ainsi justifier la présence de ces revues dans notre corpus.

L'aspect de comparaison le plus important est évidemment la fréquence des séries. Comme les longueurs des articles varient significativement selon les revues (cf. le Tableau 1 à la page 94), il semble inadéquat de comparer les simples fréquences de série dans chaque article. Pour cette raison, nous avons, pour chaque article, calculé le nombre de séries incluses dans un segment de 1 000 mots¹⁰³. Cette figure facilite la comparaison des fréquences des séries. La Figure 44 ci-dessous présente les fréquences des séries simples et enchaînées dans le sous-corpus éducation et la Figure 45 les mêmes dans le sous-corpus histoire.

¹⁰³ (Nombre de séries / Nombre de mots de l'article) * 1 000

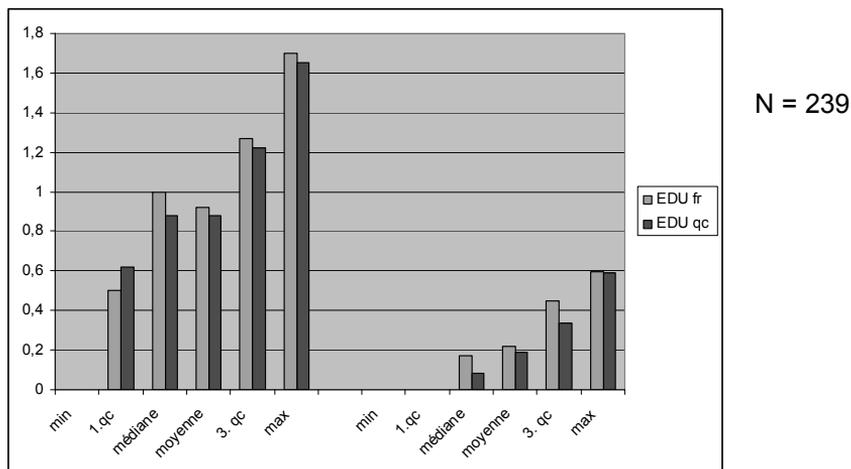


Figure 44 Les séries simples (à gauche) et les enchaînées (à droite) dans les revues françaises et canadiennes dans le sous-corpus éducation.

Au premier abord de la Figure 44 et de la Figure 45, les fréquences des séries dans les revues françaises et canadiennes semblent similaires. Or, dans la Figure 44 représentant les articles en éducation, les figures sont presque systématiquement légèrement plus élevées dans les françaises que dans les canadiennes. La différence est difficile à examiner statistiquement du fait des articles sans aucune série et du nombre limité d'occurrences qui donc dans ce cas correspondent à des articles. La médiane des simples étant de 1,01 pour les françaises et de 0,88 pour les canadiennes et celle des enchaînées de 0,17 pour les françaises et de 0,08 pour les canadiennes, les différences sont cependant minimes. Pour cette raison, l'existence d'une différence réelle nous semble cependant improbable, au moins dans notre corpus. En outre, la fréquence supérieure des séries dans la partie française irait à l'encontre des résultats rapportés dans le projet KIAP (Fløttum *et al.* 2006) ; selon Dahl (2004 : 1821), entre autres, le métadiscours utilisé en français ne représente que 50 % de celui employé en anglais. Ainsi, l'influence des normes anglophones sur les revues canadiennes devrait être patente dans la fréquence supérieure des séries dans celles-ci, et non l'inverse.

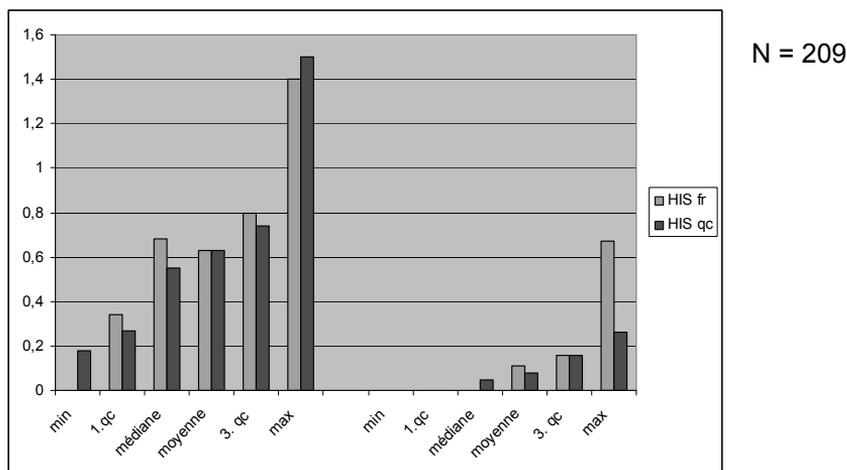


Figure 45 Les séries simples (à gauche) et les enchaînées (à droite) dans les revues françaises et canadiennes dans le sous-corpus histoire.

Pour le sous-corpus histoire aussi, certains chiffres de fréquences semblent légèrement plus élevés dans la partie française que dans la canadienne ; la médiane des simples est de 0,68 pour la première et de 0,55 pour la seconde. À nouveau, la différence est tout de même minime et très probablement irréaliste. Les moyennes identiques des parties et la faible fréquence d'articles soutiennent aussi cette idée. Le faible nombre d'occurrence dérange également l'examen des séries enchaînées. En effet, elles ne sont incluses que dans 10 articles de la partie française et dans trois de la partie canadienne. Les fréquences semblent cependant similaires malgré le chiffre maximal très élevé de la partie française.

Dans notre corpus, pour le moins, les fréquences des séries linéaires semblent donc similaires dans les revues françaises et canadiennes. Évidemment, une étude plus vaste s'impose afin de confirmer globalement cette tendance. Surtout, l'examen détaillé des séries enchaînées nécessiterait un corpus plus important. Cette section devant se limiter à démontrer les tendances de notre corpus, nous n'irons pas plus loin dans cette étude.

Dans les sections précédentes, nous avons comparé, outre les simples fréquences, la qualité du marquage des séries dans les différentes disciplines. Pour terminer cette section et démontrer à nouveau la similarité de l'usage des séries linéaires, nous présentons à la suite les distributions des catégories de série dans les deux types de revue. Comme les séries enchaînées sont très dispersées et présentent peu d'occurrences, nous nous privilégions les séries simples.

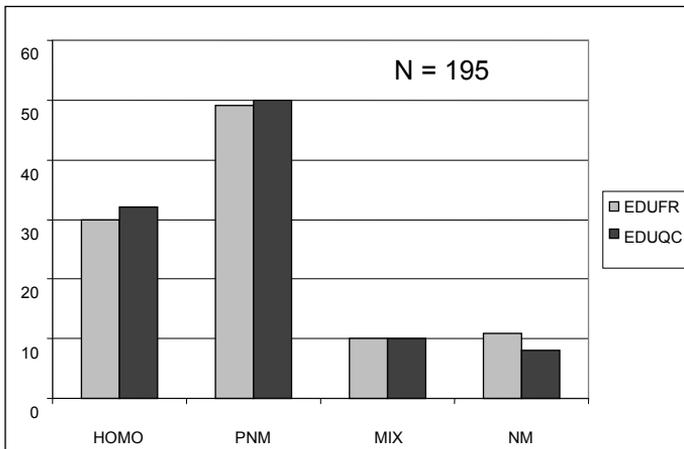


Figure 46 Catégories de séries selon les deux types de revue en éducation.

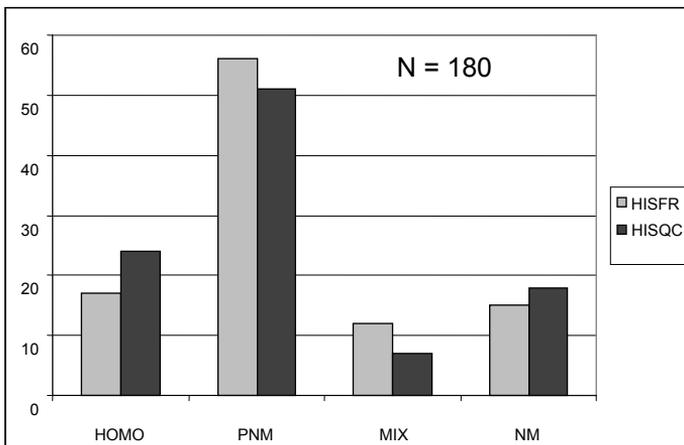


Figure 47 Catégories de séries selon les deux types de revue en histoire.

Les qualités du marquage des séries linéaires des deux sous-corpus sont présentées ci-dessus dans Figure 46 et dans la Figure 47. Comme nous le notons, les distributions dans le sous-corpus éducation, surtout, sont très similaires. Bien que les fréquences des séries non marquées semblent plus élevées dans les revues françaises que dans les canadiennes, la différence n'est pas statistiquement significative¹⁰⁴. Dans le sous-corpus histoire non plus, les variations des catégories des séries simples ne sont pas significatives¹⁰⁵. Ainsi, la façon dont les séries linéaires sont marquées ne semble pas différer entre les revues françaises et canadiennes.

¹⁰⁴ Avec le test X2, la valeur $p = 0,6616$, $X2 = 0,19$, $N = 195$.

¹⁰⁵ Avec le test X2, aucune des différences des catégories n'est statistiquement significative.

Entre les homogènes, la valeur $p = 0,3791$, $X2 = 0,78$, $N = 180$;

entre les PNM la valeur $p = 0,6654$, $X2 = 0,19$, $N = 180$;

entre les séries mixtes, la valeur $p = 0,4838$, $X2 = 0,49$, $N = 180$;

entre les NM, la valeur $p = 0,8123$, $X2 = 0,06$, $N = 180$.

8.5 BILAN : VARIATION INTERDISCIPLINAIRE ET VARIATION INTERCONTINENTALE

Cette section s'est proposée comme but de comparer la fréquence et la variation des séries linéaires dans les trois sous-corpus de notre étude, soit en linguistique, en éducation et en histoire.

Le premier aspect examiné est la fréquence des séries dans les trois sous-corpus. Comme prévu, les séries linéaires sont plus rares en histoire et plus utilisées en linguistique et en éducation. De plus, les séries temporelles s'organisant sur un axe chronologique sont aussi très fréquentes dans les articles d'histoire. Le style de ce domaine ayant souvent été décrit comme narratif et ses articles discutant d'événements du passé, ces deux résultats sont attendus. Or, simultanément, ils indiquent que les séries linéaires sont toujours aussi relativement utilisées en histoire et que l'organisation textuelle s'y montre marquée, malgré le style et la structure différents de ces textes. L'importance des séries linéaires ne doit donc pas être sous-estimée dans ce domaine non plus.

Outre la fréquence du marquage, nous avons également comparé la variation de la qualité des séries linéaires. Selon cette perspective également, l'histoire semble se distinguer des deux autres domaines, bien qu'aussi l'éducation diffère de la linguistique sous certains aspects. En histoire, les séries longues s'étendant sur plusieurs sections titrées, surtout, sont rares. Les séries section dont les items correspondent aux sections titrées sont absentes et les séries méta composées d'une série section et d'une série métadiscursive présentant explicitement la structure de la série section sont moins utilisées. D'une part, la faible fréquence de ces deux types de série peut s'expliquer par la structure différente des articles en histoire ; si les sections ne suivent pas un principe d'énumération, il leur est impossible de former une série linéaire. D'autre part, le manque des séries métadiscursives, surtout, peut refléter la rareté du métadiscours. En effet, parmi toutes les séries linéaires, ce sont elles qui guident le lecteur et se réfèrent au texte lui-même le plus explicitement.

La présence des items non marqués et des séries signalées par le même type de marqueurs, également, reflète la variation de la qualité des séries linéaires et la différence de l'histoire par rapport aux deux autres domaines. Les séries avec des items non marqués, surtout, sont très fréquentes dans ce domaine. Les séries dont tous les items sont explicitement signalés sont, elles, plus rares. Outre leur rareté, les séries linéaires sont donc, dans ce domaine, moins explicitement signalées. La fréquence des séries sans aucune amorce et de celles sans aucune clôture en histoire conforte également cette idée.

C'est également la qualité du marquage qui sépare les articles de la linguistique et de l'éducation. En effet, certaines caractéristiques de l'éducation la rapprochent de l'histoire et, par conséquent, des sciences humaines. Dans ce domaine aussi, les séries section sont plus rares qu'en linguistique et les séries temporelles aussi fréquentes qu'en histoire. Dans

le cas de l'éducation, elles décrivent souvent les phases de l'analyse effectuées pour l'article. Pareillement, le marquage des séries linéaires dans le sous-corpus éducation ressemble à celui de l'histoire ; les séries avec des items non marqués y sont plus fréquentes qu'en linguistique.

Des disciplines étudiées, la linguistique se caractérise par un marquage plus exact ; les séries homogènes composées d'items explicitement signalés et celles non marquées dont le nombre d'items est signalé avant ou après la série y sont très fréquentes, alors que les séries avec items non marqués sont plus rares. L'importance des séries linéaires en linguistique est visible également dans la variation des types d'amorce précédant les séries. En effet, tandis que les amorces en éducation et en histoire suivent les types les plus fréquents, celles en linguistique adoptent des formes plus variées.

Bref, les domaines examinés diffèrent donc en fréquence et en qualité des séries linéaires. La Figure 48 résume ces caractéristiques.

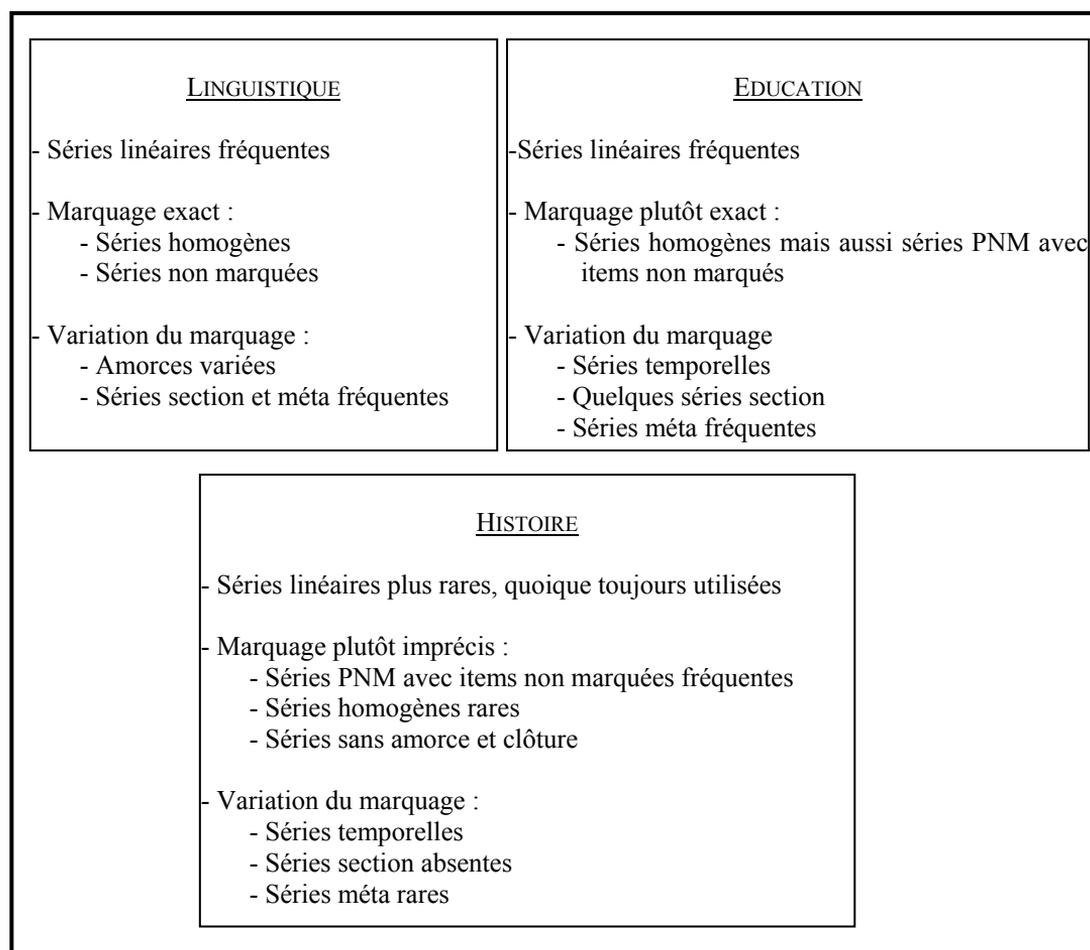


Figure 48 Résumé des caractéristiques des domaines examinés.

Le résultat de l'histoire était attendu, comme nous l'avons déjà constaté. La différence entre la linguistique et l'éducation, en revanche, concerne plus la qualité du marquage et est plus imprévue. Comme les travaux sur le métadiscours et le marquage de l'organisation textuelle en général se concentrent sur la fréquence des deux, nos résultats s'avèrent cependant difficilement comparables. Serait-il possible que la nature plus implicite du marquage soit corrélée avec la rareté des séries linéaires, comme c'est le cas pour l'histoire ? Ou s'agit-il de deux caractéristiques indépendantes ? Nos résultats en éducation conforteraient cette dernière hypothèse : la fréquence des séries y est similaire à celle de la linguistique mais la qualité du marquage, en revanche, est moins explicite.

9. POUR CONCLURE

La question centrale dans cette thèse a été l'organisation textuelle et les façons dont on la marque (ou non). Afin d'aborder cette question, nous avons examiné un mode d'organisation textuelle, les séries linéaires, qui sont des structures organisant un segment textuel en items qui se suivent d'une façon linéaire dans le texte et dont au moins une partie est signalée par des indices linguistiques. Ces indices marquent l'organisation de la série de sorte que le lecteur ne doive pas faire l'effort de l'interpréter lui-même.

Dans l'examen des séries linéaires, nous avons adopté une perspective fonctionnelle et quantitative. Ainsi, au lieu d'examiner la façon dont elles devraient être utilisées, nous nous sommes focalisée sur l'usage des séries linéaires dans la pratique. De plus, nous avons examiné les séries à partir de la fonction qu'elles ont dans l'organisation textuelle et non à partir d'un groupe de marqueurs spécifique (cf. Péry-Woodley 2000 :9, 134-135). Cette approche nous permet un regard plus global sur l'organisation textuelle et sur la variété des marqueurs utilisés à cette fin. Par ailleurs, elle rend possible l'examen des endroits où les marqueurs explicites ne sont pas utilisés et où le lecteur doit lui-même interpréter la façon dont la série est structurée.

9.1 QUESTIONS, HYPOTHESES ET REPONSES

Notre thèse s'est proposé de répondre à trois questions de recherche, dont la première concerne les descriptifs généraux des séries linéaires, notamment leur fréquence, marquage et longueur.

1. Comment les séries linéaires participent-elles à l'organisation textuelle ?

A ces premières questions, nous avons répondu dans la section 6. En particulier, nous avons montré que les séries linéaires présentent un mode d'organisation fréquent et varié, au moins dans les articles de recherche que nous avons examinés. Dans ce sens, nos résultats valident donc notre hypothèse 1.

La variété des séries repérées est manifeste notamment dans la variation du marquage et de la longueur des séries, ainsi que dans la fréquence des séries avec une étiquette particulière organisant le texte d'une façon spécifique. Tout d'abord, comme nous l'avons présumé, la longueur des séries varie de très courte à très longue. Bien que la majorité des séries soit plutôt courte et n'excède pas 130 mots, notre corpus comprend également 111 séries qui dépassent 500 mots, ce qui correspond à une page environ. Cela dit, les séries dont la longueur excède 2000 mots semblent, dans notre corpus, se composer de sections titrées. Cette limite paraît donc la longueur maximale que les séries et leurs marqueurs peuvent atteindre dans les articles de recherche sans que le marquage soit soutenu par des titres.

