



HAL
open science

Éléments initiaux : combinaisons et schémas préférentiels dans un corpus d'articles scientifiques

Frédérique Mélanie-Becquet, Sophie Prévost

► To cite this version:

Frédérique Mélanie-Becquet, Sophie Prévost. Éléments initiaux : combinaisons et schémas préférentiels dans un corpus d'articles scientifiques. *Corpus*, 2014, Éléments initiaux dans la phrase : approches inter-genres et inter-langues, 13, pp.29-60. halshs-01250220

HAL Id: halshs-01250220

<https://shs.hal.science/halshs-01250220v1>

Submitted on 5 Jan 2016

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

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

Frédérique Mélanie-Becquet et Sophie Prevost

Éléments initiaux : combinaisons et schémas préférentiels dans un corpus d'articles scientifiques

Avertissement

Le contenu de ce site relève de la législation française sur la propriété intellectuelle et est la propriété exclusive de l'éditeur.

Les œuvres figurant sur ce site peuvent être consultées et reproduites sur un support papier ou numérique sous réserve qu'elles soient strictement réservées à un usage soit personnel, soit scientifique ou pédagogique excluant toute exploitation commerciale. La reproduction devra obligatoirement mentionner l'éditeur, le nom de la revue, l'auteur et la référence du document.

Toute autre reproduction est interdite sauf accord préalable de l'éditeur, en dehors des cas prévus par la législation en vigueur en France.

revues.org

Revues.org est un portail de revues en sciences humaines et sociales développé par le Cléo, Centre pour l'édition électronique ouverte (CNRS, EHESS, UP, UAPV).

Référence électronique

Frédérique Mélanie-Becquet et Sophie Prevost, « Éléments initiaux : combinaisons et schémas préférentiels dans un corpus d'articles scientifiques », *Corpus* [En ligne], 13 | 2014, mis en ligne le 01 mai 2015, consulté le 14 mai 2015.
URL : <http://corpus.revues.org/2427>

Éditeur : Bases, corpus et langage - UMR 6039

<http://corpus.revues.org>

<http://www.revues.org>

Document accessible en ligne sur : <http://corpus.revues.org/2427>

Ce document est le fac-similé de l'édition papier.

© Tous droits réservés

Éléments initiaux : combinaisons et schémas préférentiels dans un corpus d'articles scientifiques

Frédérique MELANIE-BECQUET
Sophie PREVOST
Lattice-UMR 8094 (CNRS / ENS /
Université Sorbonne Nouvelle Paris3)

1. Introduction

Zone préverbale (Rainsford *et al.* 2012), position initiale (Ho Dac 1997), périphérie gauche (Beninca 2006), « complex beginnings » (Smits 2002), zone thématique (Halliday 1994)... : la diversité des dénominations pour désigner la zone qui précède le verbe et/ou le sujet traduit à la fois l'intérêt porté à celle-ci et la pluralité des approches qui tentent de l'appréhender, tant du point de vue de la perspective (morpho-syntaxique, syntaxique, sémantique, informationnelle) que de la posture théorique (linguistique fonctionnelle, grammaire générative...).

Notre objectif ici est d'étudier de manière globale, dans un corpus d'articles scientifiques en sciences humaines, les combinaisons et les schémas préférentiels des éléments initiaux, désormais EI, c'est-à-dire qui occupent une position avant le sujet (et avant le verbe lorsque le sujet est inversé).

L'étude s'appuie sur un travail collectif de réflexion et d'annotation accompli au sein du groupe EIOMSIT (*Eléments Initiaux, Ordre des Mots, Structuration Informationnelle et Textuelle*) du laboratoire Lattice de 2010 à 2013¹.

La détermination des schémas possibles est envisagée en relation avec les caractéristiques des séquences sujet-verbe. Il s'agit donc de rendre compte de la structuration de ce que nous appellerons ici la zone préverbale, entendue dans un sens descriptif. Quels types d'EI peut-on y trouver ? Comment se

¹ Quatre autres articles de ce numéro présentent des études issues de ce travail : il s'agit des contributions de C. Fuchs, A. Celle et L. Lansari, E. Havu et S. Carter-Thomas.

combinent-ils (c'est-à-dire combien peut-on en trouver, et de quel type) ? Leur agencement est-il lié à la nature du sujet ? Nous nous attacherons en outre à la mise au jour d'éventuelles divergences entre textes, selon la discipline dont ils relèvent.

Voici un exemple dans lequel 3 EI précèdent un sujet nominal :

- (1) [Mais], [en même temps], [du fait même de leurs tribulations],
les héros traversent tous les continents et tous les milieux.

Nous présenterons dans un premier temps (2) les aspects méthodologiques de cette étude, puis l'analyse du corpus (3), en envisageant tout d'abord quelques caractéristiques générales des sujets et des EI (3.1), puis en nous attachant plus spécifiquement à la combinaison des EI dans les chaînes (3.2). L'approche adoptée dans cet article assume son caractère général, quantitatif et descriptif. Nous tenterons cependant en conclusion (4) de replacer les tendances et les éventuelles régularités mises au jour dans une perspective interprétative plus générale, qui permettra d'évaluer l'apport d'une étude de corpus à la problématique du numéro.

2. Aspects Méthodologiques

2.1 Présentation du corpus

Notre analyse a porté sur une partie du corpus *Chambers-Le Baron corpus of Research Articles in French*². Il s'agit d'une ressource libre de droits, constituée d'articles scientifiques écrits entre 1998 et 2006. Ce corpus a été choisi en raison de son homogénéité : il comporte des articles scientifiques en français, qui relèvent des humanités et des sciences sociales au sens large. L'hypothèse de départ était que, en dépit de l'homogénéité résultant de l'appartenance à un même « genre », celui de l'article scientifique, des différences pourraient émerger selon les disciplines, et ainsi traduire une certaine hétérogénéité. Le corpus comprend en effet des articles relevant de 10 domaines différents, parmi lesquels nous avons sélectionné les suivants : Littérature

² <http://ota.ahds.ac.uk/desc/2527>

Eléments initiaux : combinaisons et schémas préférentiels

(Litt), Droit (Droit), Economie (Eco) et Histoire (Hist)³. Sur l'ensemble des articles de ces 4 domaines, nous avons annoté 36 articles, soit 209 629 mots.

Tableau 1. Dimensions du corpus

	Eco	Hist	Droit	Litt	Total
Nbre de textes	8	10	8	10	36
Nbre de mots	51 142	71 884	34 523	52 080	209 629

2.2 Structures et modèle d'annotation

Notre unité d'étude de référence est la proposition. Nous n'avons traité que les propositions déclaratives avec un verbe conjugué et un sujet exprimé, qu'elles soient suivies et/ou précédées d'une subordonnée, ou non. Ne sont donc pas prises en compte les propositions coordonnées sans sujet exprimé.

Les structures que nous étudions sont définies par la présence d'un sujet et du ou des éventuel(s) EI qui le précède(nt). Nous avons écarté les structures avec sujet postverbal (1.6 % de l'ensemble des sujets), qui sont plus spécifiquement étudiées dans l'article de C. Fuchs ici-même (« Les éléments initiaux dans les phrases à sujet inversé »).

Nous avons codé et analysé 7 631 structures. Nous revenons plus en détail ci-dessous sur les critères de délimitation des structures, sur les unités et sur les catégories annotées.

Deux catégories ont été annotées : les sujets et les éléments initiaux.

2.2.1 Les sujets

Pour ce qui est du sujet, nous avons caractérisé sa nature et sa position. Pour la première, nous avons distingué quatre catégories : les syntagmes nominaux, avec ou sans expansion, les pronoms personnels (dont le pronom *on*), le pronom impersonnel *il*, et les autres pronoms (indéfinis et démonstratifs). Pour ce qui

³ Des annotations ont été effectuées sur des articles de 3 autres domaines – anthropologie, communication et linguistique, soit un total de 9 articles (3 articles par domaine). Il ne sera pas question de ces annotations ici : elles ont servi à tester et valider la structure d'annotation, mais le volume par domaine n'est pas suffisant pour qu'il en soit rendu compte ici.

est de la position, nous avons distingué les sujets préverbaux et les sujets postverbaux. Seuls les premiers sont retenus pour la présente analyse.

