



HAL
open science

An empirical approach to the signalling of enumerative structures

Lydia-Mai Ho-Dac, Cécile Fabre, Marie-Paule Péry-Woodley, Josette Rebeyrolle, Ludovic Tanguy

► **To cite this version:**

Lydia-Mai Ho-Dac, Cécile Fabre, Marie-Paule Péry-Woodley, Josette Rebeyrolle, Ludovic Tanguy. An empirical approach to the signalling of enumerative structures. *Discours - Revue de linguistique, psycholinguistique et informatique*, 2012, 10, (publication en ligne). halshs-00954182

HAL Id: halshs-00954182

<https://shs.hal.science/halshs-00954182>

Submitted on 28 Feb 2014

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.

Lydia-Mai Ho-Dac, Cécile Fabre, Marie-Paule Péry-Woodley, Josette Rebeyrolle et Ludovic Tanguy

An Empirical Approach to the Signalling of Enumerative Structures

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

Lydia-Mai Ho-Dac, Cécile Fabre, Marie-Paule Péry-Woodley, Josette Rebeyrolle et Ludovic Tanguy, « An Empirical Approach to the Signalling of Enumerative Structures », *Discours* [En ligne], 10 | 2012, mis en ligne le 16 juillet 2012, consulté le 13 octobre 2012. URL : <http://discours.revues.org/8611> ; DOI : 10.4000/discours.8611

Éditeur : Laboratoire LATTICE, UMR 8094 ENS/CNRS

<http://discours.revues.org>

<http://www.revues.org>

Document accessible en ligne sur : <http://discours.revues.org/8611>

Ce document PDF a été généré par la revue.

© Discours - Presses universitaires de Caen



Revue de linguistique, psycholinguistique et informatique

<http://discours.revues.org/>

An Empirical Approach to the Signalling of Enumerative Structures

Lydia-Mai Ho-Dac

CLLE-ERSS, CNRS & Université de Toulouse (UTM)

Cécile Fabre

CLLE-ERSS, CNRS & Université de Toulouse (UTM)

Marie-Paule Péry-Woodley

CLLE-ERSS, CNRS & Université de Toulouse (UTM)

Josette Rebeyrolle

CLLE-ERSS, CNRS & Université de Toulouse (UTM)

Ludovic Tanguy

CLLE-ERSS, CNRS & Université de Toulouse (UTM)

.....
Lydia-Mai Ho-Dac, Cécile Fabre, Marie-Paule Péry-Woodley, Josette Rebeyrolle et Ludovic Tanguy, «An Empirical Approach to the Signalling of Enumerative Structures», *Discours* [En ligne], 10 | 2012, mis en ligne le 16 juillet 2012.

.....
URL : <http://discours.revues.org/8611>

.....
Titre du numéro : *Multidisciplinary Perspectives on Signalling Text Organisation*

Coordination : Shirley Carter-Thomas & Frédéric Landragin

revues.org
CENTRE POUR L'ÉDITION ÉLECTRONIQUE OUVERTE
CENTRE FOR OPEN ELECTRONIC PUBLISHING

 discours

 Presses
universitaires
de Caen

An Empirical Approach to the Signalling of Enumerative Structures

Lydia-Mai Ho-Dac

CLLE-ERSS

CNRS & Université de Toulouse (UTM)

Cécile Fabre

CLLE-ERSS

CNRS & Université de Toulouse (UTM)

Marie-Paule Péry-Woodley

CLLE-ERSS

CNRS & Université de Toulouse (UTM)

Josette Rebeyrolle

CLLE-ERSS

CNRS & Université de Toulouse (UTM)

Ludovic Tanguy

CLLE-ERSS

CNRS & Université de Toulouse (UTM)

This paper presents a data-intensive study of the signalling of enumerative structures. In contrast with semasiological studies of specific markers, the approach described here takes as its starting point annotated structures and cues, seeking to identify recurrent patterns in these data. To do so, it exploits a new resource for French, the ANNODIS resource, a large corpus of written texts manually annotated at discourse level. The data analysed —first quantitatively with large populations, then qualitatively on selected examples— allows the authors to illustrate how cues involved in signalling text organisation combine in complex ways metadiscourse and propositional content, or the textual and ideational metafunctions.