Les séries *section* dont les items correspondent aux sections titrées sont un type de série auquel nous avons, en raison de son marquage particulier, attribué une étiquette particulière. Ces étiquettes sont également des signes de la variété des structures repérées dans notre corpus. D'autres types de série avec une étiquette sont les séries *méta*, composées d'une série section et d'une autre série qui la précède et en présente les items d'une façon cataphorique, les séries *temporelles*, énumérant des items ayant eu lieu dans le monde extérieur dans un ordre chronologique, les séries *accumulatives*, composées d'items dont une partie correspond au principe d'énumération de la série et une autre partie y présente des contre-arguments, des concessions ou des explications, et les séries *à deux temps*, composées de deux séries inférieures dont la seconde dépend thématiquement de la première. De plus, comme nous l'avons présumé dans l'introduction, nous avons repéré quelques séries *argumentatives* composées d'une thèse originale, d'une antithèse et d'une conclusion. En tout, les séries étiquetées couvrent 19 % du total de séries dans notre corpus, les plus fréquentes étant les séries section et méta ainsi que les séries temporelles. Par conséquent, malgré cette variété, les séries typiques sont tout de même les plus fréquentes.

Deuxièmement, en ce qui concerne le marquage des séries repérées, nos résultats soulignent la fréquence du marquage partiel des séries linéaires et de l'organisation textuelle. Dans les séries simples, presque la moitié (47%) des structures incluent au moins un item non marqué, et seulement un tiers (29%) ressemble au modèle souvent donné des séries linéaires, dans lequel tous les items sont signalés par le même type de marqueurs. Ces résultats sont très similaires à ceux de Taboada (2006, aussi Mann & Taboada 2007) sur des textes journalistiques et des conversations. Ils indiquent que 60 à 70 % des relations de cohérence ne sont pas signalées. Le marquage de l'organisation des articles de recherche ne semble donc pas différer de celui de ces genres.

Contrairement aux séries simples, les séries enchaînées, composées de plusieurs séries inférieures, sont plus explicitement signalées. Dans celles-ci, le marquage explicite de tous les items et l'annonce du nombre d'items de la série par une amorce ou clôture sont plus fréquentes, ainsi que les expressions anaphoriques et les séries dans lesquelles tous les items sont non marqués. D'une part, nous estimons que cette exactitude s'explique par la structuration relativement complexe de ces séries ; vu qu'elles sont composées d'un nombre d'items plutôt élevé mais que leur longueur n'excède pas celle des séries simples, le lecteur peut avoir besoin de guidage afin de faciliter la lecture. D'autre part, la fréquence des séries non marquées et des expressions anaphoriques illustre la structuration interne de ces séries, qui se composent souvent des parties inférieures dont l'une fonctionne comme point de référence pour l'autre.

Comme l'hypothèse 2 concernant le marquage, nous avons avancé que le marquage des séries linéaires serait accompli avec une rigueur certaine en raison du processus d'édition que les articles ont parcouru et en raison du rôle important que le marquage d'organisation textuelle a dans les articles de recherche. Pour ces raisons, nous avons estimé que l'usage

des marqueurs questionnables, tel que *d'une part sans d'autre part*, ainsi que les séries non parallèles dans lesquels la longueur des items, le type de marqueurs et la structure interne des items varient seraient plutôt rares. De ce point de vue, nos résultats soulignant le marquage partiel sont plutôt surprenants.

Or, bien que nos résultats démontrent de la fréquence du marquage partiel, ils attestent simultanément que le non marquage ne rend pas la série directement non parallèle. Ainsi, contrairement à ce que nous avons présumé, les séries consistant en items complètement ou partiellement non marqués peuvent tout de même être faciles à repérer et se composer d'items équivalents. Elles ne sont donc pas par défaut non parallèles. Plutôt, dans ces séries, l'ordre n'est pas nécessairement un critère d'organisation important, et la structure ressemble plus aux énumérations listant des items non ordonnés. Or, il n'empêche que les séries avec des items non marqués peuvent également être non parallèles. En particulier, notre analyse a montré que la variation de la longueur des items et leur différence syntaxique ou thématique, ainsi que le changement du principe d'énumération, peuvent avoir cet effet. Outre les séries non marquées et partiellement non marquées, ces caractéristiques peuvent également avoir lieu dans les séries signalées par différents types de marqueurs. De celles-ci, surtout les structures dans lesquels au moins un des marqueurs est un connecteur et un autre une expression anaphorique ou une construction numéro+nom fonctionnant comme sujet ou objet de la phrase sont souvent, selon, notre analyse non parallèles.

Contrairement aux séries incluant des items non marqués, l'usage des marqueurs questionnables est pratiquement absent dans notre corpus. Ce résultat s'explique pourtant sans doute par notre méthode d'annotation. Comme nous l'avons remarqué en définissant le schéma d'annotation, nous avons exclu de l'analyse les séries dont nous n'avons pas été capable d'interpréter au moins deux items. Ainsi, l'absence des marqueurs questionnables dans le corpus ne reflète pas nécessairement leur présence réelle parce qu'ils peuvent être exclus de l'annotation dès le départ. La fréquence des séries non parallèles et du non marquage dans les articles qui ont été révisés et édités par la revue, par contre, semble signaler que les lecteurs ne les considèrent pas comme incohérentes. Sinon, les revues les auraient édités et ces structures ne seraient plus présentes dans la version finale de l'article. Ainsi, l'usage de ces structures qui ne suivent pas le modèle typique du marquage n'est pas questionné. Ce résultat correspond à la remarque de Péry-Woodley (2000 : 139), qui constate que les séries sont facilement acceptables bien que leur marquage et structure soient « déviants ».

Ensuite, notre deuxième question de recherche a eu comme objectif d'explorer le marquage des séries linéaires plus en détail, et d'examiner des facteurs textuels qui puissent expliquer l'usage d'un marqueur particulier ou le non marquage de l'item. De plus, nous avons abordé l'identification des items non marqués. Leur examen peut également expliquer le non marquage, car la présence des indices complémentaires qui

contribuent à l'identification des items mais ne sont pas annotés comme marqueurs explicites peut rendre l'usage des marqueurs explicites inutile.

2. Pourquoi les séries linéaires sont-elles signalées comme elles le sont?

Notre hypothèse sur les raisons derrière le marquage, soit l'hypothèse 3, a concerné uniquement les types de marqueurs utilisés dans des segments textuels de longueurs différentes. Suivant Ho Dac (2007) et Jackiewicz et Minel (2003), nous avons présumé que les connecteurs fonctionnent pour signaler des segments d'une longueur plutôt restreinte qui s'étend sur un seul, ou au maximum sur quelques paragraphes. De même, nous avons avancé comme hypothèse que, comme l'a noté Schnedecker (2000), les points de référence des anaphoriques se trouvent dans la proximité, et par conséquent les séries introduites par ces marqueurs seraient plutôt courtes.

Nos résultats ont validé ces deux suppositions ; les expressions anaphoriques fonctionnent dans les séries courtes, et aussi les séries introduites par des connecteurs sont au maximum de longueur intermédiaire. De plus, nous avons également démontré que les connecteurs peuvent se distinguer entre eux ; tandis que les marqueurs en *-ment* fonctionnent pour signaler des segments plus longs, les connecteurs d'origine temporelle surtout introduisent des segments plus courts. Les segments les plus longs, par contre, ne sont pas signalés dans notre corpus par des connecteurs mais par des constructions numéro+nom composées en général d'un ordinal et d'un nom. Le plus typiquement, ces marqueurs sont de forme *le premier exemple* ou *un autre problème*.

Or, nos résultats les plus intéressants sur le marquage des séries concernent son exactitude et la présence des items non marqués. En particulier, notre analyse a montré que toutes les séries ne visent pas à guider le lecteur de la même manière et que certaines séries ont comme objectif de le faire de façon moins exacte et moins explicite que d'autres. On peut donc dire que les séries réalisent des stratégies textuelles différentes, que l'auteur les utilise dans des contextes variés afin que l'organisation du texte corresponde le mieux au but communicatif de l'auteur (Virtanen 1992 : 51 ; Enkvist 1975 : 18-22, 1987). Les séries dans lesquelles tous les items sont marqués par le même type de marqueurs sont souvent attachées à une amorce ou clôture, qui répètent encore la façon dont la série est organisée. De plus, les marqueurs d'item exacts, précisant la place exacte de l'item dans la liste, sont plus fréquents dans ces séries. Ces séries guident donc le lecteur de façon exacte et explicite. Dans les séries incluant des items non marqués ou des types de marqueurs différents, par contre, les amorces et clôtures sont plus rares bien qu'elles puissent y être utiles afin de préciser le marquage manquant de la série proprement dite. De même, les marqueurs d'item additifs, ne signalant que l'addition de l'item dans la liste, sont plus utilisés dans ces séries. L'organisation pourvue par ces séries est ainsi plutôt imprécise et moins explicite.

La stratégie textuelle que la série réalise est donc un facteur qui selon notre analyse influence son marquage. De même, les types de marqueurs sont utilisés pour signaler des

items des longueurs différentes. De plus, notre analyse indique que la façon dont la série est marquée se corrèle également avec la longueur de la structure ; les séries les plus courtes sont non marquées (ou un des items est introduit par une conjonction de coordination), les séries de longueur intermédiaire sont composées d'items signalés par le même type de marqueurs, et les séries les plus longues sont partiellement non marquées ou introduites par des marqueurs différents. La faible longueur des séries non marquées est logique vu que l'organisation de ces séries est déjà indiquée dans la proximité par une amorce et / ou une clôture. Par contre, la longueur plus importante des séries partiellement non marquées dont le marquage est aussi autrement plutôt imprécis, est plutôt contre-intuitive. Elle est également contraire aux résultats d'un travail antérieur (Jackiewicz 2005). Intuitivement, on pourrait croire que le marquage de l'organisation des segments plus vastes devrait être plus exact et plus explicite pour que le lecteur puisse la mémoriser plus facilement.

Sans essayer d'expliquer les raisons derrière le marquage plus imprécis et implicite des séries plus longues, nous avons remarqué que ce phénomène se corrèle bien avec le fonctionnement des constructions numéro+nom. Selon notre analyse, ces marqueurs introduisent les segments les plus longs dans notre corpus, et ils nous semblent également plus discrets en tant que marqueurs que les adverbiaux et les connecteurs en général placés à l'initiale de la phrase. Ainsi, si nous acceptons nos résultats sur le marquage plus imprécis et implicite des segments plus longs, et notre hypothèse sur la nature plus discrète des constructions Num+N, nous pourrions proposer que ces marqueurs sont peut-être les mieux adaptés pour fonctionner dans ce contexte, dans lequel le marquage est aussi autrement moins exact et moins explicite.

Enfin, la dernière partie de notre analyse sur les facteurs textuels influençant le marquage des séries linéaires a été consacrée aux items non marqués et à leur identification. En effet, nos résultats soutiennent ceux que nous avons obtenus sur les différentes stratégies textuelles adoptées par les séries. Les indices complémentaires de sériation pouvant faciliter l'identification des items non marqués, tels que la répétition de mêmes items lexicaux devant tous les items, sont les plus rares dans les séries partiellement non marquées, qui sont aussi autrement signalées de façon plus implicite et plus imprécise. Dans ces structures, ces indices ne peuvent donc pas expliquer le non marquage. En revanche, dans les séries courtes composées d'items non marqués, dont le nombre d'items est précisé par une amorce, une clôture ou l'autre partie de la série enchaînée, les indices complémentaires aussi sont plus fréquents. Dans ces séries courtes, ces indices peuvent également expliquer le non marquage ; vu que les séries sont courtes, les items facilement repérables grâce à ces indices et leur nombre indiqué dans la proximité, le marquage explicite pourrait être répétitif ou inutile.

Finalement, pour notre troisième question de recherche, nous avons adopté une perspective qui diffère légèrement de celle des deux premières questions ; nous avons visé

à comparer l'usage des séries linéaires dans trois sous-domaines de notre corpus, la linguistique, l'éducation et l'histoire.

3. L'usage des séries linéaires varie-t-il selon la discipline scientifique ?

Comme hypothèse 3, nous avons présumé que les séries linéaires sont moins fréquentes dans les articles appartenant au domaine de l'histoire que dans ceux des deux autres disciplines. Nous avons basé cette supposition sur les travaux antérieurs soulignant la fréquence des structures narratives dans le domaine (par exemple Holmes 1997 ; Silver & Bondi 2004 ; Bondi 2009a), ce qui pourrait suggérer que l'organisation explicite du texte en items qui se suivent y serait plus rare. En ce qui concerne la différence potentielle de la linguistique et de l'éducation, nous n'avons pas pris position sur la question parce que les deux disciplines semblent se ressembler sous plusieurs aspects.

Nos résultats sur la variation des séries linéaires selon les disciplines ont bien validé notre hypothèse : les séries linéaires sont plus rares dans les articles en histoire qu'en linguistique et en éducation, et aussi la qualité des séries y est différente. Or, de plus, quelques aspects qualitatifs des séries rapprochent l'éducation de l'histoire et ainsi suggèrent qu'elle est plus typique comme science humaine que la linguistique.

Tout d'abord, le domaine de l'histoire se caractérise par la fréquence des séries temporelles s'organisant sur un axe chronologique, par la rareté des séries très longues composées de sections titrées, et par un marquage d'organisation textuelle plutôt implicite, qui dans la pratique signifie la rareté relative des séries signalées par le même type de marqueurs et la fréquence des séries avec des items non marquées. Pour ce qui est des séries temporelles, leur fréquence est prévisible vu l'objectif de l'histoire de décrire les événements du passé. Quant au marquage plutôt implicite des séries, il peut s'expliquer par le fait que le marquage de l'organisation textuelle est moins important dans ce domaine. Toutefois, malgré ces résultats, les séries linéaires de ces articles nous ont surprise par leur fréquence. Bien qu'elles soient plus rares que dans les autres domaines étudiés, un article en histoire inclut en moyenne sept séries linéaires. Il ne serait donc pas adéquat de dire que le marquage de l'organisation textuelle y serait extrêmement rare ou absent.

Une explication potentielle pour la fréquence relative des séries linéaires en histoire peut être que l'auteur ne les considère pas comme indices métadiscursifs explicites mais les utilise comme partie inhérente du texte en tant que marqueurs de cohésion. La rareté des séries méta présentant explicitement la structure et la thématique des sections de l'article soutient cette hypothèse, car contrairement aux séries linéaires typiques, ces structures sont extrêmement rares dans ce domaine. La différence possible entre les séries guidant le lecteur de façon explicite et celles qui feraient plutôt partie de la cohésion doit pourtant être étudiée plus en détail.

Comme deuxième résultat de la comparaison interdisciplinaire, nous avons constaté que l'éducation se rapproche de l'histoire sous plusieurs aspects, et qu'elle semble ainsi une science humaine plus typique que la linguistique. En particulier, de même qu'en histoire, le marquage des séries linéaires y est moins exact et les séries longues couvrant des sections titrées plus rares qu'en linguistique. De plus, les séries temporelles sont également plutôt fréquentes dans ce domaine, où elles décrivent souvent des phases d'analyse ayant eu lieu l'une après l'autre, dans un ordre chronologique.

Enfin, la linguistique, en revanche, diffère clairement des deux autres disciplines étudiées : c'est elle qui se caractérise par le marquage le plus exact, les séries signalées par le même type de marqueurs y étant les plus fréquentes et les séries avec des items non marqués les plus rares. De plus, la variété des séries différentes y est la plus importante, ce qui semblerait suggérer que les séries y ont un rôle plus important. Notamment, à l'exception des séries temporelles, les types de série particuliers, tels que les séries section, méta et argumentatives, y sont les plus utilisés. Les différentes amorces sont aussi plus fréquentes que dans les autres domaines. Il est cependant difficile d'estimer si ces caractéristiques sont typiques uniquement de la linguistique ou si elles sont communes à un groupe de disciplines plus grand. Cette question, ainsi que la position plus exacte de cette discipline à l'intersection des sciences sociales et humaines, constituerait un objet d'étude très intéressant.

9.2 DISCUSSION

Notre étude a visé à examiner le marquage de l'organisation textuelle dans une perspective textuelle et fonctionnelle, et surtout à aborder ce sujet du point de vue de l'usage de la langue. Pour accomplir ces objectifs, nous avons examiné les séries dans un corpus dans lequel nous avons manuellement annoté toutes les structures qui ont correspondu à nos critères et à notre définition.

A notre avis, ce sont surtout l'usage du corpus et l'examen de la façon dont les séries linéaires sont utilisées dans la pratique qui nous ont apporté les résultats les plus intéressants. En particulier, la variété et les proportions des séries linéaires repérées, les différentes stratégies textuelles réalisées par elles ainsi que le marquage plus implicite et imprécis des séries plus longues révèlent des aspects importants de l'organisation textuelle que les travaux antérieurs n'ont pas encore expliqués de façon détaillée. Ainsi, notre travail a démontré également le potentiel et la nécessité des travaux quantitatifs ; sans l'appui d'un corpus, nos résultats seraient difficiles à présumer ou à reproduire.

Or, l'examen du corpus et notamment le processus d'annotation qu'il requiert nous ont également posé des problèmes, et cette méthode n'est pas sans inconvénients. En particulier, la façon dont nous avons défini les séries linéaires et dont nous avons distingué celles qui sont incluses dans l'analyse et celles qui ne le sont pas est une question problématique qui risque d'avoir un effet sur les résultats du travail mais pour laquelle il

n'existe pas de solution idéale. Bien que nous soyons relativement contente de notre choix d'utiliser dans l'annotation des critères de surface basés sur la fonction des marqueurs explicites, nous admettons que cette méthode peut être questionnée. Surtout, nous avons trouvé problématiques les séries marginales dont l'identification ou le fonctionnement comme série linéaire paraît douteux. Quoique nous les ayons parcourues ensemble avec notre directeur de thèse, une meilleure solution aurait été une double-annotation et la discussion des difficultés de définition dans un groupe de recherche. Cette méthode aurait également permis de mesurer la qualité du corpus avec un accord inter-annotateur. Malheureusement, elle n'a pourtant pas été possible pour nous. Notre seul réconfort est la rareté relative des structures douteuses dans un corpus de taille plutôt importante ; grâce au nombre élevé de séries dans notre corpus, nous pouvons estimer que les séries douteuses n'ont pas d'effet réel sur l'ensemble de nos résultats.

Néanmoins, si nous avons la possibilité, nous ne changerions pas les critères de surface que nous avons établis pour les séries linéaires. Outre l'absence d'un groupe de recherche et d'une annotation du corpus par plusieurs personnes, le seul défaut majeur de notre corpus, à notre avis, est l'absence d'information sur les paragraphes dans le XML. Ce manque nous semble injustifié surtout parce que la conservation de cette information aurait été possible si nous avions été plus informée des aspects techniques de l'annotation. De plus, évidemment, on peut toujours questionner l'influence des articles et des revues choisis sur les résultats. Comme notre comparaison des revues canadiennes et françaises a montré que les différences entre celles-ci sont improbables, nous ne croyons pourtant pas que leur changement ait eu d'effet remarquable sur nos résultats.