2.2.2 Les éléments initiaux (EI)

Pour ce qui est de la délimitation des EI, nous avons opté pour un principe de segmentation maximale (et sans tenir compte de la ponctuation), y compris dans des cas de juxtaposition ou de coordination de syntagmes relevant d'une même catégorie sémantique, et non séparés par une virgule⁴. Ainsi dans *En France(.) à Paris*, on considère que l'on a 2 EI. Il en va de même dans *et pourtant* ou *mais néanmoins*.

La conjonction *et* a cependant fait l'objet d'un traitement spécifique : lorsque *et* coordonne un élément, c'est-à-dire qu'il n'est pas en tête de proposition, il est inclus dans l'élément qu'il coordonne, et ne constitue pas un EI à part. Voici les différents cas possibles de traitement de *et* :

[Et] il sortit sans se retourner : 1 EI

[Et], [s'il bouge], il y va de sa vie : 2 EI

[Soulagé d'avoir fini] [et heureux d'être en vacances], il était rayonnant : 2 EI

Une fois l'EI délimité, nous avons pris en compte deux aspects : son rang et sa catégorie.

Nous avons en effet spécifié pour chaque EI son rang, qu'il s'agisse d'un EI isolé (1/1) ou faisant au contraire partie d'une série d'EI, par exemple 2/3 (2^e EI sur 3).

Le classement des EI qui a été adopté s'appuie sur des critères sémantiques et morpho-syntaxiques. Il apparaît en effet que les compléments dits circonstanciels, par exemple, ont un fonctionnement différent selon leur sémantisme (manière, spatio-temporel...) au regard de la position initiale, de leur combinaison, voire des caractéristiques du sujet. Il en est de même pour

⁴ Deux arguments ont guidé ce choix : d'une part, il nous semblait moins subjectif de dissocier des EI que de les associer ; d'autre part, dans une perspective de réutilisation de la ressource constituée avec possiblement d'autres choix de segmentation, il est plus aisé de réunir des EI qui auraient été séparés que de segmenter un EI en plusieurs EI.

Eléments initiaux : combinaisons et schémas préférentiels

certaines éléments « périphériques » (connecteurs, prédications secondes) Nous avons au final déterminé 11 catégories :

a) Eléments logico-pragmatiques

Il s'agit d'éléments, majoritairement des adverbes ou des conjonctions, qui possèdent une fonction connectrice forte. Celle-ci peut s'exercer au niveau de l'énoncé (*et, mais, car, pourtant, en effet, de même,...*), et au niveau de l'énonciation sous la forme de marques phatiques ou conatives, dont relèvent les apostrophes et interpellations, qui assurent une connexion avec l'interlocuteur. Notre corpus, composé d'articles scientifiques, ne comporte cependant pas d'EI de ce dernier type.

- (2) [Mais] *le caractère exemplaire de ces cas* se heurte à une multitude d'exemples opposés, notamment en Afrique ou en Amérique latine. (Eco)

b) Commentaires énonciatifs

Ces éléments ont en commun de dénoter l'attitude du locuteur vis-à-vis de l'énoncé (*vraiment, peut-être, sans doute, évidemment... par exemple, en particulier, certes*) :

- (3) [Bien sûr], *le rituel du mariage* a surtout une fonction symbolique. (Hist)

ou de l'énonciation (*franchement, à vrai dire, sans rire, en clair...*) :

- (4) [Comme nous l'avons indiqué plus haut], *les premières critiques* furent formulées par les autorités métropolitaines dès qu'on les avisa de l'adoption d'une politique de prime. (Hist)

La distinction que nous avons opérée entre les éléments logico-pragmatiques et les commentaires énonciatifs s'appuie largement sur celle proposée par Nølke (1993). Pour le classement des adverbiaux « contextuels » (c'est-à-dire non spatio-temporels), Nølke retient deux paramètres : le type de contexte (textuel et non-textuel, c'est-à-dire le cotexte et la situation énonciative), et la nature de ce que les adverbiaux commentent (l'énoncé ou l'énonciation). La prise en compte de ces deux paramètres conduit à la constitution de trois classes :

- i) non textuel/énonciation : adverbiaux d'énonciation

- ii) non textuel/énoncé : adverbiaux d'énoncé
- iii) textuel/énonciation et énoncé : connecteurs

La différence majeure entre les trois types est que la fonction de connexion est première chez les connecteurs alors qu'elle est seconde chez les adverbiaux d'énonciation et d'énoncé. Les premiers s'apparentent dans notre classification aux éléments logico-pragmatiques, les seconds et les troisièmes aux commentaires énonciatifs⁵.

c) Eléments exprimant une relation logique

Cette catégorie rassemble l'ensemble des relations logiques.

Cause :

- (5) [Avec les événements de 1848], *la Préfecture de police, absorbée par des tâches de surveillance politique*, relâcha la surveillance sur les chemins de fer. (Hist)

On notera que, dans cet exemple, « avec les événements de 1848 » a une valeur de base temporelle ; c'est toutefois la valeur causale qui prévaut ici. L'hésitation entre valeur temporelle et causale n'est pas rare : nous avons toujours choisi la valeur qui s'imposait dans le contexte, avec, inévitablement, une part de subjectivité propre aux annotateurs⁶.

Hypothèse / condition :

- (6) [Si ce déficit alimente la demande dans un contexte de ralentissement de l'activité économique], *il* peut avoir un effet positif de relance de type keynésien. (Eco)

But / conséquence :

- (7) [De manière à faciliter l'application de cette résolution], *l'Organisation* créa un Département Maladies transmissibles. (Droit)

⁵ Voir ici-même l'article de S. Carter-Thomas dédié aux commentaires énonciatifs dans une perspective contrastive anglais-français.

⁶ Les textes ont été annotés par un seul annotateur, mais, d'une part, les cas ambigus ont été discutés collectivement, d'autre part nous avons opéré des vérifications par catégorie, et des corrections si nécessaire, afin d'obtenir la meilleure homogénéité possible.

Éléments initiaux : combinaisons et schémas préférentiels

Comparaison / conformité :

- (8) Ainsi, [de même qu'on offrait de nombreux présents lors de cérémonies périodiques afin d'entretenir les alliances], *on* en distribuait également autant au départ des partis de guerre. (Hist)

Opposition / concession :

- (9) En réalité, [alors que la loi de séparation avait, au moins en partie, l'intention de brimer l'Eglise], *c'est* celle-ci qui sort, appauvrie certes, mais victorieuse, du combat entre les deux institutions. (Hist)

On notera que « alors que la loi de séparation avait, au moins en partie, l'intention de brimer l'Eglise » a une valeur première temporelle. C'est cependant la valeur d'opposition qui s'impose ici. Comme pour la valeur causale, l'hésitation entre valeur temporelle et valeur d'opposition n'est pas rare.

Additif :

- (10) [Outre des dispositifs spatiaux], *il* décrit un système d'enregistrement permanent sous forme de rapports hiérarchisés. (Hist)

Les catégories a) et c) de notre classement peuvent sembler proches. Elles se distinguent cependant d'un point de vue formel et fonctionnel. Sur le plan formel, les éléments logico-pragmatiques correspondent surtout à des adverbes, des locutions verbales figées, des syntagmes nominaux (Sn) ou pronominaux figés (*à cet effet, pour cela*), contenant généralement un élément anaphorique. Dans la catégorie Relations logiques, on trouve des Sn non figés, des subordonnées, des propositions infinitives ou participiales. Cette différence formelle induit une double différence fonctionnelle. D'une part, le « poids » (longueur et complexité) de ces éléments n'est pas le même, ce qui a des conséquences sur leur position et sur leur possible multiplication. D'autre part, la fonction de connexion est forte pour les éléments logico-pragmatiques, elle est secondaire pour les Relations logiques.

d) Prédications secondes

Cette catégorie regroupe des éléments de nature différente : apposition nominale, participe passé, ou adjectif, parfois participe présent. Voici un exemple :

- (11) [Décrite en fonction de mécanismes d'appropriation ou de contestation] *la notion de « temps politique »* peut prêter à confusion. (Hist)

Souvent, les participe présents ou gérondif ont une valeur causale, ils sont alors classés comme relation logique⁷ :

- (12) [En refusant la loi], *l'Eglise* a perdu la disposition des édifices du culte et la propriété des biens. (Hist)

e) Eléments « détachés »

On ne retient dans cette catégorie que les éléments détachés non introduits par un marqueur de topicalisation (*quant à, en ce qui concerne...*). La plupart de ces éléments ont une marque de coréférence ultérieure dans la proposition :

- (13) Dès lors, [ce qui importe], *ce* n'est pas le silence de Nathalie Sarraute quant aux arts de la table mais, plus globalement, ce dont ce silence témoigne. (Litt)

Mais on classe aussi dans cette catégorie un cas tel que :

- (14) [les enfants], *tu* fais une croix sur la grasse matinée (exemple construit)

f) Compléments essentiels

L'extension donnée à cette catégorie excède les seuls objets directs et indirects, compléments locatifs et attributs. Elle inclut tout élément (ou partie d'élément) dépendant du verbe, selon un lien plus ou moins étroit :

- (15) [De la politique], *Luc Ferry* a longtemps pensé avoir saisi l'essentiel. (Hist)

- (16) [De cette prochaine guerre annoncée], *il* sortira des gagnants et des perdants mais au-delà, il y aura surtout des Etats et

⁷ Voir ici même l'article de A. Celle et L. Lansari consacré aux prédications secondes, dans une perspective contrastive anglais-français.

Eléments initiaux : combinaisons et schémas préférentiels

leurs populations qui seront encore plus victimes de ses effets. (Droit)

Les catégories qui suivent définissent les limites représentationnelles des univers de discours (Charolles 1997) et ont une fonction adverbiale.

g) Eléments spatiaux et temporels

- (17) [En Italie,] la raréfaction de l'offre de titres publics a coïncidé avec l'introduction dans le paysage financier des sociétés de gestion. (Eco)

h) Eléments spatiaux « abstraits »

- (18) [Dans une démarche économétrique de cette nature], *le signe attendu des coefficients des variables* est fonction des pré-supposés théoriques des auteurs. (Eco)
- (19) A partir de cette ordonnance, [en droit], *le préfet de police* exerçait les attributions qui lui étaient dévolues. (Hist)

Il nous a semblé utile de distinguer les cas relevant d'une spatialité concrète (cf. 17) et ceux relevant d'une spatialité abstraite, même si la limite entre spatiaux concrets et spatiaux abstraits n'est pas toujours aisée à établir.

i) Compléments de manière

- (20) [Implicitement], le service de la dette lié aux flux de capitaux entrant et sortant est intégré. (Eco)

j) Elément Champ notionnel

Contrairement aux éléments détachés (e), ils sont introduits par un marqueur de topicalisation :

- (21) D'autre part, [en ce qui concerne la connotation positive (concorde) du terme de concordat], *Bruno Neveu* montre qu'à partir de 1911 un rapprochement s'est opéré. (Hist)
- (22) Ainsi, [pour le trafic « banlieue »], *la présence de trains à intervalles réguliers et rapprochés, le cadencement*, est un argument pour attirer les voyageurs. (Eco)

k) Point de vue

Il s'agit des éléments qui spécifient la source d'énonciation :

- (23) En effet, [selon la loi de 1905], *le culte* ne peut être exercé publiquement sans associations culturelles. (Hist)
- (24) [Comme le rappelle Blake (2000)], *cette tendance* s'est très rapidement affirmée, à tel point que 85 % des fonds de retraite d'entreprise qui ont été ouverts au cours de l'année 1997 étaient à cotisations définies. (Eco)

Nous avons pris le parti de n'attribuer qu'une seule valeur de catégorie à chaque EI, et donc de résoudre le caractère ambigu, voire indéterminé, de certains EI du point de vue de notre classement.

Il est en effet des cas difficiles (porosité entre certaines valeurs) pour lesquels on peut hésiter entre deux étiquettes (cf. certains des exemples mentionnés ci-dessus) : nous avons choisi dans ce cas celle des deux qui nous semblait prévaloir.

Pour clore la présentation de notre classement, il convient d'insister sur un point : tout codage résulte d'une classification, et toute classification est conventionnelle, et l'aboutissement de divers partis-pris. Elle est donc discutable et contestable. Nous proposons ici une grille de lecture des éléments initiaux, parmi d'autres possibles.

2.3 Une solution technique adaptée à l'annotation des chaînes d'EI

Des caractéristiques énoncées ci-dessus résulte une structure d'annotation constituée de 5 propriétés, et de 34 valeurs. Il s'agit de 3 propriétés associées au sujet : type (4)⁸, présence ou non d'une expansion (2), position (2), et des 2 propriétés associées aux EI : catégorie (11) et rang (15). L'analyse n'a pas uniquement porté sur les éléments indépendamment les uns des autres, mais aussi, et surtout, sur la chaîne EI-Sujet.

La nécessité de travailler en groupe, le besoin d'analyser des chaînes d'éléments nous a menés à choisir comme

⁸ Entre parenthèses est indiqué le nombre de valeurs de la propriété concernée.

outil d'annotation *Analec*⁹ (Landragin *et al.* 2012), logiciel particulièrement adapté à notre démarche de travail.

La structure d'annotation définie ci-dessus a été implémentée sous *Analec*. Il est possible, sous *Analec*, d'enrichir et de modifier une structure d'annotation au fil du travail, et de lui attribuer un formalisme, nommé « vue ». Une mise en forme a ainsi été associée à chacun des éléments : les sujets sont grassement mis en surbrillance, les EI mis en surbrillance. De plus une mise en forme a été attribuée à chacune des 34 valeurs (un code couleur est ainsi associé à chacune des catégories d'un EI).

A tout moment de l'annotation, l'outil fournit des statistiques sur les annotations effectuées, à l'échelle du texte ou du corpus. Il est en effet facile de concaténer dans *Analec* un ensemble de textes ayant une même structure d'annotation, ce qui permet de dégager des tendances et des problèmes d'annotations communs à l'ensemble des textes.

Afin d'automatiser la tâche d'annotation, nous avons mis en place des scripts permettant soit de générer, compter et observer nos chaînes d'annotation (une chaîne étant composée d'un ou plusieurs éléments initial/ux suivi(s) d'un sujet), soit de lister et corriger les erreurs d'annotation (telles que l'oubli d'annotation d'un élément de la chaîne, la mauvaise attribution de la valeur du rang de l'EI ou la double annotation d'un même élément). Par ailleurs, si *Analec* permet de concaténer les données, il ne possède pas de fonction de partition de corpus offrant la possibilité de faire des calculs, non sur l'ensemble du corpus mais sur des sous-parties du corpus, afin de les comparer les unes aux autres. Les scripts permettent de pallier ce manque.

3. Analyse

3.1 Données générales

Nous avons annoté 7631 structures, une structure étant composée d'un sujet, et, le cas échéant, du ou des EI qui le précède(nt).

⁹ Le logiciel *Analec*, développé au Lattice par B. Victorri, est téléchargeable à l'adresse suivante : <http://www.lattice.cnrs.fr/Telecharger-Analec>.

3.1.1 Les sujets

Sur l'ensemble des 7 631 sujets, 4 559 (59.7 %) ne sont précédés d'aucun EI, tandis que 3 072 (40.3 %) sont précédés d'un ou plusieurs EI.

Cette prévalence globale du sujet en position initiale se retrouve au sein de chacun des 4 domaines : les sujets non précédés d'un EI restent dans tous les cas majoritaires, ils représentent de 55.7 % à 61.8 % de l'ensemble des sujets. C'est dans les articles de droit que l'on trouve le plus grand nombre de sujets précédés d'un EI (44.3 %), ce qui est sans doute lié à la nature très argumentative des textes dans cette discipline.

Tableau 2. Fréquence respective des sujets initiaux et des sujets précédés d'EI dans les différentes disciplines

	Eco	Hist	Droit	Litt	Total
EI + Sujet	669 <i>38.6¹⁰</i>	1 198 <i>41.3</i>	433 <i>44.3</i>	772 <i>38.2</i>	3 072 <i>40.3</i>
Sujet initial	1 066 <i>61.4</i>	1 701 <i>58.7</i>	544 <i>55.7</i>	1 248 <i>61.8</i>	4 559 <i>59.7</i>
Total	1 735	2 899	977	2 020	7 631

Nous avons fait l'hypothèse que la nature du sujet pourrait varier selon qu'il est précédé ou non d'un ou plusieurs EI, ou formulé autrement, que les EI pourraient être davantage associés à tel ou tel type de sujet.

Le tableau suivant présente, de façon globale, la distribution des types de sujet selon qu'ils sont précédés ou non d'un ou plusieurs EI :

Tableau 3. Distribution des types de sujet (avec et sans EI)

	Sn	P.perso	P.imper	P.autres	Total
Sujet initial	3 172 <i>69.6</i>	816 <i>17.9</i>	251 <i>5.5</i>	320 <i>7</i>	4 559

¹⁰ Les nombres en italiques correspondent à la fréquence relative des Sujets initiaux versus Sujet précédés d'EI.