Une fois annoté et contrôlé, le corpus a cependant aussi des avantages, dont l'une des plus importantes est qu'il peut être utilisé plusieurs fois pour plusieurs études. Ainsi, amélioré ou dans son état actuel, notre corpus permettrait l'examen d'un nombre de questions linguistiques. Par exemple, pour mentionner quelques aspects intéressants, une question serait les fins des séries linéaires, qui ont été difficiles à déterminer et dont Charolles (1997) a également fait une remarque similaire en examinant les fins des univers de discours ouverts par des expressions cadratives. De plus, une autre question serait l'examen plus détaillé des séries linéaires et leur structure thématique. Lors de l'analyse, nous avons noté plusieurs fois que la série s'organise selon l'un des modèles de la progression thématique de Danes (1974). L'examen quantitatif de cette correspondance et de la fréquence du modèle à thème constant et de celui à thème dérivé dans les séries pourrait encore clarifier leur fonctionnement dans le texte.

Une autre question que notre thèse a laissée ouverte touche aux séries non parallèles, à leur fréquence et définition. Bien que nous ayons identifié plusieurs caractéristiques de ces séries et ayons estimé leur fréquence, elles méritent un examen plus détaillé. Evidemment, nous avons déjà constaté que le parallélisme semble plutôt une propriété relative qu'absolue et que la distinction des séries parallèles et non parallèles paraît difficile. Ainsi,

il se peut que la définition exacte de ces structures et de leur fréquence soit impossible ou même inutile. Tout de même, ces questions nécessitent une étude plus ample.

Une étude de corpus telle que la nôtre a comme objectif l'usage de la langue, ce qui, au moins pour nous, a signifié l'examen des usages ou des pratiques les plus fréquents. Ainsi, comme nous l'avons déjà mentionné lors de l'analyse, nous nous sommes concentrée sur les façons dont, par exemple, le marquage des séries les plus longues est signalé en général, le plus fréquemment. Par conséquent, nos résultats ne sont pas facilement comparables avec ceux issus des études se concentrant entre autres sur la portée sémantique que les expressions cadratives peuvent avoir, ou sur la capacité organisatrice que les connecteurs peuvent adopter. Notre corpus peut pourtant offrir des réponses à ces questions, qui concernent en particulier la variété des usages d'un marqueur ou ses fonctionnements extrêmes. Cette variation vaudrait également une étude.

Par ailleurs, la version actuelle de notre corpus ou au moins une version étendue permettrait également l'examen plus détaillé d'un nombre de questions sur la variation interdisciplinaire de l'usage des séries linéaires dans le domaine des études sur le discours scientifique. Comme nous venons de le mentionner, la position de la linguistique comme une science sociale ou humaine vaut une étude plus élaborée. De plus, la comparaison de l'usage des séries linéaires dans les différentes revues ainsi que l'examen de la variation personnelle de leur usage seraient possibles même avec le corpus dans son état actuel. Aussi, entre autres, l'examen plus détaillé des articles d'histoire permettrait de mieux déterminer les spécificités de ce domaine qui souvent est simplement caractérisé comme « différent ».

Pour finir, nous voulons encore brièvement discuter les perspectives d'avenir de notre corpus. Nous sommes d'avis que tous les corpus qui sont construits de façon adéquate et dont la construction a été un travail important devraient être publics et librement accessibles et utilisables pour tout le monde. De cette façon, la communauté scientifique pourrait profiter du travail déjà effectué en comparant ses propres résultats à un autre genre discursif et en testant d'autres hypothèses sur un corpus prêt. Tout simplement, l'accessibilité libre et gratuite des corpus bénéficierait à toute la communauté scientifique et ferait avancer le domaine. Pour l'instant, notre corpus n'est pourtant pas encore accessible sur l'internet. Les raisons les plus importantes pour ce statut sont les questions sur les droits d'auteur des revues et des articles que nous devons résoudre. Idéalement, tout le corpus et les annotations pourraient être rechargeables et lisibles sur une page web, et un de nos projets d'avenir vise à cet objectif.

REFERENCES

- Adam, J-M. 1990. *Eléments de linguistique textuelle*. Mardaga : Liège.
- Adam, J-M. 1992. *Les textes: types et prototypes. Récit, description, argumentation, explication et dialogue*. Paris : Nathan.
- Adam, J-M. 1999. *Linguistique textuelle : des genres de discours aux textes*. Paris : Nathan.
- Adam, J-M. 2005. *Analyse de La linguistique textuelle - Introduction à l'analyse textuelle des discours*. Paris : Armand Colin.
- Adam, J. et F. Revaz. 1989. Aspects de la structuration du texte descriptif : les marques d'énumération et de reformulation. *Langue française* 81, 59-98.
- Anscombre, J-C. et O. Ducrot 1997. *L'argumentation dans la langue*. Mardaga : Belgique.
- Apothéloz, D. 1994. *Structuration de la cohésion textuelle. Etudes sur le développement et fonctions de l'indexicalité interne*. Université de Neuchâtel.
- Apothéloz, D. 1995. *Rôle et fonctionnement de l'anaphore dans la dynamique textuelle*. Genève : Librairie Droz.
- Ariel, M. 1990. *Accessing Noun-Phrase Antecedents*. London : Routledge.
- Ariel, M. 2001. Accessibility theory: An overview. In : Sanders, T., J. Schliperoord & W. Spooren (ed.) *Text representation*. Amsterdam : John Benjamins, 29-87.
- Asher, N. 1993. *Reference to abstract objects in discourse*. Dordrecht : Kluwer Academic Publishers.
- Asher, N. & A. Lascarides 2003. *Logics of Conversation*. Studies in Natural Language Processing. Cambridge University Press.
- Auricchio, A. et al. 1995. L'anaphore démonstrative à fonction résomptive. *Pratiques* 85, 27-52.
- de Beaugrande, R. et W. Dressler 1981. *Introduction to Text Linguistics*. Longman : London.
- Becher, T. et P.R. Trowler 2001. *Academic tribes and territories. Intellectual enquiry and the culture of disciplines*. University of Wisconsin Press.
- Beke, R. 2005. El metadiscurso interpersonal en artículos de investigación. *Revista Signos* 38(57), 7-18.
- Beke, R. & A. Bolivar 2009. Certainty and commitment in the construction of academic knowledge in the humanities. In : Suomela-Salmi, E. & F. Dervin (ed.) *Cross-linguistic and Cross-cultural Perspectives in Academic Discourse. Vol.2*. Amsterdam: Benjamins, 33-48.
- Berrendonner, A. & M.-J. Reicher-Béguelin 1996. De quelques adjectifs à rendement anaphorique : premier, dernier, autre. In : Gross, G., P. Lerat & C. Molinier (ed.) *Studi di linguistica teorica e applicata, XXV/3*.
- Bestgen, Y. & W. Vonk 1995. The role of temporal segmentation markers in discourse processing. *Discourse Processes* 19, 385-406.
- Bestgen, Y. & W. Vonk 2000. Temporal adverbials as segmentation markers in discourse comprehension. *Journal of Memory and Language* 42, 74-87.

- Bestgen, Y., L. Degand, W. Spooren 2006. Toward Automatic Determination of the Semantics of Connectives in Large Newspaper Corpora. *Discourse Processes* 41(2), 175-193.
- Biber, D. 1995. *Dimensions of register variation : a cross-linguistic comparison*. Cambridge U.P.
- Biber, D, S. Conrad, R. Reppen 1998. *Corpus linguistics : investigating language structure and use*. Cambridge University Press.
- Bilger, M. 2003. Coordination: entre l'ordre et le désordre. In : Combettes, B., C. Schnedeker, A. Theissen (ed.) *Ordre et distinction dans la langue et dans le discours*. Paris : Champion, 29-41.
- Blanche-Benveniste, C. & A. Chervel 1966. Recherches sur le syntagme substantif. *Cahiers de Lexicologie* IX(2), 3-37.
- Bondi, M. 2009a. Polyphony in academic discourse: a cross-cultural perspective on historical discourse. In : Suomela-Salmi, E. & F. Dervin (ed.). *Cross-linguistic and Cross-cultural Perspectives in Academic Discourse. Vol.2*. Amsterdam: Benjamins, 83-108.
- Bondi, M. 2009b. Writing Economic History: The Narrator and the Arguer. In : *Commonality and Individuality in Academic Discourse*. Germany: Peter Lang, 163-190.
- Bouraoui, J-L. & N. Vigouroux 2003. Les marqueurs des structures énumératives sur le web : analyse pour la transmodalité. *Conférence Internationale sur le Document Electronique (CIDE 6)*. Paris : Europia, 199-217.
- Bras, M. & A. Le Draoulec 2009. *D'abord* marqueur de structuration du discours. *Journal of French Language Studies* 19, 229-248.
- Breivega, K.R. 2003. *Vitskaplege argumentasjonsstrategiar*. Oslo : Norsk Sakprosa.
- Brown, G. and G. Yule 1983. *Discourse Analysis*. Cambridge University Press.
- Busquets, J., L. Vieu, N. Asher. 2001. La SDRT: Une approche de la cohérence du discours dans la tradition de la sémantique dynamique. *Verbum* 231, 73–101.
- Carlson, L. D. Marcu & M. E. Okurowski 2003. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In : van Kuppevelt, J. & R. Smith (ed.) *Current Directions in Discourse and Dialogue*. Kluwer Academic Publishers.
- Carr, E. 1961. *What is History?* New York : Vintage.
- Carter-Thomas, S. 2000. *La cohérence textuelle. Pour une nouvelle pédagogie de l'écrit*. Paris : L'Harmattan.
- Chafe, W. 1976. Givenness, contrastiveness definiteness, subjects, topics, ad point of view. In : Charles, N. (ed.). *Subject and topic*. New York : Academic Press, 25-55.
- Chafe, W. 1984. How people use adverbial clauses. In : Brugman, C. & M. Macauley (ed.) *Proceedings of the tenth annual meeting of the Berkeley Linguistics Society*. Berkeley, CA : Berkeley Linguistic Society, 437-449.
- Blublitz, W., U. Uta, E. Ventola (ed.) 1999. *Coherence in spoken and written discourse : how to create it and how to describe it : selected papers from the international workshop of coherence* Augsburg, 24-27 April 1997. Amsterdam : Philadelphia.

- Chafe, W.L. 1992. Discourse: an overview. In : Bright, W. (ed.) *International Encyclopedia of Linguistics, Vol 1*. New York : Oxford University Press, 355-358.
- Chafe, W.L. 1994. *Discourse, consciousness and time: the flow and displacement of consciousness experience in speaking and writing*. University of Chicago Press.
- Charaudeau, P. & D. Maingueneau 2002. *Dictionnaire d'analyse du discours*. Paris : Seuil.
- Charolles, M. 1983. Coherence as a principle in the interpretation of discourse. *Text* 3 83, 71-97.
- Charolles, M. 1988. Les plans de l'organisation textuelle : périodes, portées et séquences. *Pratiques* 57, 3-13.
- Charolles, M. 1995. Cohésion, cohérence et pertinence du discours. *Travaux de linguistique* 29, 125-151.
- Charolles, M. 1997. Encadrement du discours: univers, champs, domaines et espaces. *Cahiers de Recherche Linguistique* 6, 1-60.
- Charolles, M. 2002. *La référence et les expressions référentielles en français*. Paris : Orphys.
- Charolles, M. 2005. Framing adverbials and their role in discourse cohesion. From connection to forward labelling. *Proceedings of the Symposium on the Exploration and Modelling of Meaning*, Biarritz 14-15. 11. 2005.
- Charolles, M. 2006 Cohérence, pertinence et intégration conceptuelle. In : Lane, Ph. (éd.) *Des discours aux textes*. PUR/CNRS. Publications des Universités de Rouen et du Havre, 39-74.
- Charolles, M. et B. Combettes 1999. Contribution pour une histoire récente de l'analyse du discours. *Langue française* 121, 76-116.
- Charolles, M. et M-P. Péry-Woodley 2005. Introduction. *Langue française* 148 2005(4), 3-8.
- Chomsky, N. 1957. *Syntactic structures*. The Hague : Mouton.
- Crismore, A., R. Markkanen et M. Steffensen 1993. Metadiscourse in persuasive writing. A study of texts written by American and Finnish university students. *Written Communication* 10(1), 39-71.
- Coffin, C. 1997. Constructing and Giving Value to the Past : an Investigation into Secondary School History. In : Christie, F. & J. Martin (ed.) *Genres and Institutions*. London : Pinter, 196-230.
- Coffin, C. 2006. *Historical Discourse*. London : Pinter.
- Combettes, B. *Pour une grammaire textuelle. La progression thématique*. Bruxelles / Paris : Duculot, DeBoeck-Wesmael.
- Condamines, A. 2003. *Sémantique et Corpus Spécialisés : Constitution de bases de connaissances terminologiques*. Mémoire d'Habilitation à Diriger les Recherches. Université Toulouse Le Mirail.
- Condamines, A. 2005. Introduction. In : Condamines, A. (ed.) *Sémantique et corpus : quelles rencontres possibles ?* London : Hermes, 15-38.
- Corblin, F. 1985. Les chaînes de référence: analyse linguistique et traitement automatique. *Intellectica* 5(1), 123-143.
- Corblin F. 1995. *Les formes de reprise dans le discours*. Rennes : Presses universitaires.

- Cornish, F. 1990. Anaphore pragmatique, référence, et modèles du discours. In : Kleiber, G. et J-E. Tyvaert (ed.) *Anaphore et ses domaines*. Recherches linguistiques XIV. Etudes publiées par le centre d'analyse syntaxique. Université de Metz. Faculté de Lettres et Sciences humaines. Paris : Klincksieck, 81-96.
- Cornish, F. 1999. *Anaphora, Discourse and understanding*. Oxford University Press.
- Cornish, F. 2006. Relations de cohérence en discours : critères de reconnaissance, caractérisation et articulation cohésion-cohérence. *Corela, numéro spécial Organisation des textes et cohérence des discours*.
- Crompton, P. 2006. The effect of position on the discourse scope of adverbials. *Text and Talk* 3, 245-279.
- Coulthard, M. (ed.) 1994. *Advances in written text analysis*. London : Routledge.
- Dahl, T. 2004. Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics* 36 (2004), 1807-1835.
- Dahl, T. 2005. The academic author: Traces in Text. In : Bäcklund, I., U. Börestam, U. Melander Marttala & H. Näslund (ed.) *Text i arbete / Text at work. Essays in Honour of Britt-Louise Gunnarsson, 12 January 2005*. Institutionen för nordiska språk vid Uppsala Universitet, 289-297.
- Danes, F. 1974. Functional Sentence Perspective and the Organization of the Text. In : Danes, F. (ed.) *Papers in Functional Sentence Perspective*. Prague : Academia, 106-128.
- Day, M. 2008. *The Philosophy of History: An Introduction*. London : Continuum.
- Ducrot, O. & T. Todorov 1972. *Dictionnaire encyclopédique des sciences du langage*. Paris : Seuil.
- Ducrot, O. et al. 1980. *Les mots du discours*. Paris : Les Éditions de Minuit.
- Eggs, S., P. Wignell & J.R. Martin 1993. The discourse of History Distancing the Recoverable Past. In : Ghadessy, M. (ed.) *Register Analysis: Theory and Practice*. London : Pinter, 75-109.
- Enkvist, N.E. 1975. *Tekstlingvistiikan peruskäsitteitä*. Jyväskylä : Gaudeamus.
- Enkvist, N.E. 1984. Contrastive linguistics and text linguistics. In : Fisiak, J. (ed.) *Contrastive Linguistics, Prospects and Problems*. Berlin, etc. : Mouton, 45-67.
- Enkvist, N.E. (ed.) 1985. *Coherence and composition : a symposium*. Åbo Akademi Tidningsbokhandeln.
- Enkvist, N.E. 1987. A note towards text strategy. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 40(1), 19-27.
- Enkvist, N.E. 1989. From text to interpretability. A contribution to the discussion of basic terms in text linguistics. In : Heydrich, W., F. Neubauer, J. Petöfi & E. Sözer (ed.) *Connexity and Coherence: Analysis of Text and Discourse*. Berlin, New York : Walter de Gruyter, 369-382.
- Fries, P.H. 1983. On the Status of Theme in English: Arguments from Discourse. In : Petöfi, J.S. et E. Sözer (ed.) *Micro and macro connexity of texts*. Hamburg : Buske, 116-152.
- Fradin, B. 1994. Anaphorisation et stéréotypes nominaux. *Lingua* 64, 325-369.

- Fløttum, K. 2006. The typical research article - does it exist? In : Suomela-Salmi, E. & F. Dervin (ed.) *Perspectives inter-culturelles et inter-linguistiques sur le discours académique Cross-cultural and Cross-linguistic Perspectives on Academic Discourse Volume 1*. Department of French Studies, University of Turku, 16-44.
- Fløttum, K. T. Dahl et T. Kinn 2006. *Academic Voices. Across languages and disciplines*. Amsterdam : J. Benjamins Pub. Co.
- Givon, T. (ed.). 1983. *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam : John Benjamins.
- Gjesdal, A. M. 2008. *Étude sémantique du pronom ON dans une perspective textuelle et contextuelle*. Universitetet i Bergen.
- Greenbaum, B-N. 1999. *Studies in English Adverbial Usage*. London: Longman.
- Grice, H.P. 1975. Logic and Conversation. In : Cole, P. & J.L Morgan (éds.) *Syntax and Semantics 3*. New York : Academic Press, 41-58.
- Grize, B. 1974. Du groupement au nombre. *Problèmes de la construction du nombre, Études d'épistémologie génétique XI*. Nendeln/Lienchtenstein : K. Reprint, 69-96.
- Goutsos, D. 1996. A model of sequential relations in expository text, *Text* 16(4), 501-533.
- Guimier, C. 1996. *Les adverbs du français, les cas des adverbs en –ment*. Paris: Orphys.
- Guiraud, A-M. 2008. Les emplois corrélatifs de *parfois, quelquefois et des fois*. *Discours* 2.
- Gundel, J.K, N. Hedberg, R. Zacharski 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2), 274-307.
- Habert, B., A. Nazarenko, A. Salem 1997. *Les linguistiques de corpus*. Paris : Armand Colin.
- Halliday, M. A. K. 1977-2003. Text as semantic choice in social contexts. In : Webster, J. (ed.) *The Collected Works of M.A.K. Halliday (Volume 2): Linguistic Studies of Text and Discourse*. London: Continuum, 23-81. Reprinted from van Dijk, T., J.S. Petöfi, (ed.) *Grammars and Descriptions*. Berlin : Walter de Gruyter, 176-226.
- Halliday, M.A.K. 1985. *An Introduction to Functional Grammar*. USA, Australia : Edward Arnold.
- Halliday, M.A.K. & R. Hasan 1976. *Cohesion in English*. London : Longman.
- Halliday, M.A.K. & R. Hasan 1989. *Language, context and text: aspects of language in a social-semiotic perspective*. Oxford University Press.
- Halliday, M. A. K. & C. M. I. M. Matthiessen 2004. *An introduction to functional grammar*. London : Arnold.
- Hinds, J. 1987. Reader versus writer responsibility. A new typology. In : Connor, U. et R. Kaplan (ed.) *Writing across languages: analysis of L2 text*. Reading, MA : Addison-Wesley, 141-152.
- Ho Dac, L-M. 2007. *La position initiale dans l'organisation du discours : une exploration en corpus*. Thèse de doctorat. Université Toulouse-le Mirail.
- Ho-Dac, L.-M., M.-P. Jacques & J. Rebeyrolle 2004. Sur la fonction discursives des titres, In : Porhiel, S. & D. Klingler (ed.) *L'unité texte*. Pleyben : Perspectives, 125-152.