Eléments initiaux : combinaisons et schémas préférentiels

1 EI + Sujet	1 772 68.9	492 19.1	166 6.5	143 5.6	2 573
2 EI + Sujet	303 66.3	100 21.9	31 6.8	23 5	457
3 EI + Sujet	19 50	13 34.2	2 5.3	4 10.5	38
4 EI + Sujet	2	1	0	0	3
5 EI + Sujet	0	0	0	1	1
Total EI + sujet	2 096 68.2	606 19.7	199 6.5	171 5.6	3 072

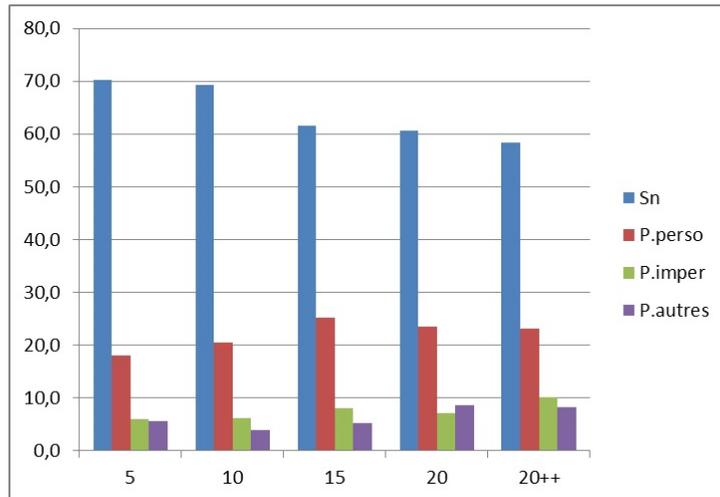
Il ressort de ces données qu'aucune différence significative n'apparaît dans la distribution des types de sujet selon qu'ils sont précédés ou non d'EI. Dans les deux cas, les Sn sont largement en tête (plus de 68 % de l'ensemble des sujets), suivis de loin des pronoms personnels (entre 17.9 % et 19.7 %), puis des pronoms impersonnels et autres pronoms (entre 5.5 et 7 %) ¹¹.

Toutefois, si l'on tient compte du nombre d'EI qui précèdent le sujet, on voit apparaître des variations notables : quand le sujet est précédé de 3 EI, la fréquence de Sn fléchit (50 %), et, à l'inverse, celle des pronoms personnels et celle des pronoms autres augmente : 34.2 % et 10.5 %.

L'affinement de ces résultats selon les différentes disciplines ne laisse paraître que des différences très ponctuelles. Ainsi en Economie et en Droit, la fréquence des Sn derrière 2 EI est supérieure à la moyenne (66.3 %), elle est respectivement de 78.3 % et 76.1 %.

Indépendamment du nombre d'EI, nous avons considéré le nombre de mots qui précèdent le sujet. Nous avons pour cela regroupé les mots par tranche de 5, et observé, pour chacun des types de sujet donné, la représentation de chacune de ces tranches.

¹¹ La faible fréquence des pronoms personnels dans les articles scientifiques français n'est pas spécifique à ce corpus et a été constatée dans d'autres études (voir en particulier (Fløttum *et al.* 2006)).



Graphique 1. Distribution des types de sujets selon le nombre de mots qui précèdent le sujet

Comme le montre le graphique 1, les variations sont mineures. Elles sont convergentes avec celles que nous avons observées en considérant le nombre d'EI qui précèdent le sujet (tableau 3)¹². En effet la fréquence des Sn baisse légèrement au fur et à mesure que le nombre de mots augmente. À l'inverse, on constate une légère hausse du nombre de pronoms.

Nous avons utilisé le calcul du khi2 pour évaluer le caractère plus ou moins significatif de la distribution des types de sujets au regard du nombre de mots qui les précèdent, et donc la liaison plus ou moins forte entre les deux variables (nombre de mots et type de sujet).

Le khi2 est de 28.94 ; la probabilité qui lui est associée est de 0.00402, bien inférieure au seuil de 0.05 au-dessus duquel on ne peut rejeter l'hypothèse nulle. La répartition observée n'a

¹² L'hypothèse est qu'un nombre croissant d'EI est lié à un nombre croissant de mots, même s'il est vrai qu'un EI peut être plus ou moins long, et qu'une séquence de 3 EI peut donc être plus courte qu'une séquence de 2 EI. Pour affiner ce point, il conviendrait en toute rigueur de combiner nombre d'EI, nombre de mots, mais aussi nombre de caractères par mot. Nous laissons pour l'instant de côté l'approfondissement de cet aspect, nous en tenant à un constat assez global et général, qui doit de ce fait être considéré avec prudence.

Eléments initiaux : combinaisons et schémas préférentiels

donc quasiment aucune chance de résulter d'une distribution aléatoire. Il y a bien une forme de dépendance entre nombre de mots et nature du sujet, qui se traduit par une attraction entre les sujets pronominaux et un nombre élevé de mots et par une répulsion entre les Sn et un nombre élevé de mots.

Il reste que, globalement, indépendamment de la présence ou non d'EI, de leur nombre, de leur longueur, les Sn constituent de loin la catégorie la plus fréquente.

3.1.2 Les éléments initiaux

Nous présentons ci-dessous quelques données générales sur les 3 628 EI annotés, avant d'aborder plus en détail (3.2) leurs combinaisons.

Il s'agit de voir, parmi les 11 valeurs que peuvent endosser les EI, quelles sont celles qui sont les plus présentes, dans l'ensemble du corpus mais aussi selon les disciplines dont relèvent les textes.

Si l'on considère l'ensemble du corpus, on constate que les EI logico-pragmatiques sont les plus fréquents, suivis des éléments spatio-temporels et des relations logiques, puis des éléments spatiaux abstraits et des commentaires énonciatifs. Viennent ensuite les autres catégories, dont la fréquence est égale ou inférieure à 5 %.

Tableau 4. Fréquence des EI selon leur catégorie

	Fréquence absolue	Fréquence relative
logico-pragmatique	1 148	31,7
spatio-temporel	779	21,5
relation logique	701	19,4
spatial abstrait	335	9,3
commentaire énonciatif	195	5,4
prédication seconde	177	4,9
point de vue	94	2,6
élément détaché	69	1,9
complément notionnel	52	1,4
complément de manière	44	1,2
complément essentiel	24	0,7
Total	3 618	100,0

Si l'on considère les 3 catégories les plus fréquentes dans chacune des disciplines, on constate que, dans toutes, on retrouve dans le peloton de tête les trois mêmes types d'éléments : logico-pragmatiques, relations logiques et spatiaux-temporels, néanmoins diversement ordonnés. Les EI logico-pragmatiques arrivent ainsi en tête dans 3 des 4 disciplines, leur fréquence variant de 31.5 % (Eco) à 40.5 % (Litt). Ils sont suivis, dans un ordre variable, des EI spatiaux temporels ou dénotant une relation logique. Les sous-disciplines Economie et Droit attribuent ainsi la seconde position aux Relations logiques, tandis que celle-ci est occupée par les EI spatiaux-temporels en Littérature. Un examen rigoureux des textes du corpus serait nécessaire, mais l'on peut néanmoins affirmer qu'il n'est pas surprenant de voir émerger certaines affinités entre les textes d'Economie et de Droit, du fait de leur contenu thématique.

En Histoire, ce sont en revanche les éléments spatiaux-temporels qui occupent la première position (26.3 %), mais ils sont suivis d'assez près par les éléments logico pragmatiques (25.5 %). La fréquence plus élevée des éléments spatiaux-temporels en histoire s'explique par des besoins de repérage spatio-temporel plus marqués, directement liés au contenu thématique des textes.

On n'approfondira pas davantage cette analyse, d'autant que, pour les éléments en 4^e ou en 5^e position, il s'agit de fréquences avoisinant les 5 %.

On retiendra de ces données le caractère relativement homogène de la distribution des EI les plus fréquents dans les différentes disciplines. On insistera aussi sur le fait que, toutes disciplines confondues, les éléments logico-pragmatiques et ceux dénotant une relation logique représentent 51.1 % de l'ensemble des EI. Cette fréquence très élevée d'éléments présentant, de manière plus ou moins marquée, une dimension argumentative, semble s'expliquer assez aisément par leur appartenance à un même « genre », celui de l'article scientifique.

3.2 Analyse des chaînes

3.2.1 Longueur des chaînes

Le corpus annoté comprend 3 618 éléments initiaux, qui se répartissent en 3 072 chaînes¹³ – c'est-à-dire la succession d'un ou plusieurs éléments initiaux suivi(s) d'un sujet. Nous reprenons ici de manière synthétique le tableau 3, en spécifiant le nombre de chaînes.

Tableau 5. Fréquence des chaînes selon leur nombre d'EI

	1 EI + sujet	2 EI + sujet	3 EI + sujet	4 EI + sujet	5 EI + sujet	total
Fréq. chaînes	2573 83.8	457 14.9	38 1.2	3 0.1	1 0	3072 100
Fréq. EI	2573	914	114	12	5	3618

Le tableau ci-dessus montre que les chaînes composées d'un seul EI s'imposent massivement, puisqu'elles représentent 83.2 % de l'ensemble des chaînes. Elles sont suivies de très loin par les chaînes à 2 EI (14.9 %), tandis que celles à 3 EI et plus s'avèrent marginales.

Si l'on considère la fréquence des différentes chaînes au sein de chaque discipline, on constate que la répartition est largement similaire : les chaînes à 1 EI représentent entre 78.1 % et 87.1 % de l'ensemble des chaînes, et celles à 2 EI entre 11.9 % et 19.7 %. On signalera toutefois la fréquence légèrement moindre des chaînes à 1 EI en Littérature (78.1 %) au profit d'une représentation un peu plus élevée des chaînes à 2 EI (19.7 %). La variation reste néanmoins marginale et ne met nullement en cause la nette prévalence des chaînes à 1 EI, suivies, de loin, de celles à 2 EI.

3.2.2 Combinaisons des EI dans les chaînes

Nous envisagerons principalement les chaînes de 2 ou 3 EI, celles de 4 ou 5 EI étant trop peu fréquentes (4 occurrences en

¹³ Nous parlerons de combinaison pour désigner un ensemble d'EI sans tenir compte de l'ordre des éléments, et à l'inverse de chaîne pour désigner une structure en contexte, en tenant compte de l'ordre des EI.

tout) pour pouvoir dégager de véritables tendances¹⁴. Pour les chaînes avec un seul EI, dans la mesure où elles représentent la très grande majorité de l'ensemble des chaînes (83.2 %), la répartition des différents types d'EI que l'on y observe rejoint la fréquence globale des différents types d'EI, présentée dans le tableau 4 : les EI logico-pragmatiques sont les plus fréquents, suivis des éléments spatio-temporels et des relations logiques, puis des éléments spatiaux abstraits et des commentaires.

Pour les exemples à un seul EI, nous avons observé la répartition des catégories d'EI selon la nature du sujet. Pour cela nous avons utilisé le calcul du khi2, qui a révélé l'existence d'une relation partielle entre les deux. Le khi2 est en effet de 194.52, et la probabilité qui lui est associée est de 5,21659E-26, soit largement inférieure à 0.05 : on peut donc rejeter l'hypothèse nulle et considérer la distribution de la catégorie d'EI au regard de la nature du sujet comme significative. En observant plus précisément les données, nous notons que les EI Relation logique sont en proportion moindre avant les Sn (ils sont moins nombreux qu'attendus) et inversement plus nombreux qu'attendus devant les pronoms (personnels et impersonnels). A l'in-

¹⁴ Les chaînes de 4 EI sont les suivantes :

[logico-pragmatique] [commentaire] [relation logique] [relation logique]
(Hist)

[logico-pragmatique] [relation logique] [relation logique] [logico-pragmatique] (Eco)

[logico-pragmatique] [relation logique] [point de vue] [complément notionnel] (Eco)

On notera que toutes commencent par un EI logico-pragmatique, et que 2 d'entre elles ont en 2^e position un élément dénotant une relation logique :

[Cependant], [même si les taux d'intérêt pratiqués par la BCE sont trop bas], [de ce point de vue], [pour les PECO], on peut mettre en avant le gain de crédibilité (Debrun, 2001) ou l'accroissement du coût de l'inflation (Rogoff, 1985) qu'apporterait l'entrée dans l'Union économique et monétaire (UEM). (Eco).

La chaîne avec 5 EI a ceci de remarquable qu'il s'agit de 5 éléments détachés :

[la structure du peuplement],[les conditions d'intégration au marché économique national],[les relations au sein du monde paysan], [les rapports avec les autres catégories socioprofessionnelles], [le degré d'instruction], tout cela influe globalement sur le mode d'entrée en politique des ruraux d'entre elles. (Hist)

verse, on observe une attraction entre les EI logico-pragmatiques et les Sn, et une certaine répulsion entre ces mêmes EI et les pronoms. Cela est vraisemblablement lié à la valeur formelle et fonctionnelle de chacune de ces 2 unités (voir à ce propos les remarques à l'issue de la présentation des catégories d'EI en 2.2.2, ainsi que celles en 3.1.1 sur la relation entre nature du sujet et longueur des EI).

Pour ce qui est des chaînes de 2 ou 3 EI, nous nous attacherons aux points suivants : quel est le nombre de combinaisons d'EI réalisées par rapport à l'ensemble des combinaisons possibles ? Quelle est la mobilité de chacun des éléments dans les chaînes constituées ? Chacun des EI a-t-il une position de prédilection ? Plus précisément, privilégie-t-on certains EI en début de chaîne, en tête de phrase donc, ou bien au contraire en fin de chaîne, au plus près du sujet et de son verbe ? Certains EI peuvent-ils occuper toutes les positions dans les chaînes ?

3.2.2.1 Chaînes à 2 EI

Dans les chaînes à 2 EI, si l'on ne tient pas compte de l'ordre, on recense 44 combinaisons des 11 types d'EI, sachant que 2 EI du même type peuvent se combiner ensemble. Ce nombre représente 66.7 % des 66 combinaisons possibles : 22 combinaisons ne sont donc pas attestées¹⁵.

La distribution des 44 combinaisons parmi les 457 binômes attestés est très inégale : ainsi seulement 5 combinaisons

¹⁵ [commentaire] [commentaire] ; [commentaire] [complément de manière] ; [commentaire] [complément essentiel] ; [complément de manière] [complément de manière] ; [complément de manière] [complément essentiel] ; [complément de manière] [complément notionnel] ; [complément de manière] [point de vue] ; [complément de manière] [prédication seconde] ; [complément de manière] [spatial abstrait] ; [complément essentiel] [complément notionnel] ; [complément essentiel] [point de vue] ; [complément essentiel] [prédication seconde] ; [complément essentiel] [relation logique] ; [complément essentiel] [spatial abstrait] ; [complément essentiel] [élément détaché] ; [complément notionnel] [complément notionnel] ; [complément notionnel] [point de vue] ; [complément notionnel] [spatial temporel] ; [complément notionnel] [élément détaché] ; [point de vue] [point de vue] ; [point de vue] [élément détaché] ; [spatial abstrait] [élément détaché]

accaparent plus de la moitié de l'ensemble des occurrences. Il s'agit des combinaisons suivantes :

[logico-pragmatique] [relation logique] :	21.4 % (98 occ)
[logico-pragmatique] [spatial temporel] :	17.9 % (82 occ)
[spatial temporel] [spatial temporel] :	9 % (41 occ)
[logico-pragmatique] [spatial abstrait] :	7.9 % (36 occ)
[relation logique] [spatial temporel] :	5.9 % (27 occ)

Dans un ordre qui peut varier légèrement, et à quelques exceptions près, on retrouve en tête ces 5 combinaisons au sein des différentes disciplines.

La prévalence de la combinaison d'un EI logico-pragmatique avec une relation logique (21.4 %) n'est pas étonnante dans des articles scientifiques, dans lesquels argumentation et structuration logique sont supposées occuper une place importante.

Les 3 EI qui entrent dans le nombre le moins élevé de combinaisons sont ceux qui sont, d'une manière générale, les moins fréquents : il s'agit des compléments essentiels, des compléments de manière et des compléments notionnels, qui ne se combinent qu'avec un nombre assez restreint d'EI.

On notera par ailleurs que, parmi ces 3 éléments, 2 ne se combinent pas avec eux-mêmes : les compléments notionnels et les compléments de manière, auxquels il faut ajouter les commentaires. On peut penser, au moins pour les deux derniers, qu'il s'agit d'un « hasard » du corpus analysé : la succession de deux compléments de manière ou de deux commentaires semble tout à fait possible. En revanche, on peut concevoir que la combinaison de deux compléments notionnels, par leur contenu même, soit, sinon exclue, en tout cas plus difficile.