- Ho-Dac, L.-M. & M.-P. Péry-Woodley 2009. A data-driven study of temporal adverbials as discourse segmentation markers. *Discours* 4.
<http://discours.revues.org/index5952.html>.
- Ho-Dac, L.-M., C. Fabre, M.-P. Péry-Woodley & J. Rebeyrolle 2009. Corpus annotation of macro-discourse structures. *1st International conference on corpus linguistics*. 7-9 may 2009, Murcia, Spain.
- Honeste, M.L. & C. Froissart 2003. Blancs, casses, puces, tirets,... In : Combettes, B., C. Schnedecker, A. Theissen (ed.) *Ordre et distinction dans la langue et dans le discours*. Paris : Champion, 251-258.
- Hoey, M. 1983. *On the surface of discourse*. London : Allen & Unwin.
- Hoey, M. 1991. *Patterns of Lexis in Text*. Hong Kong : Oxford University Press.
- Holmes, R. 1997. Genre analysis and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes* 16(4), 321-337.
- Hunston, S. 1993. Disagreement in academic discourse. In : Baker, M., G. Francis, E. Tognini-Bonelli.(ed.) *Text and technology : in honour of John Sinclair*. Philadelphia : John Benjamins, 115-134.
- Hyland, K. 1998a Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics* 30(4), 437-455.
- Hyland, K. 1998b. *Hedging in scientific research articles*. Amsterdam : John Benjamins Publishing Company.
- Hyland, K. 1999. Disciplinary discourses: Writer stance in research articles. In : Candlin, C. & K. Hyland (ed.) *Writing: Texts, processes and practices*. London : Longman, 99-121.
- Hyland, K. 2000. *Disciplinary discourses: Social interactions in Academic Writing*. Singapore : Longman, Pearson Education.
- Hyland, K. 2005a. *Metadiscourse : exploring interaction in writing*. London : Continuum.
- Hyland, K. 2005b. Stance and engagement : a model of interaction in academic discourse. *Discourse Studies* 7(2), 173-192.
- Hyland, K. & P. Tse 2004. Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics* 25, 156-177.
- Jackiewicz, A. 2004. Les séries linéaires dans le discours : marques, opérations et structures sous-jacentes. *Journées ATALA*,
http://halshs.archives-ouvertes.fr/docs/00/06/25/13/PDF/sic_00001228.pdf.
 Consulté le 2 mars 2011.
- Jackiewicz, A. 2005. Les séries linéaires dans le discours. *Langue française* 148, 95-110.
- Jackiewicz, A. & J.-C. Minel 2003. L'identification des structures discursives engendrées par les cadres organisationnels. *Actes de la 10ème Conférence Traitement Automatique du Langage Naturel (TALN 2003)* Batz-sur-Mer, 155-164.
- Jacques, M.P. et J. Rebeyrolle 2006. Titres et structuration des documents. In : Enjalbert, P. (éd.) *Schédae* 1. Presses universitaires de Caen, 1-13.
- Jaubert, A. 2002. Corpus et champs disciplinaires. Le rôle du point de vue. *Corpus et recherches linguistiques* 1.

- Jaubert, A. (ed.) 2005. *Cohésion et cohérence. Etudes de linguistique textuelle*. Lyon : ENS éditions.
- Jenkins, J. 2003. *World Englishes : a resource book for students: abcd*. London : Routledge.
- Kalela, J. 2000. *Historiantutkimus ja historia*. Helsinki : Gaudeamus.
- Karttunen, L. 1969. Pronouns and variables. In : Binnic, R.I., A. Davison, G.M. Green, J.L. Morgan (ed.) *Papers from the fifth Regional Meeting of the Chicago Linguistic Society*. University of Chicago, 108-116.
- Kleiber, G. 1990. Article défini et démonstratif. Approche sémantique versus approche cognitive. In : Kleiber, G et J-E. Tyvaert (ed.) *Anaphore et ses domaines. Recherches linguistiques XIV. Etudes publiées par le centre d'analyse syntaxique*. Université de Metz. Faculté de Lettres et Sciences humaines. Paris : Klincksieck, 199-227.
- Kleiber, G. 1992. Anaphore associative et inférences. In : Tyvaert, J-E. (ed.) *Lexique et inférence(s)*. Paris : Klincksieck, 175-201.
- Kleiber, G., C. Schnedecker, L. Ujma 1994. L'anaphore associative, d'une conception à l'autre. In : Schnedecker, C., M. Charolles, G. Kleiber et J. David (ed.) *L'anaphore associative (Aspects linguistiques, psycholinguistiques et automatiques)*. Paris : Klincksieck, 5-66.
- Laippala, V. 2008. Nature des marqueurs des séries linéaires. In : *Collection des Congrès Mondiaux de Linguistique Française. CMLF 2008*. Paris : Institut de linguistique française.
- Leech, G. 1992. Corpora and theories of linguistic performance. In : Svartvik, J (ed.) *Directions in corpus linguistics*. Berlin, New York : Mouton de Gruyter, 105-122.
- Leech, G, T. McEnery & M. Wynne 1997. Further levels of annotation. In : Garside, R, G. Leech & A. McEnery (ed.) *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London : Addison Wesley, 85-101.
- Lesznyák, Á. 2004. *Communication in English as an international lingua franca : an exploratory case study*. Norderstedt : Books on Demand.
- Luc, C. 2001. Une typologie des énumérations basée sur les structures rhétoriques et architecturales du texte. Congrès TALN 2001 Recital 2001, Tours, 2-5 juillet 2001. Luc, C. J. Virbel 2001. Le modèle d'architecture textuelle : fondements et expérimentations. *Verbum* 23(1), 103-123.
- Luc, C., M. Mojahid, & J. Virbel 2002. Le modèle d'architecture de texte. *Actes de ISLSP*. Toulouse : Prescot, 41-50.
- Lundquist, L. 1980. *La cohérence textuelle: syntaxe, sémantique, pragmatique*. Copenhagen : Nyt Nordisk Forlag Arnold Busck.
- Maillard, M. 1974. Essai de typologie des substituts diaphoriques. *Langue française* 21, 55-71.
- Mann, W. C. & S. A. Thompson 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8 (3), 243-281.
- Mann, W. C. & S. A. Thompson 2001. Deux perspectives sur la théorie de la structure rhétorique (RST). *Verbum* XXIII 2001(1), 9-30.

- Mann, William C. and Maite Taboada 2007. *RST Web Site*, from <http://www.sfu.ca/rst>.
- Marcus, M.P., M. A. Marcinkiewicz & B. Santorini 2004. *Building a Large Annotated Corpus of English: The Penn Treebank*.
<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.14.9706>
- Markkanen, R. & H. Schröder (ed.) 1997. *Hedging and discourse : approaches to the analysis of a pragmatic phenomenon in academic texts*. Research in text theory, vol. 24. Berlin : Walter de Gruyter.
- Martin, J.R. 1992. *English text: system and structure*. Philadelphia (PA) : John Benjamins Publishing.
- Martin, J.R. 2003. Making History: Grammar for Interpretation. In : Martin, J.R. & R. Wodak (ed.) *Re/reading the past*. Amsterdam: Benjamins, 19-57.
- Martin, R. 1983. *La logique du sens*. Paris : PUF.
- Mauranen, A. 1993. *Cultural differences in academic rhetoric: a textlinguistic study*. Frankfurt am Main : Peter Lang.
- Mauranen, A. 2009. Spoken rhetoric: How do natives and non-natives fare? In : Suomela-Salmi, E. & F. Dervin (ed.) *Cross-linguistic and Cross-cultural Perspectives on Academic Discourse. Vol. 2*. Amsterdam : John Benjamins.
- McCabe, A. 1998. Sentences combined: text and discourse: In : Berko Gleason, J.& N. Ratner Bernstein (ed.) *Psycholinguistics*. New York : Harcourt Brace, 275-308.
- McEnery, T. & A. Wilson 1996. *Corpus linguistics*. Edinburgh University Press.
- Meyer, P.G. 1997. Hedging Strategies in Written Academic Discourse: Strengthening the argument by weakening the claim. In : Markkanen, R. & H. Schröder (ed.), 21-41.
- Milner, J-C. 1982. *Ordres et raisons de langue*. Paris : Editions du Seuil.
- Miltsakaki, E., R. Prasad, A. Joshi & B. Webber 2004. The Penn Discourse TreeBank. *Proceedings of the Language Resources and Evaluation Conference*. Lisbon, Portugal.
- Moeschler, J. & A. Reboul 1994. *Dictionnaire encyclopédique de pragmatique*. Paris : Editions du Seuil.
- Moirand, S. 1975. Le rôle anaphorique de la nominalisation dans la presse écrite. *Langue française* 28, 60-78.
- Monte, M. 2008. Usages littéraires de l’apostrophe : fonctions textuelles et pragmatiques et spécificités génériques. *Collection des Congrès Mondiaux de Linguistique Française. CMLF 2008*. Paris : Institut de linguistique française.
- Morris, J. & G. Hirst 2004. Non-Classical Lexical Semantic Relations. *Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics, Boston, Massachusetts*, 46-51.
- Mosegaard Hansen, M-B. 2005. A comparative study of the semantics of *enfin* and *finalemt*. *Journal of French Language Studies* 15, 153-171.
- Pacelli Pekba, T. 2003. Connecteurs et relations de discours : les cas de *quand*, *encore* et *aussi*. *Cahiers de linguistique française* 25, 237-256.
- Pascual, E. 1991. *Représentation de l’architecture textuelle et représentation de texte*. Thèse de doctorat. Université Paul Sabatier, Toulouse.

- Péry-Woodley, M-P. 2000. *Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle*. Université de Toulouse-LeMirail : ERSS.
- Péry-Woodley, M-P. 2005. Discours, corpus, traitement automatiques. In : Condamines, A. (éd.) *Sémantique et Corpus*. Paris : Hermès science publications, 177 – 210.
- Péry-Woodley, M.-P. & D. Scott 2006. Computational Approaches to Discourse and Document Processing. *T.A.L.* 47 (2), 7-19.
- Petőfi, J. 1971. 'Generativity' and text-grammar. *Folia linguistica V*, 278-309.
- Piérard, S. & Y. Bestgen 2006. Validation d'une méthodologie pour l'étude des marqueurs de la segmentation dans un grand corpus de textes. *TAL Volume 47. Discours et document : traitements automatiques*.
- Porhiel, S. 2007. Les structures énumératives à deux temps. *Revue Romane* 42(1), 103-135.
- Poudat, C. 2006. Étude contrastive de l'article scientifique de revue linguistique dans une perspective d'analyse des genres. *Texte!* XI(3-4).
- Prasad, T., N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, et B. Webber 2008. The Penn Discourse Treebank 2.0. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Rastier, F. 1989. *Sens et textualité*. Paris : Hachette.
- Reboul, A. 1994. Déixis et anaphore. In : Moeschler, J. et A. Reboul (ed.) *Dictionnaire encyclopédique de pragmatique*. Paris : Editions du Seuil, 349-372.
- Rebeyrolle, J., M-P. Jacques & M-P. Péry-Woodley 2009. Titres et intertitres dans l'organisation du discours. *Journal of French Language Studies* 19, 269-290
- Reppen, R., S. M. Fitzmaurice, D. Biber 2002. *Using corpora to explore linguistic variation*. Amsterdam, Philadelphia : J. Benjamins.
- Riegel, M., J-C. Pellat & R. Rioul 1994. *Grammaire méthodique du français*. Paris : Presses Universitaires de France.
- Rinck F. 2006. *L'article de recherche en Sciences du Langage et en Lettres, Figure de l'auteur et approche disciplinaire du genre*. Thèse de doctorat, Université de Grenoble III.
- Rinne, R., J. Kivirauma. & E. Lehtinen 2004. *Johdatus kasvatustieteisiin*. Helsinki : WSOY.
- Rossari, C., A. Beaulieu-Masson, C. Cojocariu, A. Razgouliaeva 2004. *Autour des connecteurs, Réflexions sur l'énonciation et la portée*. Berne : Peter Lang.
- Sanders, T. & W. Spooren 1999. Communicative intentions and coherence relations. In : Bublitz, W., U. Lenk & E. Ventola (ed.) *Coherence in spoken and written discourse: how to create it and how to describe it*. Amsterdam : John Benjamins, 235–250.
- Sanders, T.J.M & L.G.M Noordman 2000. The role of coherence relations and their linguistic markers in text processing. *Discourse Processes* 29, 37–60.
- Schneidecker, C. 1997. *Nom propre et chaînes de référence*. *Recherches Linguistiques* 21. Université de Metz.
- Schneidecker, C. 2000. Ordres des ordinaux pronominaux. *Travaux de linguistique* 41(2000), 7-34.

- Schnedecker, C. 2001. Adverbes ordinaux et introducteurs de cadre. Aspects linguistiques et cognitifs. *Linguisticae Investigationes* 24 :2(2001), 257-287.
- Schnedecker, C. 2006. *De l'un à l'autre et réciproquement. Aspects sémantiques, discursifs et cognitifs des pronoms anaphoriques corrélés l'un / l'autre et le premier / le second*. Bruxelles : Editions Duculot.
- Schneuwly, B., M-C. Rosat, J. Dolz 1989. Organismes textuels dans quatre types de textes écrits. Etude chez des élèves de dix, douze et quatorze ans. *Langue française* 81, 40-58.
- Sinclair, J. 1996. *Preliminary recommendations on Corpus Typology*.
<http://www.ilc.cnr.it/EAGLES/corpus/corpus.html>, visité le 1.3. 2011.
- Sinclair, J. 2004. *Trust the text*. London and New York: Routledge.
- Silver, M. & M. Bondi 2004. Weaving voices: A study of article openings in historical discourse.
 In : del Lungo Camiciotti, G. & E. Tognini Bonelli (ed.) *Academic discourse - new insights into evaluation*. Bern : Lang.
- Sullet-Nylander, F. 1998. *Le titre de presse. Analyse syntaxique, pragmatique et rhétorique*. Cahiers de la Recherche 8. Département de français et d'italien, Université de Stockholm.
- Suomela-Salmi, E. 1992. Some aspects of the pragmatic organization of academic discourse. In : Lindeberg, A.-C., N.E. Enkvist & K. Wikberg (ed.) 1992. *Nordic Research on Text and Discourse. NORDTEXT Symposium 1990*. Åbo Akademis Förlag, 251-262.
- Suomela-Salmi, E. 1993. Remarques sur l'organisation pragmatique du discours - une étude de cas. In : *Romanistica Turkuensis: Melanges d'etudes romanes offerts a Lauri Lindgren a l'occasion de son 60e anniversaire*. Turun yliopiston julkaisu, 241-257.
- Suomela-Salmi, E. 1997. *Les syntagmes nominaux (SN) dans les discours économiques français : repères textuels*. Thèse de doctorat. Turku.: Annales Universitatis Turkuensis, Ser.B – Tom 218.
- Swales, J. 1990. *Genre analysis : English in academic and research setting*. Cambridge University Press.
- Swales, J. 2004. *Research genres : explorations and applications*. New York : Cambridge University Press.
- Sözer, E. (ed.) 1984. *Text, Connexity, Text Coherence*. Hamburg : Helmut Buske Verlag.
- Taboada, M. 2006. Discourse Markers as Signals (or Not) of Rhetorical Relations. *Journal of Pragmatics* 38(4), 567-592.
- Taboada, M. 2009. Implicit and Explicit Coherence Relations. In: Renkema, J. (ed.) *Discourse, of Course*. Amsterdam/Philadelphia : John Benjamins, 127-140.
- Taboada, M. and W. C. Mann 2006. Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies* 8, 423-459.
- Tadros, A. 1985. *Prediction in text*. English Language Research, University of Birmingham.

- Tadros, A. 1994. Predictive categories in expository text. In : Coulhard, M. (éd.) *Advances in written text analysis*. London and New York : Routledge, 69-82.
- Tanskanen, S-K. 2006. *Collaborating towards coherence : lexical cohesion in English discourse*. Amsterdam : John Benjamins.
- Tesnière, L. 1959. *Éléments de syntaxe structurale*. Paris : Klincksieck.
- Teufel, S. 1998. Meta-discourse markers and problem-structuring in scientific articles. *Proceedings of the Workshop on Discourse Relations and Discourse Markers at the 17th International Conference on Computational Linguistics*, 43-49.
- Tomlin, R.S. (ed.) 1987. *Coherence and grounding in discourse : outcome of a symposium*. Eugene, Oregon, June 1984. Amsterdam, Philadelphia : J. Benjamins Pub. Co.
- Toulmin, S. 1958. *The Uses of Argument*. Cambridge : University Press.
- Turco, G. & D. Coltier 1988. Des agents doubles de l'organisation textuelle, les marqueurs d'intégration linéaire. *Pratiques* 57, 57-79.
- Van Dijk, T. 1972. *Some aspects of text grammars*. The Hague : Mouton.
- Van Peteghem, M. 2000. Les indéfinis corrélatifs *autre, même et tel*. In : Bosveld, L., M. Van Peteghem, D. Van de Velde (ed.) *De l'indétermination à la qualification. Les indéfinis*. Arras : Artois Presse Université, 117, 202.
- Vande Kopple, W. 1985. Some exploratory discourse on metadiscourse. *College Composition and Communication* 36 (1), 82-93.
- Vassileva, I. 2000. Who is the author ? *A Contrastive Analysis of Authorial Presence in French, English, German, Russian and Bulgarian Academic Discourse*. Asgard : Sankt Augustin.
- Vergez-Couret M. 2009. Le rôle de l'adverbe *notamment* dans la mise en oeuvre des relations de discours. *Journal of French Language Studies* 19(2). Cambridge University Press, 249-268.
- Winter, E. O 1982. *Towards a contextual grammar of English : the clause and its place in the definition of sentence*. London : Allen & Unwin.
- Virbel, J. 1989. The Contribution of Linguistic Knowledge to the Interpretation of Text Structures. In : André, J., V. Quint & R. K. Furuta (ed.) *Structured Documents*. Cambridge : CUP, 161-181.
- Virbel, J. 1999. *Structures textuelles - planches, fascicule 1 : Énumérations*. Rapport IRIT, Toulouse.
- Virbel, J. 2002. Éléments d'analyse du titre, Inscription Spatiale du Langage : structures et processus, Toulouse, Prescot, 123-132.
- Virtanen, T. 1992. *Discourse Functions of Adverbial Placement in English*. Åbo Akademis University Press.
- Virtanen, T. 2004. Point of departure. Cognitive aspects of sentence-initial adverbials. In : Virtanen, T. (ed.) *Trends in linguistics. Approaches to Congition through text and discourse*. Berlin : Mouton de Gruyter.
- Vold, E. 2008. *Modalité épistémique et discours scientifique. Une étude contrastive des modalisateurs épistémiques dans des articles de recherche français, norvégiens et anglais, en linguistique et médecine*. Thèse pour le degré de philosophiae doctor (PhD). Université de Bergen.

ANNEXE 1 LISTE DES ARTICLES DE RECHERCHE DU CORPUS

SOUS CORPUS LINGUISTIQUE CAHIERS DE GRAMMAIRE

Cahiers de Grammaire 26(2001), « Sémantique et Discours ».

Joan Busquets & Pascal Denis: *L'ellipse modale en français : le cas de devoir et pouvoir.*

Cahiers de Grammaire 28 (2003).

Michel Roché : *De la « bonne formation » des dérivés en –on.*

Cahiers de Grammaire 29 (2004).

Hervé Lieutard : *Spécificité morphologique du pluriel languedocien : la notion de « cheville »*

Cahiers de Grammaire 29 (2004).

Michèle Olivieri : *Paramètre du sujet nul et inversion du sujet dans les dialectes italiens et occitans*

Cahiers de Grammaire 29 (2004).

Alain Viaut : *Aréologie du lexique viticole d'oc en Bordelais et pluralité des facteurs déterminants : éléments pour une approche.*

Cahiers de Grammaire 30 (2006).

Anne Condamines : *Modes de construction du sens en corpus spécialisé.*

Cahiers de Grammaire 30 (2006).

Frédéric Lambert : *Pourquoi met-on une virgule devant et ?*

CORPUS

Corpus 1(2002).

Anna Jaubert : *Corpus et champs disciplinaires. Le rôle du point de vue.*

Corpus 1(2002).

Jean-Philippe Dalbera: *Le corpus entre données, analyse et théorie.*

Corpus 1(2002).

Damon Mayaffre: *Les corpus réflexifs : entre architextualité et hypertextualité.*

Corpus 5(2006)

Jean-Marie Viprey: *Un de ces [syntagmes] qui... (à propos de la locution un(e) de ces [...] qui).*

Corpus 6(2007)

Fanny Rinck et Agnès Tutin: *Annoter la polyphonie dans les textes : le cas des passages entre guillemets.*

Corpus 6(2007)

Michelle Lecolle : *Polysignifiante du toponyme, historicité du sens et interprétation en corpus. Le cas de Outreau.*

FRENCH LANGUAGE STUDIES

French Language Studies 16(2006).

Florence Lefeuve: *La structure en de quoi.*

French Language Studies 16(2006).

Marleen Van Peteghem: *Le datif en français: un cas structural.*

French Language Studies 18 (2008).

Cecile Fabre et Didier Bourigault: *Exploiter des corpus annotés syntaxiquement pour observer le continuum entre arguments et circonstants.*

French Language Studies 18 (2008).

Bernard Laks: *Pour une phonologie de corpus.*

LINGUISTICAE INVESTIGATIONES

Linguisticae Investigationes. XXVI, I (2003), 97-122.

Denis Vigier : *Les syntagmes prépositionnels en « en N » détachés en tête de phrase référant à des domaines d'activité.*

MARGES LINGUISTIQUES

Marges linguistiques 6(2003).

Françoise Gadet : *« Français populaire » : un classificateur déclassant ?*

Marges linguistiques 9(2005).

Jacques Guilhaumou : *Où va l'analyse de discours ? Autour de la notion de formation discursive*

Marges linguistiques 9(2005)..

Alain Rabatel : *La part de l'énonciateur dans la construction interactionnelle des points de vue.*

Marges linguistiques 11(2006).

Sylvain Auroux : *Les embarras de l'origine des langues.*

Marges linguistiques 11(2006).

Jean-Louis Dessalles : *Du protolangage au langage : modèle d'une transition.*

REVUE ROMANE

Revue Romane 39(2) 2004.

Jean-Michel Gouvard : *Proverbes et marquage évidentiel : le cas des énoncés métareprésentationnels.*

Revue Romane 39(2)2004.

Dulcie M. Engel & Nathalie Rossi-Gensane : *Sémantique et fréquence. Etude de c'est / il est dans un corpus journalistique.*

Revue Romane 40(1)2005.

Marion Carel : *La construction du sens des énoncés.*

Revue Romane 40(1)2005.

Anne Beaulieu-Masson : « De ce point de vue, c'est vrai que t'as carrément raison » : un cadre anaphorique de point de vue.

Revue Romane 40(1) 2005.

Corinne Rossari : *Le leurre des significations lexicales. L'exemple de la preuve.*

Revue Romane 41(1) 2006.

Sylvie Porhiel : *Non-phrases thématiques vs. non-phrases a-thématiques – Quels critères pour les distinguer dans les textes écrits.*

Revue Romane 41(2) 2006.

Karolien Rys : *L'exclamation : assertion non stabilisée ? Le cas des exclamatives à mot qu-.*

SOUS CORPUS HISTOIRE

ANNALES HISTORIQUES DE LA REVOLUTION FRANCAISE

Annales historiques de la Révolution française 337.

Annie Duprat : *Le regard d'un royaliste sur la Révolution : Jacques-Marie Boyer de Nîmes.*

Annales historiques de la Révolution française 338.

Jean-Luc Chappey : *Enjeux sociaux et politiques de la « vulgarisation scientifique » en Révolution (1780-1810).*

Annales historiques de la Révolution française 339.

Michel Biard : *Contrainte ou liberté économique ? Les représentants du peuple en mission et le ravitaillement de Paris en l'an III.*

Annales historiques de la Révolution française 339.

Bernard Gainot : *La Décade et la « colonisation nouvelle ».*

Annales historiques de la Révolution française 342.

Fabrice Bensimon : *L'écho de la Révolution française dans la Grande-Bretagne du XIXe siècle (1815-1870)*.

CAHIERS D'HISTOIRE

Cahiers d'histoire 2000(3).

Emmanuel Filhol : *L'image de l'autre au Moyen Age. La représentation du monde rural dans le Guide du pèlerin de saint-jacques de compostelle*.

Cahiers d'histoire 45(2) 2000.

Marie-Thérèse Lorcin : *Des paysans retrouvés : les vilains du XIIIe siècle d'après quelques textes en langue d'oïl*.

Cahiers d'histoire 46(3-4) 2001.

Jérôme Lafargue : *Landes de terre, d'eau et de bois. Pistes pour une sociologie historique d'un paysage entêté et entêtant*.

Cahiers d'histoire 47(1) 2002.

Dominique Dessertine et Bernard Maradan : *Patronages catholiques, patronages laïques entre les deux guerres : Les enjeux de la socialisation des enfants*.

Cahiers d'histoire 47(1) 2002.

Olivier Faure, Sorbonne : *Corps infirmes et villes malades*.

CAHIERS DE RECHERCHES MEDIEVALES

Cahiers de recherches médiévales 7(2000).

Patrick Arabeyre : *Aux racines de l'absolutisme : Grand Conseil et Parlement à la fin du Moyen Âge d'après le Tractatus celebris de auctoritate et preeminentia sacri magni concilii et parlamentorum regni Francie de Jean Montaigne (1512)*.

Cahiers de recherches médiévales 7(2000).

Éric Bournazel : *Réflexions sur l'institution du conseil aux premiers temps capétiens (XIIe-XIIIe siècles)*.

Cahiers de recherches médiévales (7)2000.

Gérard Giordanengo : *De l'usage du droit privé et du droit public au Moyen Âge*.

Cahiers de recherches médiévales 7(2000).

Anne Lefebvre-Teillard : *Causa natalium ad forum ecclesiasticum spectat : un pouvoir redoutable et redouté*.

Cahiers de recherches médiévales 7(2000).

Corinne Leveleux-Teixeira : *Dire et interdire. Le discours juridique entre omission et action. L'exemple du blasphème (XIIe-XVIe siècles)*.

Cahiers de recherches médiévales 7(2000).

Sophie Petit-Renaud : *Le roi, les légistes et le parlement de Paris aux XIVe et XVe siècles : contradictions dans la perception du pouvoir de « faire loy » ?*

Cahiers de recherches médiévales (7)2000.

Sophie Peralba : *Des coutumiers aux styles. L'isolement de la matière procédurale aux XIIIe et XIVe siècles.*

Cahiers de recherches médiévales (7)2000.

Franck Roumy : *L'ignorance du droit dans la doctrine civiliste des XIIe-XIIIe siècles.*

REVUE D'HISTOIRE DU XIXe SIECLE

Revue d'histoire du XIXe siècle 24(2002).

Manuel Charpy : *Formes et échelles du commerce d'occasion au XIXe siècle. L'exemple du vêtement à Paris.*

Revue d'histoire du XIXe siècle 29(2004).

François Jarrige : *Une « barricade de papiers » : le pétitionnement contre la restriction du suffrage universel masculin en mai 1850.*

Revue d'histoire du XIXe siècle 29(2004).

Stéphanie Sauget : *Surveiller les gares parisiennes au XIXe siècle : police et modernité.*

Revue d'histoire du XIXe siècle 32(2006).

Isabelle Surun : *L'exploration de l'Afrique au XIXe siècle : une histoire pré coloniale au regard des postcolonial studies.*

Revue d'histoire du XIXe siècle 33(2006).

Laurent Clavier : *« Quartier » et expériences politiques dans les faubourgs du nord-est parisien en 1848.*

Revue d'histoire du XIXe siècle 33(2006).

Aude Dontenville-Gerbaud : *Les discours, acte de fondation de la République : l'interaction orateurs/publics populaires. Eugène Spuller, Charles Floquet et Louis Blanc à Troyes en 1879.*

RHAF¹⁰⁶

RHAF57(4) 2004.

Vincent Duhaime : *« Les pères ont ici leur devoir » : le discours du mouvement familial québécois et la construction de la paternité dans l'après-guerre, 1945-1960.*

RHAF 57(4) 2004.

Cynthia S. Fish : *La puissance paternelle et les cas de garde d'enfants au Québec, 1866-1928.*

¹⁰⁶ Revue d'histoire de l'Amérique française.

RHAF 57(4) 2004.

Karine Hébert: *Carabines, poutchinettes co-eds ou freschettes sont-elles des étudiantes ? Les filles à l'Université McGill et à l'Université de Montréal (1900-1960).*

RHAF 57(4) 2004.

Christine Hudon & Louise Bienvenue : *Entre franche camaraderie et amours socratiques. L'espace trouble et ténu des amitiés masculines dans les collèges classiques (1870-1960).*

RHAF 57(4) 2004.

Isabelle Perreault: *Morale catholique et genre féminin : la sexualité dissertée dans les manuels de sexualité maritale au Québec, 1930-1960.*

RHAF 58(2) 2004.

Maude Roux-Pratte : *Les élites drummondvilloises et la crise des années 1930. Une étroite collaboration autour de l'assistance aux chômeurs.*

SOUS CORPUS EDUCATION

ACTES DU COLLOQUE « INNOVATIONS, USAGES, RESEAUX »

Actes du Colloque « Innovations, Usages, Réseaux »¹⁰⁷

Zohra Bouheraoua : *Les exercices interactifs : outil d'apprentissage ?*

Actes du Colloque « Innovations, Usages, Réseaux »

André Demailly : *La pédagogie en environnement numérique, entre ivresse des mots et rudesse des choses.*

Actes du Colloque « Innovations, Usages, Réseaux »

Emilie Marchal-Parmentelat : *TIC et acquisition de connaissances artistiques ou culturelles en milieu universitaire.*

Actes du Colloque « Innovations, Usages, Réseaux »

Hani Qotb : *Pour un scénario collaboratif d'enseignement du FOS à distance.*

RECHERCHES & EDUCATIONS

Recherches & éducations 5(2003).

Philippe Rousseaux : *Fonction du silence en pédagogie : une dimension performative.*

Recherches & éducations 5(2003).

Aline Voiry-Philippe : *La pédagogie du projet dans la classe d'intégration scolaire.*

Recherches & éducations 6(2004).

Franco Morandi : *Pragmatisme et pratiques en éducation Réflexion sur le principe d'action selon le pragmatisme de Pierce, James et Dewey.*

¹⁰⁷ <http://www.atilf.fr/atilf/evenement/Colloques/Innovations2006/Innovations2006.htm>

Recherches & éducations 6(2004).

Franck Vialle : *Une pragmatique de l'alternance. L'expérience des maisons familiales et rurales.*

Recherches & éducations 11(2006).

Céline Garcia : *Action sociale par le sport.*

REVUE DES SCIENCES DE L'ÉDUCATION

Revue des sciences de l'éducation XXX(3) 2004.

André Brassard, Martine Cloutier, Sylvie De Saedeleer, Lise Corriveau, Régent Fortin, Arthur Gélinas, Lorraine Savoie-Zajc : *Rapport à l'activité éducative et identité professionnelle chez les directeurs d'établissement des ordres d'enseignement préscolaire et primaire.*

Revue des sciences de l'éducation XXX(3) 2004.

Claire Beaumont, Égide Royer, François Bowen, Richard Bertrand : *L'adaptation psychosociale des élèves en trouble de comportement agissant comme médiateurs.*

Revue des sciences de l'éducation XXX(3) 2004.

Philippe Jonnaert, Johanne Barrette, Samira Boufrahi, Domenico Masciotra : *Contribution critique au développement des programmes d'études : compétences, constructivisme et interdisciplinarité.*

Revue des sciences de l'éducation XXX(3) 2004.

Benoît Galand, Pierre Philippot, Sylvie Petit, Michel Born, Geneviève Buidin : *Regards croisés sur les phénomènes de violence en milieu scolaire : élèves et équipes éducatives.*

Revue des sciences de l'éducation XXX(3) 2004.

Alain Legardez : *L'utilisation de l'analyse des représentations sociales dans une perspective didactique. L'exemple de questions économiques.*

Revue des sciences de l'éducation XXX(3) 2004.

Franck Martin, Agnès Morcillo, Jean-François Blin : *Le vécu émotionnel des enseignants confrontés à des perturbations scolaires.*

Revue des sciences de l'éducation XXX(3) 2004.

Stéphane Martineau, Annie Presseau : *Analyse exploratoire du discours sur la pratique chez des enseignants d'un CFER.*

REVUE DES SCIENCES DE L'ÉDUCATION DE MCGILL

Revue des sciences de l'éducation de McGill 40(2) 2005.

Yamina Bouchamma et Michael David : *Le leadership et stress en milieu minoritaire francophone.*

Revue des sciences de l'éducation de McGill 42(3) 2007.

Nancy Bouchard : *L'éducation éthique au sein du programme québécois de formation : une éducation transversale ?*

Revue des sciences de l'éducation de McGill 42(3) 2007.

Marc-André Ethier, David Lefrançois : *Eduquer à la citoyenneté délibérative : le défi d'exercer un leadership sociopolitique pour rectifier les inégalités à l'école en démocratie.*

Revue des sciences de l'éducation de McGill 42(3) 2007.

Claude Lessard, Louis Levasseur : *L'école publique généraliste est-elle en train de voir sa vocation transformée ?*

Revue des sciences de l'éducation de McGill 42(1) 2007.

Mirela Moldoveanu et Donatille Mujawa Mariya : *L'éducation multiculturelle dans la formation initiale des enseignants : des politiques aux pratiques.*

Revue des sciences de l'éducation de McGill 43(2) 2008.

Jean-Pierre Pelletier et France Jutras : *Les composantes de l'entraînement à l'improvisation actives dans la gestion des imprévus en salle de classe du niveau secondaire.*

Revue des sciences de l'éducation de McGill 43(2) 2008.

Thierry Piot : *La construction des compétences pour enseigner.*

SPIRALE

Spirale 2001(27).

Charles Gardou : *Enjeux et dérivés de la relation aux personnes en situation de handicap.*

Spirale 2001(27).

Anne-Marie Jovenet : *La différence à l'école : de la théorie à la pratique, ou de la pratique à la théorie ?*

Spirale 2005(36).

Christine Berzin : *Interactions entre pairs et apprentissages à l'école maternelle. Le cas du tutorat : Intérêts et limites.*

Spirale 2005(36).

Jérôme Clerc et Claire Leconte : *L'école maternelle a-t-elle vocation à enseigner le transfert de stratégies cognitives ?*

Spirale 2005(36).

Marc Loison : *Les apprentissages à l'école maternelle. Entre structuration du temps vécu et évocation du passé.*

Spirale 2005(36).

Jean-François Marcel : *Evaluer les apprentissages à l'école maternelle. Des pratiques enseignantes à l'interface du pédagogique et du social.*

Spirale 2005(36).

Jeanine Maso-Taulère : *La construction des inégalités : déjà en maternelle.*

ANNEXE 2 EXEMPLES DES SOUS CORPUS



Du protolangage au langage : modèle d'une transition¹

Par Jean-Louis Dessalles

École Nationale Supérieure des Télécommunications,
Paris, France

Mai 2006

1. Introduction : l'anatomie d'une transition²

Le comportement langagier humain repose sur plusieurs capacités cognitives propres à notre espèce. Les animaux n'utilisent pas de système de communication qui soit directement comparable au nôtre. Les tentatives pour leur enseigner différents aspects du langage humain (Premack & Premack, 1983 ; Savage-Rumbaugh & Lewin, 1994 ; Pepperberg, 1999) révèlent que certaines des capacités cognitives qui nous permettent de manier le langage semblent ne pas être présentes chez les animaux testés. L'une de ces capacités, maintes fois soulignée, concerne l'usage d'une syntaxe. Certes, les oiseaux, par exemple les rossignols, sont connus pour produire des chants présentant plusieurs niveaux de combinatoire (Hauser, 1996, p. 286). Cependant, rien n'indique que les différentes structures ainsi produites servent à exprimer des significations différenciées. Les primates entraînés produisent parfois spontanément des groupements de deux mots sémantiquement reliés au contenu concret exprimé, avec une tendance à observer un ordre systématique (Savage-Rumbaugh & Lewin, 1994, p. 161). Ce type de performance, qui reste éloignée de ce que les êtres humains produisent avec le langage, amène certains auteurs à voir là l'expression d'une capacité protolangagière (Bickerton, 1990, 1995).

Le protolangage a été défini, par extrapolation à partir de l'observation des pidgins, comme une forme d'expression consistant à assembler quelques mots dans un énoncé court, sans support grammatical : absence de mots grammaticaux, absence de dépendance à distance dans la phrase, absence de marques de flexion, absence d'ordre défini (Bickerton, 1990). Le protolangage, vers lequel nous régressons spontanément dès que nous sommes en difficulté linguistique, est ainsi présenté par Bickerton comme un précurseur du langage, une sorte de capacité intermédiaire entre la communication spontanée des primates et le langage proprement dit qui est universellement pratiqué dans notre espèce. Une expression moderne du protolangage peut s'observer sur les moteurs de recherches utilisés sur la toile mondiale. Bien que ces moteurs de recherche permettent de s'exprimer à l'aide d'expressions booléennes et reconnaissent des groupes nominaux, les utilisateurs ont tendance à s'exprimer par des suites de mots sans liens grammaticaux entre eux (Guichard, 2002).

Quelle que soit l'attitude que l'on adopte vis-à-vis de la réalité du protolangage, la question de la transition menant, au cours de l'évolution, à la forme de langage que nous connaissons, ne peut être ignorée. Certaines de nos aptitudes cognitives nous permettent de manier des constructions langagières signifiantes et de réaliser des opérations non triviales sur ces structures, comme l'interrogation, la mise au passif, le changement modal ou temporel, etc. Sans prendre ici parti sur la question de savoir si ces dispositions cognitives sont spécifiques au langage ou correspondent à des fonctions plus générales (Piattelli-Palmarini, 1979), nous cherchons dans la structure du langage des indices susceptibles de nous renseigner sur leur ordre d'apparition. Pour tenter de reconstituer l'anatomie de la transition vers le langage, nous nous intéressons particulièrement à l'interface entre la syntaxe et la sémantique, en adoptant

¹ Je remercie Laleh Ghadakpour qui a bien voulu commenter une version antérieure de cet article. Je voudrais exprimer ma gratitude aux personnes qui sont à l'origine de l'initiative « Origine de l'Homme, du Langage et des Langues » du CNRS, qui ont permis à la problématique des origines du langage de retrouver une juste place dans le débat scientifique français.

² Ce travail a été effectué dans le cadre de l'initiative OHLL (Origine de l'Homme, du Langage et des Langues) du CNRS.

le point de vue de la modélisation. En considérant la complexité relative des opérations de synchronisation entre la structure syntaxique et le sens, nous en viendrons à proposer un ordre dans l'apparition de certaines capacités qui sous-tendent la maîtrise du langage.