A l'inverse, les éléments logico-pragmatiques se combinent avec l'ensemble des autres éléments, ce qui est en partie lié au fait qu'il s'agit des éléments les plus fréquents.

Si l'on tient compte de l'ordre des 2 EI dans les chaînes, on recense 60 chaînes attestées (sur 121 possibles, soit 49.6 %).

Éléments initiaux : combinaisons et schémas préférentiels

En effet, pour 28 des 44 combinaisons précédemment considérées, les 2 EI n'apparaissent que dans un seul ordre¹⁶, tandis que pour 16 autres les 2 EI peuvent se combiner indifféremment¹⁷.

On retrouve en tête, avec une fréquence supérieure à 5 %, les mêmes associations que pour les combinaisons non ordonnées.

Toutefois, pour chacun des binômes, l'un des deux ordres s'impose nettement : ainsi la chaîne [logico-pragmatique] [relation logique] apparaît 94 fois :

- (25) [Ainsi,] [de même qu'on offrait de nombreux présents lors de cérémonies périodiques afin d'entretenir les alliances,] *on* en distribuait également autant au départ des partis de guerre. (Hist)

alors que la séquence inverse [relation logique] [logico-pragmatique] ne se rencontre que 4 fois :

- (26) [Si les étapes de la promulgation de cette loi sont désormais bien connues,] [en revanche] *on* a moins étudié les réactions qu'elle suscita dans le pays. (Hist)

La chaîne [logico-pragmatique] [spatial temporel] apparaît 73 fois :

- (27) [Or,] [en période de conflit armé,] *les garanties* sont d'autant plus importantes qu'elles sont difficiles à appliquer. (Droit)

mais la séquence inverse seulement 9 fois :

- (28) [Jusqu'à la fin de 1849,] [toutefois,] *il* s'agit davantage de « circonvenir » le suffrage que de le remettre en cause de manière radicale. (Hist)

La chaîne [logico-pragmatique] [spatial abstrait] se rencontre 32 fois :

- (29) [En effet,] [dans ce cas,] *la valeur privée des sillons pour les opérateurs* reflète leur valeur sociale, et il convient donc

¹⁶ Sont comptabilisées parmi elles les 8 combinaisons de 2 EI identiques.

¹⁷ Ces 32 chaînes correspondent à 323 occurrences, soit 70.7 % des 457 binômes.

d'attribuer les sillons aux opérateurs qui en ont la plus forte valuation. (Eco)

mais la séquence [spatial abstrait] [logico-pragmatique] seulement 4 fois :

- (30) [Dans le cas de la République tchèque et de la Hongrie,] [au contraire,] *les investissements directs* financent le déficit courant. (Eco)

De même, on relève 14 occurrences de la chaîne [logico-pragmatique] [commentaire] :

- (31) [Mais,] [de toute évidence,] *les remontrances royales* n'eurent pas l'effet escompté sur la politique coloniale pendant l'administration de Frontenac. (Hist)

mais seulement 1 occurrence pour la séquence inverse :

- (32) [Peut-être][en effet] *ce que je cherchais en Afrique* était-il la capacité à sortir de moi-même et à m'imbiber d'une culture extérieure assez profondément pour la vivre de l'intérieur. (Litt)

Cette disparité dans l'agencement des binômes invite à considérer le rang préférentiel que peuvent occuper les différents éléments. Le tableau 6 ci-dessous indique les fréquences respectives en rang 1 ou en rang 2 de chaque type d'EI.

Tableau 6. Fréquence en rang 1 ou rang 2 des EI dans les binômes

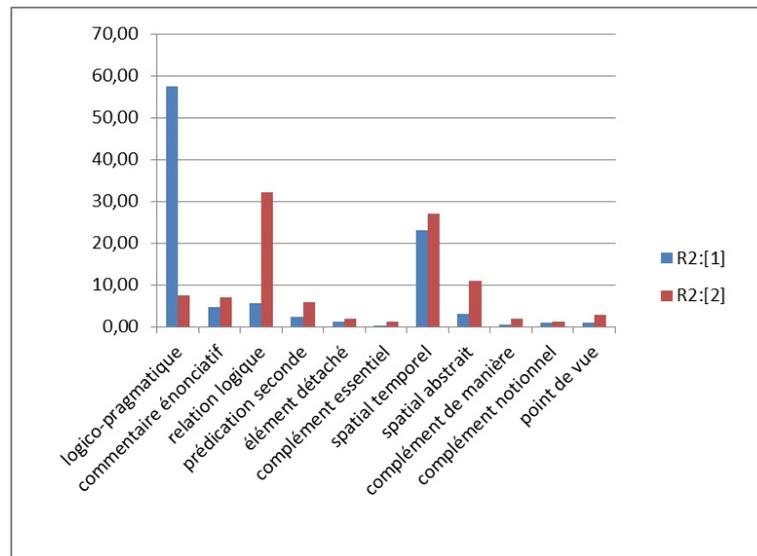
	Rang 1	Rang 2	Total
logico-pragmatique	263 (88.6)	34 <i>(11.4)</i>	297
spatio-temporel	105 <i>(45.9)</i>	124 (54.1)	229
relation logique	26 <i>(15)</i>	147 (85)	173
spatial abstrait	14 <i>(21.9)</i>	50 (78.1)	64
comment. énonciatif	21 <i>(39.6)</i>	32 (60.4)	53
prédication seconde	11 <i>(28.9)</i>	27 (71.1)	38
point de vue	4 <i>(23.5)</i>	13 (76.5)	17
élément détaché	6 <i>(40)</i>	9 (60)	15
complément manière	2 <i>(18.2)</i>	9 (81.8)	11

Eléments initiaux : combinaisons et schémas préférentiels

compl. notionnel	4 (40)	6 (60)	10
complément essentiel	1 (14.3)	6 (85.7)	7
Total	457	457	914

Il ressort du tableau 6 une opposition nette entre les EI logico-pragmatiques, qui occupent de manière écrasante le rang 1 (88.6 %), et les autres EI qui privilégient au contraire le rang 2. L'écart est assez marqué pour la majorité des EI (rapport de 1 à 3 ou 4) à l'exception des EI spatio-temporels, pour lesquels la distribution sur les rangs 1 et 2 est un peu moins tranchée. La relative fréquence des séquences de 2 éléments spatio-temporels (41 occ) peut en partie expliquer le relatif équilibre entre les rangs 1 et 2.

Le graphique 2 ci-dessous complète le tableau 6 en présentant la répartition sur les rangs 1 et 2 des différents EI.



Graphique 2. Distribution des EI sur les rangs 1 et 2

La fréquence très élevée en général des EI logico-pragmatiques et dans une moindre mesure celle des éléments spatio-temporels, la nette prévalence des premiers en 1^{re} position et la part non négligeable des seconds en cette même

position : ces données expliquent que le rang 1 est occupé dans 80.5 % des cas par l'un de ces deux EI.

Les relations logiques, les EI spatio-temporels et spatiaux abstraits représentent en revanche plus de 59.2 % des EI en seconde position.

Nous avons utilisé le calcul du khi2 pour confirmer l'existence d'une relation entre le type d'EI et son rang. Le khideux est de 305.84, et la probabilité qui lui est associée est de 9,06995E-60, soit largement inférieure à 0.05. On peut donc écarter l'hypothèse nulle et considérer la distribution des types d'EI au regard de leur rang comme significative, avec en particulier une forte attraction entre la première position et les éléments logico-pragmatiques.

3.2.2.2 Chaînes à 3 EI

On ne proposera pas pour les combinaisons de 3 EI une description aussi détaillée que pour celles de 2 EI, en raison de l'éparpillement des combinaisons actualisées dans les 38 occurrences de trinômes, et de la multitude des combinaisons non attestées.

Sur les 286 combinaisons non ordonnées possibles, 23 sont actualisées, soit 8 %. Une seule donne lieu à des chaînes différentes.

C'est la combinaison des EI [logico-pragmatique] [spatial temporel] [relation logique] qui présente la plus grande diversité, puisqu'elle se réalise sous la forme de 4 chaînes (8 occurrences). Dans aucune de ces 4 chaînes on ne trouve de relation logique en position 1, comme l'illustrent les exemples ci-dessous :

[logico-pragmatique] [spatial temporel] [relation logique] :

- (33) [Mais,] [en même temps,] [du fait même de leurs tribulations,]
les héros traversent tous les continents et tous les milieux.
(Litt)

[logico-pragmatique] [relation logique] [spatial temporel] :

- (34) L'exil qui ne peut en effet être perçu que comme passager devient en quelque sorte un espace d'attente. Le temps est ainsi par défaut ou en creux une donnée essentielle, [car] [si l'espace de l'exil peut être un jour effacé par la grâce ou la

Eléments initiaux : combinaisons et schémas préférentiels

chute du régime proscripteur,] [jamais] *son temps* ne peut être aboli. (Hist)

[spatial temporel] [logico-pragmatique] [relation logique] :

- (35) [Au début de l'année 1905] [pourtant,] [malgré des réserves,] *la majorité des évêques, les notables laïcs, les catholiques libéraux* penchaient pour l'acceptation des associations culturelles, comme en témoignent les nombreux articles que les Semaines religieuses consacrent à leur analyse. (Hist)

Sur les 1 331 combinaisons ordonnées possibles, seulement 26 chaînes sont donc actualisées (2 %). Parmi elles, 18 ne le sont qu'une seule fois, tandis que 4 chaînes le sont 3 fois, et que 4 le sont 2 fois : les trinômes attestés sont donc éparpillés dans un nombre relativement élevé de chaînes.

Pour ce qui est de la présence des différents EI dans les chaînes, tous sauf les compléments notionnels sont attestés, mais dans des proportions très variables. Ainsi les EI logico-pragmatiques sont présents dans 18 des 26 chaînes, les relations logiques dans 14 chaînes, les éléments spatio-temporels dans 12 chaînes. A l'inverse, les EI Point de vue, Commentaire et les compléments de manière n'apparaissent que dans 2 chaînes.

Par ailleurs, 4 des 10 EI présents dans les trinômes peuvent se combiner entre eux. On trouve ainsi des chaînes de 3 éléments spatiaux-temporels, de 3 éléments détachés, de 3 relations logiques ou de 3 prédications secondes. Chacune de ces chaînes n'est cependant instanciée qu'une seule fois, hormis pour les EI spatiaux-temporels (3 occurrences) :

[spatial temporel] [spatial temporel] [spatial temporel]

- (36) [Plus tard,] [lorsque je repris avec lui le travail dans les années 90] [après quelques années d'interruption,] *je commençai* par lui lire des passages de ses interventions précédentes désormais transcrites, en lui demandant des précisions ou des compléments. (Litt)

[élément détaché] [élément détaché] [élément détaché]

- (37) « [Nous séparer de l'espace et du temps,] [rompre nos attaches dans le passé et dans l'avenir,] [nous arracher brusquement à

nos occupations de chaque jour, à notre société, à nos proches,] c'est nous faire respirer le vide, c'est nous condamner à une mort pleine de lenteurs ». (Hist)

[prédication seconde] [prédication seconde] [prédication seconde] :

- (38) [Réclamée par la Préfecture de police,] [promulguée par Louis-Philippe] [et rendue exécutoire par Georges Delessert, préfet de police,] *elle* avait le mérite d'offrir un texte valable sur toutes les lignes et pour toutes les compagnies, mettant fin à la confusion des règlements sans cesse recopiés et insatisfaisants. Elle donnait force légale à l'action de police. (Hist)

Considérons maintenant le rang des différents EI dans les trinômes.

Tableau 7. Fréquence de la position 1 ou 2 ou 3 des EI dans les trinômes

	Rang 1	Rang 2	Rang 3	total
logico-pragmat.	26 (81.2)	2 (6.3)	4 (12.5)	32
spatio-temporel	7 (24.1)	13 (44.9)	9 (31)	29
relation logique	1 (3.8)	14 (53.9)	11 (42.3)	26
prédication seconde	1 (12.5)	3 (37.5)	4 (50)	8
élément détaché	1 (20)	2 (40)	2 (40)	5
comm. énonciatif	1 (25)	0	3 (75)	4
spatial abstrait	0	1 (33.3)	2 (66.7)	3
point de vue	1 (33.3)	2 (66.7)	0	3
compl. manière	0	0	2 (100)	3
compl. essentiel	0	1 (50)	1 (50)	2
Total	38	38	38	114

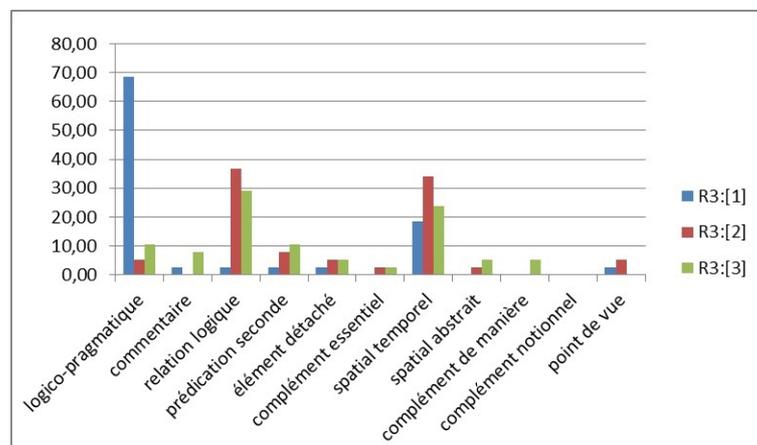
La rareté des effectifs de la majorité des EI ne permet de tirer de conclusions que pour quelques-uns d'entre eux. On notera ainsi, à nouveau, la tendance nette des EI logico-pragmatiques à privilégier la première position, tandis que les éléments spatiaux-temporels connaissent une distribution plus

Eléments initiaux : combinaisons et schémas préférentiels

homogène entre les 3 positions, avec néanmoins une prédilection pour le deuxième rang. Les relations logiques apparaissent quant à elles prioritairement en deuxième et troisième positions.

On notera par ailleurs que, alors que certains EI peuvent apparaître en toute position, dans des proportions variables, d'autres n'apparaissent que dans 2 positions. C'est le cas des éléments spatiaux abstraits, des compléments essentiels, des EI point de vue et des commentaires. Quant aux compléments de manière, ils n'apparaissent qu'en rang 3. Rappelons néanmoins que les effectifs de tous ces EI sont peu élevés.

Le graphique ci-dessous complète le tableau 7 en présentant la répartition sur les rangs 1, 2 et 3 des différents EI.



Graphique 3. Distribution des EI sur les rangs 1, 2 et 3

4. Conclusion et perspectives

A l'issue de cette analyse, différents points méritent d'être soulignés. Tout d'abord, comme nous l'avons noté à différentes reprises, le corpus se caractérise par une relative homogénéité entre disciplines, même si des divergences ponctuelles apparaissent. D'une manière générale, deux configurations s'imposent : celles avec un sujet initial (59.7 %), suivies de celles avec un seul EI : 34.16 % de l'ensemble des 7 531 séquences, et 83.7 % des séquences avec EI. Par ailleurs, 3 EI dominent par leur fréquence : les éléments logico-pragmatiques, suivis des

éléments spatiaux-temporels et des relations logiques. La prévalence des premiers et des troisièmes trouve sans doute une explication dans la spécificité des articles, qui relèvent tous du genre « article scientifique ». On notera en outre que ces trois types d'éléments apparaissent en toute position, même si les éléments logico-pragmatiques se rencontrent de manière privilégiée en première position. En effet, que ces derniers soient seuls devant le sujet ou bien associés à d'autres EI, ils occupent la première position dans 96.4 % du total de leurs occurrences (1107/1148). Nous avons observé les 41 occurrences avec un EI logico-pragmatique en seconde ou en troisième position. Dans la moitié des cas, un autre EI de ce même type occupe la première position : il ne reste donc que 21 cas (19 binômes et 2 trinômes) dans lesquels aucun EI logico-pragmatique présent n'occupe la première position. On notera que plus de la moitié de ces 21 EI correspondent à des expressions marquant une opposition (*en revanche, au contraire, cependant, pourtant, toutefois* : 11 cas).