Dans la suite, nous commençons par poser le problème en reprenant certaines caractéristiques formelles de la syntaxe qui sont généralement invoquées pour démontrer l'originalité de notre mode de communication, notamment la récursivité et les systèmes de flexion. Nous tentons de dégager un modèle simplifié du rôle de ces caractéristiques dans l'expression du sens. Nous tirerons argument de la structure de ce modèle simplifié pour proposer une succession plausible dans l'établissement des capacités cognitives qui sous-tendent ces aspects du langage. Nous concluons en replaçant ces propositions dans le cadre des recherches actuelles sur l'origine de la capacité langagière.

2. Quelques aspects notables de la capacité syntaxique

La variété des langues est grande, mais elle n'est pas infinie. En particulier, toutes les langues spontanées du monde, à l'exception des pidgins récents, possèdent une grammaire qui impose la flexion de certains mots et leur agencement dans la phrase. Aucun peuple au monde ne s'exprime spontanément en groupant, sans ordre particulier, des mots invariables¹.

Certes, les règles syntaxiques diffèrent largement d'une langue à l'autre. Certaines langues comme le latin ou le dyirbal semblent très permissives en ce qui concerne l'ordre des mots dans les phrases simples (Dixon, 1972), alors que d'autres comme le français ou l'anglais sont nettement moins tolérantes sur cet aspect de la grammaire. Certaines langues, comme le lakhsota, mettent en œuvre un système morphologique complexe affectant principalement le verbe, alors que des langues comme l'anglais ou le chinois sont très économes en flexions (Van Valin & LaPolla, 1997). Malgré ces différences flagrantes, il est possible de faire quelques observations générales sur la structure des langues qui suffiront pour la suite de notre propos.

La première observation porte sur l'existence de dépendances entre les mots. Dans un énoncé protolangagier du type *maison-voisin-feu*, chacun des trois mots entretient un rapport sémantique direct avec la scène perçue décrite. La communication peut fonctionner sans que les mots d'un tel énoncé ne dépendent les uns des autres. Les langues humaines introduisent plusieurs formes de dépendance, qui sont clairement indiquées par la grammaire. Dans la phrase *la maison du voisin est en feu*, les mots entretiennent des relations formelles, indépendamment de leur signification. Ces relations, par exemple, rapprochent la préposition *en* et le nom *feu*, ou les noms *maison* et *voisin*, alors qu'elles laissent *voisin* et *feu* sans lien direct. Ces dépendances priment souvent sur le sens : la grammaire d'une expression comme *la souris mange le chat* impose une relation sémantique qui va à l'encontre de ce que suggère la simple association entre les sens évoqués par les mots *souris*, *chat*, *manger*.

La deuxième observation concerne le fait que les dépendances syntaxiques sont asymétriques. Dans une expression latine comme *domus domini*, le génitif du mot *dominus* indique sans ambiguïté que ce mot est un satellite du mot *domus*. En français, les deux expressions *la maison du maître* et *le maître de la maison* se distinguent par la position relative des mots par rapport à la préposition *de*. Cette dépendance asymétrique s'interprète de diverses manières, comme résultant d'une opération de fusion syntaxique (Chomsky, 1995) ou d'une dépendance sémantique (Van Valin & LaPolla, 1997). Du fait de la présence de dépendances asymétriques, les énoncés langagiers comportent des syntagmes, dont la tête domine, au sens structurel, l'élément dépendant.

La troisième observation concerne la structure « moléculaire » des énoncés langagiers. Comme les relations de dépendance peuvent s'enchaîner, les syntagmes se retrouvent emboîtés les uns dans les autres comme des poupées russes. Dans une phrase simple comme *J'ai moyennement aimé l'acoustique du concert auquel j'ai assisté hier*, on trouve maints emboîtements : le syntagme verbal centré sur le verbe *assister* dépend (indirectement) du syntagme nominal centré sur le mot *concert*, qui fait partie intégrante du syntagme nominal centré sur

¹ Les pidgins de création récente constituent clairement une exception. Toutefois, ils sont utilisés comme langue principale seulement dans des conditions sociales anormales qui imposent un brassage d'individus adultes de communautés linguistiques différentes (Bickerton 1990). L'histoire de la formation des créoles et l'exemple des enfants sourds du Nicaragua (Kegl, Senghas & Coppola, 1999) indiquent que les pidgins sont une forme de communication instable qui disparaît en tant que moyen de communication principal dès que la structure sociale permet aux enfants de moins de sept ans de former des communautés.

acoustique, lui-même inclus dans le syntagme verbal dont la tête est le verbe *aimer*. Les différentes théories syntaxiques peuvent différer par le choix précis des types de syntagmes, mais toutes admettent le principe de l'imbrication. Il en résulte une structure qui peut être qualifiée d'arborescente (si l'on représente le graphe des dépendances), de moléculaire (par analogie avec la chimie des macro-molécules, qui sont constituées d'autres molécules), de fractale (par référence aux structures vivantes ou physiques invariants par changement d'échelle) ou de récursive. Ce dernier qualificatif fait référence à une propriété de la procédure qui permet de lire ou d'engendrer la structure syntaxique. Les procédures récursives ont la propriété d'appeler une copie d'elles-mêmes au cours de leur exécution. C'est donc la même procédure qui permet d'analyser le syntagme nominal *acoustique du concert auquel j'ai assisté hier* et le syntagme nominal inclus *concert auquel j'ai assisté hier*. D'un point de vue technique, le fait remarquable est que la récursivité à l'œuvre dans le langage humain est une récursivité centrale, c'est-à-dire une récursivité qu'il est impossible de ramener à une simple itération répétitive.

La dernière observation que nous faisons ici à propos des dépendances syntaxiques est qu'elles s'exercent à distance. Dans *l'acoustique du concert auquel j'ai eu la chance d'assister hier*, on comprend que le nom *concert* sert de complément à la fois au nom *acoustique* et au verbe *assister*. Pourtant, *concert* et *assister* sont séparés de sept mots. La distance de ce type de dépendance peut être arbitrairement grande (*le concert auquel la fille du copain que j'ai connu quand j'étais à la fac a eu la chance d'assister avait une mauvaise acoustique*). L'interprétation de ces dépendances à distance varie d'une théorie à l'autre, selon que l'on parle de mouvement (le complément de *assister* migrant en tête de clause) ou de duplication silencieuse.

Pour plusieurs auteurs, certaines de ces propriétés syntaxiques sont indissociables. L'existence de dépendances asymétriques, l'emboîtement des syntagmes et les différentes formes de dépendance à distance iraient de pair, elles constitueraient une sorte de paquet cadeau que la Nature nous aurait offert en nous dotant de la capacité syntaxique. L'ensemble de ces propriétés résulterait de la seule opération de fusion (Berwick, 1998) ou, de manière équivalente, de l'apparition fortuite dans notre espèce d'un accès à la récursivité (Hauser, Chomsky & Fitch, 2002). De fait, il est difficile d'imaginer des intermédiaires dans l'accès à la récursivité. Dans ces conditions, la transition du protolangage (ou du non-langage) vers le langage ne peut être que soudaine. Certains ont pu invoquer l'intervention d'une improbable macro-mutation (Bickerton, 1990). Sa cause aurait été purement fortuite, et ne devrait rien à un quelconque accroissement du pouvoir communicationnel (Chomsky, 1975, p. 75 ; Piattelli-Palmarini, 1989 ; Lightfoot, 2000). La nouvelle capacité récursive, peut-être à la faveur d'un décloisonnement fonctionnel (Hauser, Chomsky & Fitch, 2002), aurait entraîné l'ensemble des propriétés de la syntaxe qu'il nous est donné d'observer dans les différentes langues, avec parmi elles les propriétés de dépendance que nous avons soulignées ici.

La question de la transition vers la capacité syntaxique, s'il s'est agi d'un changement abrupt et fortuit, perd beaucoup de son intérêt scientifique, car elle n'a pu produire qu'un dispositif peu adapté aux utilisations que les êtres humains en font.

Language design as such appears to be in many respects « dysfunctional, » yielding properties that are not well adapted to the function language is called upon to perform (Chomsky, 1995, p. 162).

[...] full-blown evolutionary novelty can also suddenly arise, so to speak, for no reason, because novelty caused by sheer proximity between genes is not governed by function and it, therefore, eludes strict adaptationism (Piattelli-Palmarini, 1989, p. 8).

Le point de vue selon lequel l'évolution pourrait ainsi produire des structures ou des compétences complexes peu fonctionnelles apparaît comme non fondé sur le plan biologique (Desalles, 1996), et ne saurait s'appliquer au langage (Pinker & Bloom, 1990 ; Pinker & Jackendoff, 2005). La théorie de la sélection naturelle et l'étude cognitive du langage suggèrent toutes deux que la syntaxe remplit une fonction précise, qui est celle de l'expression du sens sur le canal de l'expression vocal, et que sa structure est la conséquence évolutive de cette fonction. Il est donc légitime d'étudier la fonction de la syntaxe pour tenter de proposer une reconstitution des grandes étapes par lesquelles est passée l'évolution du langage.

Dans la suite, nous allons nous attacher à remettre en question ces deux propriétés supposées de la transition du protolangage au langage que seraient son caractère abrupt et son caractère fortuit. En observant, du point de vue de la modélisation, l'interface entre la syntaxe

L'ellipse modale en français : le cas de *devoir* et *pouvoir*^{*}

Joan Busquets & Pascal Denis**

*Le but de cet article est d'entamer une recherche sur le statut de quelques constructions elliptiques modales en français. En particulier, celles associées aux verbes *devoir* et *pouvoir*. Nous tenterons de démontrer que ces constructions partagent un ensemble non trivial de propriétés qui sont étroitement liées à l'ellipse en anglais (VPE), et non avec certains processus anaphoriques comme l'anaphore de complément nul (NCA). En outre, nous examinerons cette hypothèse évaluant les possibilités des constructions elliptiques modales et certaines distinctions sémantiques telles que l'interprétation épistémique VS déontique ou l'ambiguïté de *dicto* vs *de re*. L'analyse de l'ellipse modale dans la perspective d'une théorie de la structure du discours fera l'objet de nos futurs travaux.*

*The aim of this paper is to begin investigating the status of some modal elliptical constructions in French. In particular, those associated with *devoir* and *pouvoir*. It will be argued that these constructions share a set of non trivial properties closely related to English verb-phrase ellipsis (VPE), and not with anaphoric processes like null complement anaphora (NCA). We will furthermore test this assumption BY evaluating the possibilities of modal elliptical constructions and semantic distinctions between epistemic vs deontic or *de dicto* vs *de re* interpretations. It is left for future research to place the approach in the perspective of a theory of discourse structure.*

* Nous tenons à remercier A. Le Draoulec, F. Lambert et H. Portine pour leurs suggestions et commentaires.

** Université Bordeaux-3, University of Texas at Austin.

1. Introduction

Une caractéristique importante des langues naturelles est que la relation allant de la forme au contenu est rarement univoque (ou, plus précisément, injective). Ainsi, à une même signification correspond généralement une classe plus ou moins large de structures syntaxiques, équivalentes en termes vériconditionnels. Nous nous intéresserons ici à un cas particulier de cette relation non-univoque entre forme et contenu ; à savoir : le cas de l'*ellipse*, c'est-à-dire, le cas où différentes structures sémantiquement équivalentes sont construites à partir des mêmes items lexicaux et ne diffèrent qu'en termes de « longueur ».

Parmi les différents phénomènes elliptiques répertoriés par les linguistes, c'est l'ellipse verbale (ou VPE pour *Verb Phrase Ellipsis*), illustrée en anglais par la paire minimale (1)-(2), qui a sans nul doute suscité la littérature la plus abondante (*inter alia*, Hankamer et Sag, 1976 ; Napoli, 1985 ; Lobeck, 1995 ; Johnson, 1997) :

- 1) *Jerry isn't going/ hasn't gone/ (didn't/ won't/ couldn't) go to the party, but Kramer is going/ has/ (did/will/could) go to the party.*
- 2) *Jerry isn't going/ hasn't gone/ (didn't/ won't/ couldn't) go to the party, but Kramer is/ has/ (did/will/could) Ø.*

‘Jerry ne vient pas/n’est pas venu/ne viendra pas/ne peut pas venir à la soirée, bien que Kramer, lui, vienne/soit venu/vienne/puisse venir à la soirée’

Dans les exemples ci-dessus, les phrases présentées en (2) sont des variantes tronquées des phrases (1) ; elles diffèrent de ces dernières en ce qu’elles font l’économie du syntagme verbal (SV) *go to the party* (c.-à-d., en termes sémantiques, de la *propriété* [[venir à la soirée]]) dans la seconde proposition¹.

56

¹ Les autres phénomènes généralement repris sous l’étiquette d’ellipse sont le *Gapping* (i), le *Pseudo-Gapping* (ii) et le *Sluicing* (iii) :

(i) *Jerry bought a Saab and Joe a Porsche.*

‘Jerry a acheté une Saab et Joe une Porsche’

(ii) *Joe hasn't bought a Saab, but he has a Porsche.*

‘Jerry n’a pas acheté une Saab, mais il a acheté une Porsche’

(iii) *Harry is supposed to meet Sally, but he doesn't know when.*

‘Harry est censé rencontrer Sally, mais il ne sait pas quand’.

Ces constructions diffèrent en termes du matériel qui est visé par l’ellipse. En (i), l’ellipse vise le verbe fini et non le SV tout entier, laissant comme « résidus » le SN sujet et le SN complément. L’ellipse présentée en (ii) est encore plus minimale, puisqu’elle ne porte ici que sur le participe passé. Enfin, en (iii), c’est la proposition entière qui a été éliminée ; ce qui explique la présence du pronom interrogatif *when* en fin de phrase. Notons qu’en français, seuls le *Gapping* et le

Définir l'ellipse comme une relation entre deux phrases vériditionnellement équivalentes qui diffèrent en termes de longueur, comme nous l'avons fait jusqu'à présent, reste très approximatif. Cette définition ne dit rien, notamment pour (2), des principes généraux qui gouvernent l'éllision du SV. Dans de nombreux travaux issus de la grammaire générative (voir, par exemple, Sag 1976 ; Hankamer et Sag 1976 ; et nombre de travaux ultérieurs tels que Fox 2000 ; Heim et Kratzer 2000), on considère que la VPE est sujette à la condition suivante : le SV de la proposition « cible » (c.-à-d. le site elliptique) et le SV de la proposition « source » (c.-à-d. la proposition antécédente) sont *identiques* (*modulo* un renommage des variables liées) en termes de représentations LF (pour *Logical Form*). Cette condition est illustrée ci-dessous pour la phrase *Jerry didn't go to the party, although Kramer did* :

- 3) Jerry PAST not λx [_{VP} x go to the party] although Kramer PAST λy [_{VP} ~~y go to the party~~]

Ce qui est présupposé par une telle condition, c'est que les deux propositions sont identiques au niveau de LF et ne diffèrent qu'en termes de leur réalisation phonologique/phonétique ; ou encore que le second SV ne disparaît que lors du « mappage » de la S-structure vers PF (pour *Phonological Form*). Vue sous cet angle, l'ellipse verbale correspond donc à un cas extrême de dégradation phonétique. Notons que cette analyse *dérivée* de l'ellipse verbale est loin de faire l'unanimité. Ainsi, on notera en particulier que d'autres travaux (par exemple, Klein 1987 ; Asher 1993 ; Hardt 1992) proposent un traitement purement *sémantique* de l'ellipse verbale, qui diffère de l'approche dérivée sur deux points importants : (i) le site elliptique contient, non pas une représentation syntaxique, mais une forme pronominale (ou *pro-forme*) nulle, et (ii) la condition d'identité est définie en termes de la théorie des modèles et se traduit par un mécanisme de récupération de la propriété présente dans la proposition source.

Comme nous l'avons noté, le phénomène de l'ellipse verbale a fait l'objet de nombreuses études pour l'anglais. Malheureusement, ce phénomène a été largement ignoré pour d'autres langues telles que le français. A priori, une excellente raison pour justifier cette « négligence » consisterait à dire que le français n'autorise pas ce type particulier de constructions elliptiques, alors qu'il en permet d'autres comme le Gapping et le Sluicing. Ce point de vue est notamment défendu par Lobeck (1995) et semble conforté par une série de faits inhérents au français. D'une part, on a

Sluicing sont admis (voir cependant la section 3.2 pour de potentiels exemples de *Pseudo-Gapping* en français).

ENJEUX ET DÉRIVES DE LA RELATION AUX PERSONNES EN SITUATION DE HANDICAP¹

Résumé : L'auteur analyse comment certaines de nos attitudes fabriquent du handicap et développe trois conditions qui, si elles étaient réalisées, permettraient une véritable rencontre avec ceux qui vivent sur un autre mode l'humaine condition : reconnaître la différence sans la radicaliser, prendre toute la mesure de l'influence de notre regard sur la construction identitaire de l'autre, évacuer préjugés et stéréotypes. A ce compte seulement, pouvons-nous reconnaître derrière le « handicapé », une personne au sens plein du terme.

Mots-clés : catégorisation, image de soi, préjugés, reconnaissance, regard, situation handicapante, stigmaté.

« Nous ne serons jamais une seule momie sous l'antique désert et les palmiers heureux », telle était la manière de Mallarmé de rappeler que nous n'habitons pas seuls le monde. Ces propos, qui peuvent passer pour un truisme ne méritant pas de retenir l'analyse, sont susceptibles d'ouvrir sur une problématique qui s'étend jusqu'aux limites de la réflexion. En effet, l'ambiguïté de la présence de l'autre nous confronte à un double danger : soit être objet pour lui, soit le traiter comme objet. S'il est vrai, selon les mots de Sartre, que l'on ne constitue pas autrui mais qu'on le rencontre, notre relation à lui peut en revanche, s'instaurer de bien des manières : d'égal à égal, de pourvoyeur à dépendant, de supérieur à subordonné... Ainsi, la confirmation mutuelle des uns par les autres revêt-elle les formes les plus variées, sinon les plus contradictoires.

Pour qui observe le spectacle social et ne reste pas aveugle à ses propres réactions, il est aisé de constater que, feignant de sortir de nous-mêmes et croyant traiter des autres et de leurs difficultés, nous nous préoccupons d'abord, en réalité, des problèmes qui sont les nôtres. Là où l'hostilité manifeste s'estompe, l'indifférence larvée s'étale. La communication est paralysée par le besoin de soumettre, d'ordonner systématiquement autour de soi et d'étalonner selon son propre modèle : chaque partenaire se trouve alors dominant ou dominé, conforme ou étrange, accepté ou rejeté... Le monde des autres ne s'apparente en rien, chacun le sait, à un jardin de délices, mais provoque constamment à la lutte, à l'adaptation, au dépassement. Il

¹ Cette contribution correspond à un extrait remanié de l'ouvrage *Handicaps, handicapés : le regard interrogé*, sous la direction de Charles Gardou, diffusé en 1991 par les Éditions Érès. Nous la publions ici avec leur aimable autorisation.

réintroduit sans cesse le risque et la souffrance. Cela s'avère particulièrement vrai dans la relation à ceux qui portent la double blessure du handicap : celle qui a endommagé leur corps, leur sens ou leur esprit, et l'autre, bien plus intolérable, que constitue le rejet dont ils sont victimes. Or, ce n'est qu'en prenant conscience, sans concession ni esquivance, de nos attitudes face à eux que nous leur permettrons d'accéder à la place entière qui leur revient : à la pleine citoyenneté, dit-on. Dans cette perspective, trois conditions me paraissent requises : reconnaître d'abord leur différence sans la radicaliser ; prendre conscience de l'influence de notre regard sur leur construction identitaire ; tenter enfin d'évacuer les préjugés et stéréotypes entravant leur reconnaissance.