Dans 9 des 19 binômes, la première position est occupée par un EI spatio-temporel (ex. 28), dans 4 elle l'est par un EI spatial abstrait (ex. 30), dans 4 autres par une relation logique (ex. 26), et dans 2 cas par un EI notionnel ou un commentaire (ex. 32). Dans les 2 trinômes concernés, on trouve devant l'EI logico-pragmatique un EI spatio-temporel seul ou accompagné d'une relation logique. Cela signifie que les types d'EI susceptibles de se placer devant un EI logico-pragmatique sont assez limités, et que, parmi eux, les éléments spatio-temporels ou spatiaux abstraits dominent nettement (ce qui est sans doute en partie dû à leur fréquence élevée d'une manière générale). On signalera par ailleurs que ce sont dans les textes historiques ou d'économie que l'on rencontre le plus de configurations de ce type (respectivement 6 et 11 occurrences contre 2 en droit et 2 en littérature).

Par ailleurs, la position initiale, quand elle n'est pas occupée par le sujet, l'est par un élément de ce type dans 36 % des cas. Si l'on ne considère que les séquences de 2 EI ou plus, cette fréquence augmente, la position initiale étant alors occupée par un élément logico-pragmatique dans 58.5 % des cas. Cette

présence massive des EI logico-pragmatiques en tête de chaîne trouve une explication probable dans leur dimension fortement connectrice. Il serait assurément intéressant de considérer de plus près les cas dans lesquels ces éléments occupent une position plus reculée dans la phrase.

Bien des aspects ont été laissés ici, du fait de la masse des données à traiter et de l'espace imparti à cette étude. Il conviendrait ainsi, en approfondissant l'analyse des données examinées, de regarder de plus près la nature de certains EI (Sn, proposition subordonnée...), et, évidemment d'étudier de manière plus fine le fonctionnement des différents EI (les études ici-même de S. Carter-Thomas et de A. Celle et L. Lansari s'attachent précisément à deux types d'EI : les commentaires et les prédications secondes, et ce dans une perspective contrastive anglais-français). Il serait par ailleurs souhaitable de mettre en regard l'analyse menée avec des études portant sur d'autres genres de textes. C'est précisément l'objectif que s'est fixé E. Havu ici-même (« L'emploi d'éléments initiaux dans quatre romans français contemporains. Comparaison avec un corpus de textes d'économie »), son étude permettant une comparaison entre différents romans à partir d'une même classification des EI.

Au-delà de l'étude ici menée, d'autres aspects mériteraient d'être explorés. Il serait ainsi intéressant d'observer quels éléments, parmi ceux que l'on rencontre en position préverbale, peuvent aussi occuper une position après le sujet et/ou le verbe (voir dans cette perspective Hannay 2007, et Virtanen 2004¹⁸). Cette piste nous semble prometteuse en particulier pour les éléments logico-pragmatiques et les commentaires, ainsi que pour les éléments spatiaux-temporels.

Il conviendrait enfin de dépasser la dimension strictement morpho-syntaxique et sémantique qui a prévalu ici, et de placer notre approche dans une perspective plus large, en

¹⁸ « Once 'professionalized', connectors and adverbials indicating manner of speaking, beliefs, attitudes and the like, can, for various reasons, even yield the very first position to other elements and still keep their discourse organizing or affect-marking status. » (Virtanen 2004 : 95).

particulier en envisageant les dimensions cognitives, informationnelle et textuelle de la structuration de la zone préverbale.

Si la zone préverbale joue un rôle essentiel, vers l'amont, pour l'instauration et le maintien de la cohésion et de la cohérence textuelle, elle remplit aussi une fonction organisatrice évidente, vers l'aval textuel (voir entre autres Virtanen 2004, et Ho-Dac (2007 : 18), qui parle de position « pivot »). Il serait néanmoins réducteur de n'envisager que ces deux fonctions : on sait que cette zone est à même, en raison du fort potentiel stratégique et cognitif qu'elle renferme, d'accueillir des éléments aux fonctions variées, y compris, sous certaines conditions, porteurs d'informations nouvelles (*Même pas huit ans elle avait*, cf. (Abeillé *et al.* 2008)).

C'est plus spécifiquement vers l'exploration des statuts cognitif et informationnel¹⁹ qu'il nous semble intéressant de poursuivre la recherche, afin de voir si la forte présence d'éléments connecteurs, que nous avons pu observer dans notre corpus, est renforcée par celle d'éléments cognitivement actifs ou accessibles²⁰, et peu informationnels, ou si elle permet au contraire plus facilement le positionnement d'éléments inactifs, éventuellement porteurs d'information nouvelle.

Références

- Abeillé A., Godard D. & Sabio F. (2008). « Deux constructions à Sn antéposé en français », in J. Durand, B. Habert, B. Laks (éd.) *Congrès Mondial de Linguistique Française - CMLF'08*, ISBN 978-2-7598-0358-3. Paris : Institut de Linguistique Française.
- Ariel M. (1990). *Accessing Noun-Phrase Antecedents*. London : Routledge.
- Benincà P. (2006). « A detailed map of the left periphery of Medieval Romance », in R. Zanuttini, H. Campos,

¹⁹ Pour une distinction limpide entre les deux concepts, voir (Lambrecht 1994). Pour une présentation détaillée des échelles d'accessibilité, voir (Prince 1981) et (Ariel 1990).

²⁰ On gardera à l'esprit que la notion de statut cognitif, en toute rigueur, n'est pertinente que pour les éléments référentiels.

- E. Herburger, P. H. Portner (éd.) *Georgetown Round Table on Languages and Linguistics (2004). Crosslinguistic Research in Syntax and Semantics : Negation, Tense and Clausal Architecture*. Washington, D.C. : Georgetown University Press, 53-86.
- Charolles M. (1997). « L'encadrement du discours : Univers, Champs, Domaines et Espaces », *Cahier de Recherche Linguistique, LANDISCO* 6 : 1-73.
- Charolles M. & Prévost S. (éd.) (2003). *Travaux de linguistique* « Adverbiaux et topiques » 47.
- Fløttum K., Dahl T. & Kinn T. (2006). *Academic Voices Across Languages and Disciplines*. Collection *Pragmatics and Beyond New Series*. Amsterdam/Philadelphia : John Benjamins.
- Halliday M.A.K. (1994). *An introduction to functional grammar*, 2nd éd. Londres : Edward Arnold.
- Hannay M. (2007). « Patterns of multiple theme and their role in developing English writing skills », in C. Butler, R. Hidalgo, J. Lavid (éd.), *Functional perspectives on grammar and discourse*. Amsterdam & Philadelphia PA : Benjamins, 257-278.
- Ho-Dac L-M. (2007). *La position initiale dans l'organisation dans l'organisation du discours : une exploration en corpus*. Thèse de doctorat. Toulouse : Université Toulouse le Mirail.
- Lambrecht K. (1994). *Information structure and sentence form : Topic, focus, and the mental representation of discourse referents*. Cambridge : Cambridge University Press.
- Landragin F., Poibeau T. & Victorri B. (2012). « ANALEC : a New Tool for the Dynamic Annotation of Textual Data », *Eighth International Conference on Language Resources and Evaluation*. Istanbul, 357-362.
- Nølke H. (1993). *Le regard du locuteur. Pour une linguistique des traces énonciatives*. Paris : Editions Kimé.
- Prince E.F. (1981). « Toward a taxonomy of given/new information », in P. Cole (éd.) *Radical pragmatics*. NY : Academic Press, 223-255.

- Rainsford T., Guillot C., Lavrentiev A. & Prévost S. (2012). « La position préverbale en ancien français : apport de corpus annotés », in F. Neveu, V. Muni Toke, P. Blumenthal, T. Klingler, P. Ligas, S. Prévost, S. Teston-Bonnard (éd.) *Actes du 3^e Congrès Mondial de Linguistique Française*, 159-176 [publication en ligne, DOI : <http://dx.doi.org/10.1051/shsconf/20120100246>].
- Sarda L & Carter-Thomas S. (2009). « Rôle informationnel et organisationnel des adverbiaux détachés en tête de phrase : étude contrastive français / anglais », in L. Florea, C. Papahagi, L. Pop, A. Curea (éd.). *Directions actuelles en linguistique du texte*. Cluj : Casa Cartii de Stiinta, 311-328.
- Smits A. (2002). *How writers begin their sentences. Complex beginnings in native and learner English*. PhD, University of Amsterdam.
- Virtanen T. (2004). « Point of departure : Cognitive aspects of sentence-initial adverbials », in T. Virtanen (éd.), *Approaches to cognition through text and discourse*, Berlin / New York : Mouton de Gruyter, 79-97.
- Widlöcher A. & Mathet Y. (2009). « La plate-forme Glozz : environnement d'annotation et d'exploration de corpus », *16^e Conférence sur le Traitement Automatique des Langues Naturelles*, Senlis.