1 – RECONNAÎTRE LEUR DIFFÉRENCE SANS LA RADICALISER

La relation à nos semblables blessés se trouve prise dans une tension entre postulat d'égalité et reconnaissance de leur singularité. D'un côté, la négation de la différence au nom de l'égalité, qui dissimule la propension à « camoufler l'altérité au fur et à mesure qu'elle surgit dans le champ de l'expérience »², finit par conduire au rejet par indifférenciation et à des formes d'égalitarisme béat. Ici, la différence se dilue dans l'identique. À l'inverse, la revendication inconditionnelle du droit à la différence aboutit à la radicalisation des particularités : à force de souligner ce qui sépare, on l'hypertrophie et on le pérennise. Là, en assignant l'autre à une identité simplificatrice, on ne fait que traduire un désir de ne repérer que les seules dissimilarités afin de maintenir intactes les distances interpersonnelles³.

La focalisation sur la différence constitue un des moteurs des pratiques d'exclusion et justifie la mise à l'écart des « non-conformes ». À ce propos, Michel Maffesoli montre que, dans notre société où le travail, la performance, la réussite individuelle, la rentabilité et la productivité s'imposent comme valeurs dominantes, toute personne, qui ne peut pas ou ne veut pas obéir à ces injonctions, est définie par son incapacité à se plier à ce devoir-être et se trouve absorbée dans cette caractéristique⁴. Ainsi voit-on se développer, de manière insidieuse, un intégrisme de la différence. La ségrégation au nom du droit à la singularité, l'exclusion au nom de la tolérance : c'est en débusquant cette contradiction, au cœur de l'idéologie actuelle, que l'on peut repenser les bases mêmes de nos discours et pratiques.

Différencier amène, de manière subreptice, à étiqueter, puis à cantonner dans un espace séparé les victimes du handicap et, finalement, à légitimer leur marginalisation, voire leur enfermement. La tendance à les saisir d'abord par leur caractéristique la plus perceptible socialement et la polarisation sur leur différence « négative »

² Certeau (Michel de), *L'étranger ou l'union dans la différence*, Paris, Desclée de Brouwer, 1969, p. 182.

³ C'est ce que met en lumière, dans le champ de l'interculturel, les travaux de Martine Abdallah-Preteuille. On se reportera avec profit à son ouvrage *Vers une pédagogie interculturelle*, Paris, Anthropos, 1996.

⁴ Maffesoli (Michel), « La réconciliation des différences », *Prospective et Santé*, n°26, 1983, p. 48.

LA RELATION AUX PERSONNES EN SITUATION DE HANDICAP

les marquent, de manière indélébile, du sceau du particularisme. Et même les attitudes les plus généreuses restent traversées, d'une part, par la dérive classificatrice qui les stigmatise et, d'autre part, par la dérive ségrépatrice qui propose, pour chaque catégorie, des structures particulières. Le marquage non seulement les rassemble sous une étiquette commune, mais les voue à un même destin, de sorte qu'il les constitue en groupe d'appartenance, au sens où Kurt Lewin disait que celui-ci réunit tous ceux qui partagent une certaine communauté de sort. Toutefois, dans la mesure où il n'a d'autre fondement que la destinée, sur laquelle ils n'ont aucune prise et à laquelle ils se trouvent subordonnés, leur groupement existe essentiellement comme lieu d'exclusion. La fonction du terme « handicapé » n'est pas de dire en quoi un individu est différent mais seulement qu'il est porteur d'une différence qui l'entrave. Ce mot ne parle pas de la personne, il la désigne et la repère. Et cette division des êtres humains, selon ce qui les distingue, aboutit à l'impasse joliment formulée par Geza Roheim : « Vous êtes complètement différent de moi, mais je vous pardonne »⁵.

Le handicap est un mixte de l'unité et du contradictoire : sa présence est le signe patent d'un va-et-vient permanent entre la norme et le hors-norme, l'équilibre et le déséquilibre, la santé et la maladie. Il bouscule nos références mimétiques, nos visions normosées et nos certitudes préétablies. Celui qui en est atteint représente l'autre par excellence, l'autre trop radicalement différent : « il est cette déchirure de notre être qui ouvre sur son inachèvement, son incomplétude, sa précarité »⁶. En participant au grand jeu du monde, il marque les limites de notre finitude.

Face à lui, la difficulté demeure de canaliser nos réactions archaïques, de rompre l'imperméabilité qui nous éloigne, de dépasser notre inclination à repousser la différence ou à l'utiliser en termes de séparation et de faire le pari que nous pouvons communiquer malgré la diversité des « allures de la vie »⁷. Erick H. Erickson voyait à raison dans le courage de la différence un signe de plénitude à la fois chez les individus et dans les civilisations⁸. Certes, il est bien plus aisé de classer « handicapé », de marginaliser, de faire disparaître en coulisses que d'admettre qu'il existe divers modes d'être ensemble et de promouvoir un regard empêchant la réification de l'autre.

Les victimes du hasard nous donnent à comprendre à la fois que la bigarrure, la dissonance, la rupture sont immanentes à la nature humaine, à l'« ordre » social et que les hommes peuvent vivre leur confrontation sans que l'existence de normes entraîne une exigence d'uniformité ou fasse réputer déviant ce qui n'est que différent. C'est seulement en entrant en relation avec eux que je reconnais leur qualité de sujet, comparable à celui que je suis moi-même. Aussi paraît-il indispensable que

⁵ Roheim (Geza), *Psychanalyse et anthropologie*, Paris, Gallimard, 1967, p. 46.

⁶ D'après Sticker (Henri-Jacques), *Corps infirmes et sociétés*, Paris, Aubier-Montaigne, 1982, p. 18, Nouvelle édition, Paris, Dunod, 1997.

⁷ Canquillien (Georges), *Le normal et le pathologique*, Paris, PUF, 1991 (1ère édition, 1966).

⁸ Erickson (Erik H.), « Identité individuelle et sentiment de déracinement à notre époque », *Éthique et psychanalyse*, Paris, Flammarion, 1971.

Les cercles littéraires entre pairs en première secondaire: étude des relations entre les modalités de lecture et de collaboration

Manon Hébert, professeure
Université du Québec en Outaouais

Résumé – Cette étude décrit la dynamique des interactions de nature sociale, métacognitive et cognitive dans les cercles littéraires en première secondaire ainsi que les différences entre les groupes « fort » et « régulier ». L'analyse des transcriptions de huit discussions révèle, sur le plan cognitif (ou du contenu discuté), que la compréhension littérale est le mode de lecture le plus employé, même si les modes esthétique et analytique entraînent plutôt des réflexions plus élaborées. Sur le plan des interactions sociales, il appert que près de la moitié des échanges entre acteurs ont pour but de gérer la tâche et de rétroagir aux propos des pairs. Les groupes « forts » auraient tendance à adopter une modalité plus divergente de coélaboration du sens.

Introduction

Les approches collaboratives (ou enseignement-apprentissage par les pairs), qui tentent d'intégrer les multiples niveaux de fonctionnement de l'élève (cognitif, affectif, métacognitif et social), auraient des effets nettement supérieurs aux approches individuelles traditionnelles pour l'apprentissage des mathématiques et de la langue maternelle (Peklaj et Vodopivec, 1999; Fuchs, Fuchs et Kazdan, 1999). Il ressort, entre autres des travaux de Vygotski (1985), que les processus mentaux de haut niveau, sur lesquels repose en grande partie la lecture littéraire, s'acquièrent par imitation et guidage. Ainsi, les tâches encourageant le dialogue et la collaboration (comme le journal dialogué et les cercles littéraires), en permettant aux élèves de générer, de comparer et de confronter leurs propres opinions et questions, favoriseraient la compréhension, l'interprétation et l'appréciation des textes littéraires en classe (Almasi, 1995; Pressley, Brown, El-Dinary et Afflerbach, 1995; Eeds et Wells, 1989). Mais comment de telles activités d'enseignement plus décentralisées peuvent-elles réellement soutenir l'apprentissage de la lecture littéraire dans le cas plus spécifique de l'œuvre intégrale en première secondaire ?

Problématique

La lecture littéraire à l'école: bref survol des différentes approches d'enseignement

D'abord, qu'est-ce que l'apprentissage de la lecture littéraire? Car si la littérature en classe constitue pour certains « la voie d'accès par excellence pour la formation de lecteurs réfléchis et créateurs » (Giasson, 1996, p. 53), il s'agit là d'un domaine d'enseignement encore insuffisamment circonscrit (Petitjean, 1990; Reuter, 1996). Ainsi, de nombreuses approches d'enseignement se sont chevauchées au cours des cinquante dernières années, selon que l'on a privilégié l'auteur, le texte ou l'élève lecteur. De manière très schématique, avant la démocratisation de l'enseignement, c'est une approche « sacralisante » qui domine dans les écoles d'enseignement secondaire au Québec: une approche essentiellement centrée sur les « bons » auteurs et les « beaux » textes, précieux objets utiles à la formation culturelle, morale et rhétorique d'une élite.

Puis, et bien que toutes ces approches se chevauchent dans la réalité, on peut avancer que les années 1970 ont été marquées au sceau de la narratologie. Cela a entraîné l'abandon de l'histoire littéraire et de la rhétorique au profit d'une approche structuraliste ou techniciste, essentiellement axée sur l'étude des structures narratives, souvent détachée de tout contexte (Langlade, 1991; Lebrun et Roy, 1999), et ce, au total détrimement des aspects socioaffectif et culturel qui justifient pourtant en grande partie, selon nous, l'enseignement de la lecture littéraire à l'école. Parallèlement, l'approche communicative se préoccupe davantage des besoins langagiers des élèves. Le texte littéraire devient peu à peu un discours parmi d'autres et souvent un simple prétexte pour faire « communiquer » les élèves plutôt qu'un objet d'étude en lui-même.

Enfin, les années 1990 sont dominées par l'approche stratégique issue des théories cognitives, ce qui aboutit à une autre forme d'approche techniciste dans les classes, mais cette fois axée sur les structures cognitives du lecteur et où l'objet-texte est plus ou moins considéré comme un ensemble neutre d'informations à traiter et à mémoriser.

Actuellement, les derniers programmes ministériels de français du primaire et du secondaire (ministère de l'Éducation du Québec [MEQ], 2001, 2003) tentent d'intégrer l'approche esthétique de la lecture, surtout centrée sur les réactions du lecteur. Dans cette approche développée par l'américaine Louise Rosenblatt (1994), à partir de certains principes issus du pragmatisme américain et des théories littéraires de la réception, le terme « esthétique » désigne cette posture qu'un lecteur adopte lorsqu'il s'engage dans la lecture afin de vivre une expérience immédiate et imaginaire à travers une création artistique de langage. Ce type d'approche vise l'enrichissement du vécu et de l'imaginaire du lecteur et encourage fortement l'expression des réactions personnelles, qui sont considérées comme des indices d'engagement et de compréhension. Conséquemment, l'enseignant devrait d'abord amener l'élève à « vivre » le texte plutôt qu'à le résumer et à l'analyser, et ce, en incitant l'élève à faire des associations avec sa vie personnelle, à éprouver des sentiments, à s'identifier aux personnages et à leur témoigner de l'empathie, à prédire la suite des événements, à imaginer des paysages, des visages, etc.

Outre le fait que, depuis 1995, la lecture d'au moins cinq œuvres complètes par année est prescrite au secondaire, tout semble indiquer que l'on cherche aujourd'hui à mieux

définir et évaluer la compétence littéraire dans les programmes ministériels. Cependant, et notamment dans le cas plus spécifique de l'œuvre intégrale qui nous occupe, la jonction et l'opérationnalisation des approches stratégique et esthétique ne vont pas sans soulever plusieurs problèmes en classe de lecture du secondaire.

Habiletés et habitudes de lecture des élèves de première secondaire

L'un des principaux problèmes qui se pose aujourd'hui en didactique de la littérature est, selon Jocelyne Giasson, « de concilier, dans la pratique, les interventions visant la compréhension des textes et la réaction à ceux-ci. Comment arriver à enseigner des stratégies de lecture tout en ne perdant pas de vue l'objectif réel de l'appréciation du texte littéraire? » (2000, p. 12). Car, faut-il le rappeler, selon des résultats d'enquêtes françaises qui s'appliquent à dresser des portraits détaillés des élèves à leur entrée au secondaire, environ 50% de ces élèves en seraient encore au stade de la reconnaissance des mots ou du décodage et 40% ne dépasseraient pas la compréhension d'un texte simple (Leclercq, 1981; ministère de l'Éducation nationale, 1993). Par ailleurs, l'enquête internationale PISA indique que 27% des élèves québécois âgés de 15 ans ne dépassent pas le niveau de l'inférence simple et que seulement 45% d'entre eux peuvent réussir des tâches de lecture complexes, comme interpréter le sens à partir de nuances de la langue et évaluer de manière critique un texte (PISA, 2001).

Pour ce qui est de la lecture d'œuvres littéraires, même si la plupart des élèves de la fin du primaire et du début du secondaire disent aimer lire, plus de 50% avouent ne pas lire du tout ou rarement des romans en dehors de l'école (Gervais, 1997; Soussi, 1995; MEQ, 1994a). De plus, comme l'école primaire privilégie surtout la lecture/plaisir, les élèves arrivent au secondaire en ayant très peu lu de textes plus résistants, c'est-à-dire de nature à susciter un travail sur le texte ou un début de distanciation (Sorin, 2001). Ce décalage exigerait d'eux un « saut qualitatif que la plupart ne peuvent franchir » (Tauveron, 1999, p. 12).

Journal dialogué et cercle de lecture: des dispositifs didactiques encore mal connus pour la construction progressive du sens en lecture

Selon les principes de l'approche esthétique, le premier objectif d'enseignement en lecture littéraire devrait être d'encourager la prise de conscience par l'élève *et* par le groupe de la nature évolutive de leurs processus de compréhension et de la pluralité des interprétations (Rosenblatt, 1982). Or il n'est pas certain que le type de tâches traditionnellement associées à la lecture littéraire dans les classes, comme les questionnaires de compréhension factuels et les résumés de lecture, favorisent l'atteinte de cet objectif ou le passage au « qualitatif » (Veck, 1997; MEQ, 1994b).

L'adoption de l'approche esthétique a par conséquent entraîné dans de nombreuses classes de littérature américaines et dans quelques classes du Québec — signalons à ce propos les travaux pionniers de Lebrun (1996a, 1996b) — la mise en application de nouveaux dispositifs didactiques, tels que le journal de lecture et les cercles littéraires (CL) en petits groupes de pairs, afin de favoriser un dialogue lecteur-texte-lecteurs. Les CL regroupent cinq ou six élèves aux habiletés variées qui ont lu le même roman et qui, souvent,

● ● Le roi, les légistes et le parlement de Paris aux XIV^e et XV^e siècles : contradictions dans la perception du pouvoir de « faire loy » ?

Sophie Petit-Renaud

... Texte intégral



- 1 Avant Bodin au XVI^e siècle, les juristes médiévaux inspirés par le droit romano-canonique ont fait de la loi un instrument privilégié de la puissance du prince. Le pouvoir de *condere legem* est reconnu au roi de France dès la fin du XII^e siècle¹ par les canonistes favorables à l'émergence de royaumes indépendants face à un pouvoir impérial déclinant. Le roi devenu *lex animata*², se voit attribuer la faculté de *dare, solvere, condere leges* par le droit romano-canonique, véhicule de l'exaltation d'une souveraineté législative. En commentant l'adage d'Ulpien, *Princeps legibus solutus est*, qui proclame l'indépendance du prince à l'égard de la loi³ et son corollaire la constitution *Digna Vox* des empereurs Théodose II et Valentinien III, qui affirme au contraire la soumission de l'empereur à la loi⁴, les légistes établissent une théorie de la loi⁵. Le prince est en principe délié des lois, les siennes et celles de ses prédécesseurs, mais s'y soumet volontairement. Les juristes tempèrent la portée de la maxime *Princeps legibus solutus est* par la constitution *Digna vox*, en recourant à un élément subjectif, la *voluntas principis*. Le prince, dégagé des liens de la loi, se lie lui-même, volontairement, à elle. Il faut souligner que le sens de la maxime *princeps legibus solutus est* se distingue de celui de la *potestas absoluta*, dégagée par Hostiensis à la suite des théologiens, à partir du concept de *plenitudo potestatis*, expression de l'omnipotence des papes⁶. Ce que le prince ne peut faire de sa puissance ordonnée, à laquelle s'applique la *Digna vox*, il le peut de sa puissance absolue.

1 A. Gouron, « La double naissance de l'État législateur », *Théologie et droit dans la science politique*(...)

2 L. Mayali, « *Lex animata*. Rationalisation du pouvoir politique et science juridique (XII^e-XIV^e siècle)(...)

3 D. 1, 3, 31.

4 C. 1, 14, 4.

5 K. Pennington, *The Prince and the Law. 1200-1600. Sovereignty and Rights in the Western Legal Tradit*(...)

6 K. Pennington, « Loi, autorité législative et théories du gouvernement, 1150-1300 », *Histoire de la*(...)

- 2 Le roi de France, empereur en son royaume, assimilé au *princeps* du droit romain, bénéficie des mêmes prérogatives ⁷. Le roi légifère en raison de la *majestas* ⁸, et parce qu'il est titulaire de l'*auctoritas* et de la *plenitudo potestatis* ou *plena potestas*. Sous le règne de Philippe le Bel, l'auteur de la *Disputatio inter clericum et militem* affirme clairement le droit du roi de France d'innover, de promulguer une norme nouvelle : le roi peut, comme l'empereur, *leges condere et addere eis vel demere* ⁹. Les choses sont encore plus claires dans le *Songe du vergier* (1378), parfaite expression de la doctrine royale. Son auteur, Évrard de Trémaugon, martèle le droit du roi de « faire loys et constitucions » qui lient les sujets, d'« ajouter aux dittes loys ou lez diminuisier ou en tout revoquier et rappeler, ainssi le roy puet faire loys ou constitucions toutes novellez entre sez subjés » ¹⁰. Dans la *Somme rural* (vers 1393), Jean Boutillier pose lui aussi avec vigueur le principe selon lequel le souverain, parce qu'il est « roy et empereur en son royaume », peut « faire loy et edict à son plaisir » ¹¹. On reconnaît là la formule d'Ulpian *Quod principi placuit legis habet vigorem*, tirée de la *lex regia* ¹² conférant au prince le pouvoir détenu originellement par le peuple. L'autorité du législateur est source de la loi. La volonté du prince est loi. Dans un ordre juridique où les lois divine et naturelle et le *ius gentium* fournissent les critères suprêmes, le roi est *supra ius positivum* ¹³.
- 3 La faculté du roi de France de modeler l'ordre juridique semble donc sans limite. C'est là consacrer une rupture avec la pensée traditionnelle qui faisait du roi le gardien, le conservateur d'un ordre immuable, d'un droit préexistant, extérieur à sa volonté, constitué par les coutumes et les ordonnances anciennes dont le temps a éprouvé la valeur ¹⁴. Au XIV^e siècle, cette conception traditionnelle est encore défendue par certains auteurs, tel Nicole Oresme, qui fait de la soumission de principe à un ordre juridique préexistant le trait essentiel du souverain. Oresme, traducteur des œuvres d'Aristote, se fait l'apôtre d'une monarchie tempérée, marquée par la souveraineté de la loi sur le prince. Le souverain doit selon lui « gouverner selon les estatuts et loys et ordenances qui luy sont baillés » ¹⁵. C'est là le reflet de la peur et de l'hostilité de principe à l'égard des « novelletés », par essence mauvaises ; nouveautés qui, non seulement, troublent les droits des sujets, mais nuisent aussi à l'efficacité des lois. La modification incessante de celles-ci risque, à terme, d'affaiblir le pouvoir royal lui-même. Mais tous, cependant, savent bien qu'une bonne loi doit être en conformité avec les mœurs d'un peuple ¹⁶. Quand une société évolue, la règle juridique doit suivre cette évolution. La prérogative législative ne s'analyse alors qu'en une nécessaire et prudente adaptation des lois anciennes lorsque les circonstances l'exigent, pour des raisons justes et nécessaires.

⁷ Nous nous permettons de renvoyer à notre thèse, « Faire loy » au royaume de France de Philippe VI à (...)

⁸ Concept exprimant la souveraineté avant Bodin, Y. Thomas, « L'institution de la majesté », *Revue de(...)*

⁹ *Disputatio inter clericum et militem*, N. Erickson, éd., *American Philosophical Society*, 111, 1967, p(...)

¹⁰ Le *Songe du vergier*, M. Schnerb-Lièvre, éd., Paris, 1982, t. I, chap. 36, § 49, pp. 56-57.

¹¹ Jehan Boutillier, *Somme rural ou le grand coutumier general de oractique civil et canon*, revu et an(...)

¹² D. 1, 4, 1.

¹³ C'est bien ainsi que l'entend Bodin qui ne cesse de rappeler la soumission aux lois divine et nature(...)

¹⁴ O. Guillot montre que le principe d'une soumission du prince à la loi est formulé dès le VI^e siècle(...)

¹⁵ Nicole Oresme, *Le livre de Politiques d'Aristote*, A. D. Menut éd., *Transactions of the American Phil(...)*

¹⁶ Gratien, Décret, D. 4, c. 2 : « Erit autem lex honesta, iusta, possibilis, secundum naturam, secund(...)

- 4 Il y a ainsi une tension entre, d'une part, l'affirmation d'une souveraineté législative sans limite, et, d'autre part, le rejet des innovations et la soumission de principe du roi à la loi, celle-ci étant alors entendue non pas comme un produit de la volonté du législateur *supra ius*, mais comme la prudente amélioration de ce qui existe déjà. L'affrontement doctrinal, entre les tenants du droit romano-canonique véhiculé de l'exaltation de la toute-puissance royale et leurs adversaires, les *politiques*, influencés par Aristote ¹⁷, a des résonances aussi bien dans la pratique, tout au long des actes normatifs royaux, que dans les plaidoiries et les arrêts du parlement de Paris ; deux sources qu'il faut examiner simultanément pour réfléchir sur quelques aspects de la perception du pouvoir de « faire loy » au Moyen Âge.

¹⁷ Il faut cependant souligner que les interprétations de l'œuvre d'Aristote ne sont pas uniformes comm(...)

Pratique royale

- 5 Le principe d'un nécessaire respect des bonnes lois des prédécesseurs, que leur ancienneté rend vénérables, n'est pas rejeté par le législateur. La volonté du roi de France de conserver les ordonnances anciennes, bonnes et raisonnables ne peut être ignorée tant elle est rappelée dans de nombreux actes. Philippe VI évoque ainsi les *plura statuta et arresta pro bono statu regni* de ses prédécesseurs dont il s'inspire pour interdire la création de juridictions d'appel par les nobles et les ecclésiastiques sans autorisation royale ¹⁸. Elu par les grands du royaume après la mort du dernier Capétien direct, il importe au Valois d'insister sur la continuité dynastique. De même, Jean II confirme le 1^{er} juin 1363 des lettres prises par son père sur le nombre des examinateurs au Châtelet de Paris, *Nos igitur factum genitoris nostri cupientes imitari ac roboris habere firmitatem* ¹⁹. Dans une ordonnance sur le statut des officiers royaux datant du 5 février 1389 (n. st), Charles VI souligne, lui aussi, sa volonté de « ensuyr les traces de [ses] predecesseurs, lesquels ont sur ce fait plusieurs ordenances et status » ²⁰.

¹⁸ *Ordonnances des rois de France de la troisième race* (désormais citées *Ord.*), E. de Laurière et alii, (...)

¹⁹ *Ord.*, t. IV, p. 233.

²⁰ Isambert, Jourdan et Decrusy, *Recueil général des anciennes lois françaises depuis l'an 420 jusqu'à* (...)



Formes et échelles du commerce d'occasion au XIX^e siècle. L'exemple du vêtement à Paris.

Manuel Charpy

p. 125-150

Texte intégral



Texte intégral en libre accès disponible depuis le 20 juin 2005.

- 1 Afin de décrire la culture matérielle et notamment celle des XVIII^e et XIX^e siècles, l'historien est tenté d'emboîter le pas au notaire et de procéder à l'inventaire méticuleux des biens matériels. Compilés, les inventaires offrent une image séduisante qui nous informe de la quantité des objets mais, plus rarement, de leur qualité.
- 2 Krzysztof Pomian proposait d'opérer une "coupe synchronique dans l'ensemble des objets visibles présents dans notre société" ¹. L'image résultant de cette opération s'apparente elle aussi à un instantané, mais elle offre l'avantage de souligner la diversité des statuts et des usages d'un même objet, bref de s'attacher autant aux qualités des objets qu'aux quantités. Outre qu'elle permet de récuser le temps processionnel des modes, cette image invite à s'interroger sur les déplacements, les voisinages et la permanente redéfinition des objets et de leurs statuts. Dans les inventaires, quelques indications sommaires font laconiquement mention de l'état des objets ² et la valeur que leur attribuent les experts *ad hoc* peut nous permettre d'inférer leur état, la rareté des matériaux ou encore l'engouement qu'ils suscitent. Hors de ces maigres indications, les inventaires restent silencieux sur l'inscription toujours renouvelée des objets dans le temps des usages sociaux. Entre le temps lisse et successif de la mode et le temps mort des inventaires, il s'agit de restituer en mouvement la complexité des formes et des modalités de consommation.

1 . Krzysztof POMIAN, *Sur l'histoire*, Paris, Éditions Gallimard, 1999, p. 228. Qu'il me soit ici permis(...)

2 . G. BENOÛ dans son *Code et manuel du commissaire-priseur*, Paris, D'Ocagne, 1835, recommandé de cons(...)

- 3 Procédé maintes fois éprouvé dans l'étude du cadre bâti, cette attention aux processus d'aménagements, d'affectation, de désaffectation, de contextualisation, de décontextualisation et de recontextualisation apparaît plus malaisée lorsqu'il s'agit d'objets mobiliers. Les circulations dans l'espace, dans les strates du temps, les processus de réactivation ou d'abandon comme les changements de propriétaire deviennent alors déterminants ³. Pour interpréter cette vie matérielle il est donc nécessaire de décrire et comprendre les circulations et non plus seulement les possessions. La difficulté, on le comprend, réside dans le silence des sources qui n'enregistrent que les acquisitions et accumulations et peu ou pas les processus complexes qui façonnent les modalités de consommation.
- 4 Le marché de l'occasion, souvent éclipsé par la nouveauté du grand magasin et pourtant loin d'être un fait résiduel, intéresse l'ensemble des secteurs de la consommation et s'offre comme un biais à qui veut saisir les processus et les modes de consommation. Régulièrement croisé dans l'analyse de la consommation du XIX^e siècle ⁴, il n'a pas encore réellement fait l'objet d'une analyse. Il permet pourtant d'interpréter la consommation à l'aide des questionnements croisés de l'usure et de l'obsolescence, des modes et des usages tout en évitant de confondre consommation et production.
- 5 Parmi les nombreux secteurs qu'il concerne, le commerce du vêtement apparaît comme un terrain propice à une telle analyse, tant par son importance que par son inscription dans le commerce social du XIX^e siècle.
- 6 Au-delà de l'image pittoresque d'un commerce au rabais, grouillant dans les interstices du commerce institué, sa position reste à déterminer par rapport au commerce d'alors et à ses innovations. Considéré de l'approvisionnement aux modes de distribution, il éclaire les modalités d'achat, d'usage et de vente. Enfin, la mise en lumière à la fois des lignes de partage qui se dessinent entre le vieux, le neuf et l'ancien, et de la distinction aujourd'hui établie entre le commerce de l'occasion et le commerce institué, nous donne à comprendre la nature des rapports qu'entretient le XIX^e siècle avec sa production matérielle.
- 7 Observer l'évolution matérielle d'un objet et les représentations qui s'y attachent n'est pas, à la différence de l'étude du bâti, chose aisée. Il apparaît cependant possible de reconstituer ici quelques séquences de ces circulations, déplacements et changements de destinations, qui constituent les linéaments de la consommation.

3 . Dominique POULOT, "Une nouvelle histoire de la culture matérielle ?", dans *Revue d'histoire modern(...)*

4 . Voir Jean-Paul ARON, "Débris", dans *Le mangeur du XIX^e siècle*, Paris, Éditions Michel Laffont, 197(...)

Les formes du commerce de l'occasion

Un commerce mélangé

- 8 Pour qui examine les images de la rue parisienne du XIX^e siècle à la recherche du commerce de vêtements d'occasion, la figure du marchand ambulant s'impose. Il est invariablement représenté chargé de vieux habits fripés et rapiécés, de débris d'uniformes militaires, et de quantité d'objets hétéroclites. Dans la tradition des "cris de Paris", ces images accusent l'archaïsme et la désuétude de cette activité. Traités comme motifs pittoresques, ces marchands apparaissent comme les figures emblématiques d'une société citadine instable et fragile. Elles marquent la distance qui sépare cette activité du commerce établi ⁵.
- 9 Le fatras des objets proposés trahit le rôle du hasard dans ce négoce. Le succès iconographique et littéraire de cette figure ne rend pas plus facile son interprétation. Les transactions de ces marchands, sans boutique, sans capital et par obligation légale toujours en mouvement, nous restent pratiquement inconnues. Quelques certificats délivrés aux indigents indiquent qu'un grand nombre se tournent vers le métier de "brocanteur d'habits ambulant" ⁶ ; le faible investissement nécessaire à la création d'un tel commerce est sans doute décisif. On est frappé de la très faible présence des fripiers dans les archives notariales et commerciales. Comme les fripiers niortais du XIX^e siècle, évoqués par Jean-Clément Martin, les fripiers parisiens apparaissent "en-deçà de la faillite" : ils sont très peu faillis et encore plus rarement créanciers ⁷. Ce quasi-silence des archives dessine en creux un commerce incertain et précaire, situé à la marge du commerce institué.
- 10 Au XIX^e siècle, les marchandes à la toilette, pendant féminin des marchands ambulants, font quant à elles, l'objet de fréquentes procédures administratives. C'est le cas de la veuve Fernel, installée rue des Fontaines-du-Temple, qui fait faillite en 1850 pour seulement 299 francs de créances que ne recouvrent pas les 191 francs de marchandises prisées ⁸. Sa marchandise est constituée par un bric-à-brac de vieux vêtements et d'accessoires de mode d'occasion difficile à distinguer de ses effets personnels. L'imbrication étroite des espaces de travail et des espaces domestiques ainsi que la confusion de ses dettes privées et professionnelles prolonge ce désordre. La situation est similaire pour de nombreux marchands. Ils cohabitent avec une théorie de marchés de quartier ; plus d'une quinzaine dans le Paris intra-muros de 1850. Malgré l'aspect moins informel de leurs éventaires, la marchandise paraît tout aussi médiocre et mélangée. Au Carreau Beauvau-Saint-Antoine, les "chiffons et les hardes sont pêle-mêle" ⁹. Le marché de Notre-Dame au début de la Monarchie de Juillet est un "marché de Vieux-Linges et de Bric-à-brac" ¹⁰. Celui des Carmes, se compose en 1837 "d'amas de hardes" et quarante ans plus tard, on y voit encore une "ignoble accumulation de défroques sans nom et de ferrailles de toutes sortes" ¹¹. Quant au Carreau du Temple dans le 6^e arrondissement d'alors, lieu hautement pittoresque en 1835 avec 1 550 marchands, il séduit encore l'historien d'aujourd'hui ¹². L'article de *l'Illustration*, publié en 1847, dont les termes connaîtront une fortune considérable, nous donne à voir des étalages de marchandises hétéroclites, crasseuses et grouillantes de vermine ¹³. Les auteurs se complaisent à décrire un monde interlope, où règnent escroquerie et dissimulation. L'aspect et la nature des articles qui peuplent ces marchés témoignent de leur précarité et de leur obsolescence.
5. "Les modes et la politique/Ont cent fois rempli ma boutique" déclare ainsi le marchand d'habits du(...)
6. Arch. préf. de police (archives de la préfecture de police de Paris) de Paris, Certificats d'indig(...)
7. Jean-Clément MARTIN, "Le commerçant, la faillite et l'historien", dans *Annales, Économies, société*(...)
8. Arch. de Paris (archives de Paris), D11U3/126. Dossier de liquidation judiciaire, 11 avril 1850.(...)
9. Arch. préf. de police, DA 178, rapport, 1837.
10. Arch. préf. de police, DA 307.
11. *La France*, 23 mars 1877.
12. Philippe PERROT, *Les Dessus et les dessous de la bourgeoisie, une histoire du vêtement au XIX^e siècle*(...)
13. *L'Illustration*, n° 233, 14 août 1847.

INDEXE

accord inter-annotateur 94, 249
amorce 79-81, 102, 140
amorce exacte 102, 106, 140
amorce implicite 102, 140, 154
amorce question 102, 140, 234
amorce vague 102, 140, 229
Annodis 71
annotation 77-78, 95-99
anaphore 40-51
anaphore associative 50
anaphore conceptuelle 49
anaphore fidèle 49-50
anaphore infidèle 49-50
anaphore résomptive 50-51, 84, 178
anaphorique coréférentielle 48
anaphorique co-significante 48
antécédent 49-51
article de recherche 26-33
auteur responsable 27
Brown Corpus 85
cadratif 53-55
Callisto 94-96
catégorie de série 125-127
chaîne de référence 51
clôture 79-84, 102, 140
clôture exacte 102, 110
clôture vague 102, 140
cohérence 35-38
cohésion 42-43, 63-64, 83
cohésion lexicale 51-52, 194, 196
communauté scientifique 22, 26, 29, 31, 91
conjonction 43-49
conjonction de coordination 101, 104
connecteur 43-49, 82-83, 100-101
connecteur de reformulation 44, 103
connecteur externe 25, 113-114
connecteur interne 25, 113-114
construction numéro + nom 101, 114, 180-181
contiguïté sémantique 195-198, 201-208, 211-212

contiguïté sémantique implicite 202, 205-208
 continuation (Goutsos 1996) 58, 62
 continuité par défaut 55, 58
 corpus 87-93
 correspondance des items aux exemples numérotés 201, 208-209
 déclencheur d'antécédent (*antecedent trigger*) 48-49
 deixis 47, 49
 déplacement (Goutsos 1996) 58-61
 discours 39-41, 56-57
 discours de recherche 22-34
 éducation 31-32
 encadrement (Goutsos 1996) 59, 60-61
 encadrement du discours 53-55
 énumération 74-77, 86
 énumération hétérogène (Luc 2001) 70-71
 énumération homogène (Luc 2001) 70
 énumération hybride (Luc 2001) 70-71
 énumération paradigmatique (Luc 2001) 70-71
 énumération syntagmatique (Luc 2001) 70-71
 équivalence structurelle des items 206, 214
 espace de discours 54, 173
 expression cadrative 53-55, 100, 174-175
 expression référentielle 44-51, 84, 101
 Frantexte 67, 88
 fonction discursive 15-16, 21
 grammaire du texte 37
 histoire 32-33, 217-239
 hyperthème (Danes 1974) 72-73, 104, 207
 IMRAD (introduction – méthode – résultats – discussion) 29, 33, 56
 indice complémentaire 99-103, 190-213
 indice complémentaire primaire 201-202
 indice de cohésion 43-56, 63-64, 83
 introduction (Goutsos 1996) 58, 60-61
 KIAP 22, 23, 27, 28, 29, 30
 langue maternelle 90-91
 lecteur responsable 27
 linguistique 30, 217-234
 linguistique de corpus 20-21
 marqueur adversatif 111, 112
 marqueur exact 1002-103, 125
 marqueur externe 25, 113-114
 marqueur de cohésion 43-56, 63-64, 83
 marqueur d'intégration linéaire (MIL) 65-66, 173-178

marqueur interne 25, 113-114
 marqueur relatif 103-104
 maxime de quantité (Grice 1975) 154, 182
 métadiscours 24-26
 métadiscours interactionnelle 25
 métadiscours interactive 25
 métadiscours interpersonnelle 24-25
 métadiscours textuelle 24-25
 Modèle d'architecture textuelle (MAT) 69-71
 modèle d'argumentation de Toulmin (1958) 103, 111
 modèle de Goutsos (1996) 57-64
 non marquage 63-64, 99-100, 101
 non parallélisme 70-71, 81, 113, 123-124, 130-134, 145-147
 ordre 73-74, 84-86, 96, 98-99
 ordre absolu 85, 98, 100-102
 ordre relatif 85, 98, 100-102
 organisation textuelle 56-86
 parallélisme des items 70-71, 75, 79-81, 130-134
 paraphrase 201-203, 211
Penn Discourse Treebank 44, 89, 91
Penn Treebank 89
 prédiction 60, 79-80
 principe d'énumération (PE) 68, 79-81, 125, 195-199, 206, 208
 principe d'énumération implicite 202, 206-208
 processus d'édition 18, 23, 91
 progression à thème constant (Danes 1974) 72-73, 166
 progression à thème dérivé (Danes 1974) 72-73
 progression à thème éclaté (Danes 1974) 72-73
 progression à thème linéaire (Danes 1974) 72-73
 progression thématique (Danes 1974) 72-73
 référence endophorique 47
 référence exophorique 47, 49
 répétition lexicale 51-52, 101, 192-194, 201-205
 répétition des éléments différents 102, 204, 211
 répétition partielle 194, 202-204, 211
 répétition proprement dite 201-204, 211
Rhetorical Structure Theory (RST, Mann et Thompson 1988) 16, 39, 42, 45, 63
 sciences humaines 28-29
 sciences naturelles 28-29
 sciences sociales 28-29
 segmentation (du texte et du discours) 56-57
 série additive 207-208
 série à deux temps 110-111, 115

série argumentative 110-111
série enchaînée 105-106, 110-111
série homogène (HOMO) 125
série isolée (Luc 2001) 70
série liée (Luc 2001) 70
série linéaire 74-76, 84-86
série méta 110-111, 115, 185-187
série métadiscursive 115, 177
série mixte (MIX) 125
série MIXAL 147
série MIXREST 147
série non marquée (NM) 126-127
série NMet 127
série partiellement non marquée (PNM) 126
série section 110, 181-186
série simple 104-105, 108-111
série temporelle 113-114, 179
Segmented Discourse Representation Theory, la Théorie des représentations discursives
segmentées (SDRT, Asher 1993 ; Asher & Lascarides 2003) 43, 71
sous thème 72-73
stratégie textuelle 57-59
technique de déplacement 60-61
technique de continuation 62-63
Text-strategic continuity (TSC) 59
texte 39-42
titre 181-182
univers de discours 53-54
XML 93, 97
zone de continuation 58-59
zone de transition 58-59