Keywords: discourse structures, enumerative structures, text organisation signals, corpus linguistics, discourse annotation

Nous présentons dans cet article une analyse extensive sur corpus de la signalisation des structures énumératives. Notre étude se distingue des travaux antérieurs principalement caractérisés par une approche sémasiologique de marqueurs spécifiques, car elle se fonde sur une annotation manuelle systématique des structures et des indices. C'est à partir de ces données annotées que nous cherchons des motifs récurrents de signalisation. Nous exploitons une ressource récemment créée, la ressource ANNODIS, corpus de français écrit enrichi d'annotations discursives. Les données analysées – de manière quantitative d'abord pour embrasser des effectifs importants, puis de manière qualitative sur une sélection d'exemples – nous permettent de montrer que les indices qui

contribuent à l'organisation textuelle peuvent associer métadiscours et contenu propositionnel, en d'autres termes les fonctions textuelle et idéationnelle.

Mots clés : *structures de discours, structures énumératives, signalisation de l'organisation textuelle, linguistique de corpus, annotation discursive de corpus*

1. Introduction

1 In its scene-setting initial sentence, the call for papers for this thematic issue states that “certain features or elements seem to have a special instructional role with regard to text organisation”. This observation leaves room for different conceptions of signalling devices: they may be construed as being separate from the text’s main concern —i.e. its propositional content—, and identifiable as specialised discrete signals; this is the view conveyed by terms such as signposting language, metatext, or textual metadiscourse. The opposite view rejects the idea of specialised dedicated devices, and considers that text organisation is achieved and signalled by a variety of multi-functional devices, which may also be involved with conveying content. Simplistic as it may be, this dichotomous presentation is a useful thinking tool because of the methodological counterparts of each endpoint: the first view is associated with semasiological approaches —studies based on specific “markers”—, while the second view gives rise to empirical studies attempting to identify the linguistic correlates of particular structures or functions.

2 The study described here is in a position to situate itself in this second current thanks to the recent availability of a corpus of written French texts manually annotated at discourse level, the ANNODIS resource. Our focus is on a very frequent and diversely realised organisational pattern: enumerative structures, a particularly interesting feature of which is that they are multi-level structures, i.e. they appear at different levels of granularity. In the next section, we look in more detail at the theoretical and methodological dichotomy just outlined, and review existing studies related to enumerative structures in the light of these opposing views. This leads to the formulation of a hypothesis concerning the nature of discourse markers, seen as complex markers composed of bundles of cues. The third section describes the ANNODIS resource and the data concerning enumerative structures. The fourth section is devoted to the signalling of enumerative structures in the ANNODIS resource, with a particular interest in how item cues may contribute to both text organisation and propositional content.

2. Text organisation signals

2.1. Some questions on “signalling” and “metadiscourse”

3 The notion of metadiscourse posits that there are separate things going on: discourse on the one hand and discourse about discourse on the other, with a further distinction

between interpersonal metadiscourse (hedges, certainty markers, attributors, attitude markers, and commentary) and textual metadiscourse (text markers and interpretive markers) (Crismore, Markkanen & Steffensen, 1993; Hyland, 1998; Hyland & Tse, 2004, *inter alia*). The notion has inspired a mass of studies in the field of academic writing, whilst continuously arousing controversy: Ifantidou (2005) for instance questions this view which divides metadiscourse from propositional content; in a recent study focusing on spoken academic language, Aguilar (2008: 108) admits that “it is at times difficult to draw a clear line between primary discourse and metadiscourse and between textual and interpersonal categories”. Hempel and Degand (2008) is a study of sequencers in different genres which presents an interesting case in point, and its focus is of specific interest to us in relation to enumerative structures. The authors squarely situate their contribution within studies of metadiscourse, adopting Mao’s definition: “it [metadiscourse] refers to various kinds of linguistic tokens that an author employs in her text to guide or direct the reader as to how to understand her, and her stance towards it” (Mao, 1993: 265). Having stated that “metadiscursive items are distinct from the propositional content with which they occur”, they further define textual metadiscourse: “[it] concerns the understanding of the ideational meaning and serves to organise the discourse by structuring the propositional content, by introducing sequences or by referring to the source of the propositional material” (Hempel & Degand, 2008: 679). Sequencers are defined as “those linguistic items introducing new sequences in a text” (Hempel & Degand, 2008: 681) and as such constitute one type of organisational metadiscourse markers¹. Three types are distinguished and illustrated as follows:

- spatial sequencers, e.g. *On the one hand... on the other hand*;
- temporal sequencers, e.g. *First... then... finally*;
- numerical sequencers, e.g. *Firstly... secondly...*

4 Although the mention of specific linguistic items is limited and they are characterised in terms of a shared function, the approach can still be seen as largely semasiological: its starting point is a set of lexical items called “sequencers”, which even if they constitute an open set can be identified so as to be examined in different genres. More significantly, the structure they are taken to signal is observationally secondary, i.e. a structure is observed where the markers defined as sequencers are found. There is a potential risk of circularity in this approach, which we will illustrate with one example.

5 Hempel and Degand root their characterisation of sequencers in Charolles’ work on discourse frames, yet in their lexically-oriented approach, sequencers are singled out as textual metadiscourse markers. Through the made-up examples in Figure 1, we seek to suggest that there is in fact a functional similarity, in terms of

1. The others being “topicalisers (elements introducing a new subject), illocution markers (elements indicating the illocutionary act(s) the author has been realising) and reviews/previews (elements anticipating or repeating a stage in the text)” (Hempel & Degand, 2008: 680).

text organisation, between the structures created by sequencers on the left and by “ordinary” temporal and spatial frame introducers on the right. The introductory phrase on the right (“...several major periods/zones”) states “the main federative idea with the help of a quantifier and a classifier” just as “several reasons” does on the left, and the temporal/spatial frame introducers on the right signal a sequence in a manner different but functionally similar to the sequencers on the left.

Sequencers	Frame introducers (time, space)
...several reasons: Firstly, ... Next, ... Finally,several major periods/zones: During the 1950s, ... / In most of Europe, ... In the 60s and 70s, ... / In the Middle East, ... From 1985, ... / Right across North America, ...

Figure 1. Functional similarity between sequencers and temporal/spatial frame introducers

6 There is however an important difference as regards the notion of metadiscourse: the temporal and spatial frame introducers on the right contribute to propositional content, they therefore cannot be seen as belonging to textual metadiscourse if it is defined as quite distinct from textual content. If we return to Charolles’ formulation of the framing hypothesis, it is that “preposed adverbials not only contribute to the propositional content of the sentence in which they occur, but assume specific organisational functions at the discourse level” (Charolles, 2005: 16). Charolles therefore a) puts propositional content first, and b) considers that contributing to propositional content and taking on organisational functions go hand in hand.

7 Our own work on spatial and temporal adverbials (Ho-Dac & Péry-Woodley, 2009), on discourse framing (Le Draoulec & Péry-Woodley, 2005), and on headings (Ho-Dac, Jacques & Rebeyrolle, 2004; Rebeyrolle, Jacques & Péry-Woodley, 2009) has led us to explore the ways in which a range of textual elements combine propositional content and organisational role. As do Hempel and Degand (2008), we root our approach in the Hallidayan formulation of three language components (called *metafunctions*: Halliday, 1977): interpersonal (interaction between speaker/writer and addressee); ideational (construing experience: participants, processes and circumstances); textual (presentation of ideational and interpersonal meaning as information in text unfolding in context: theme, cohesion). However the approach we wish to define here differs from theirs in three major ways:

- we start from the hypothesis, closer to Halliday’s view as we interpret it, that all three metafunctions are present in a text element, though one or other may be more prevalent; this hypothesis provides the basis for our account of some peculiarities of temporal framing in narratives (Le Draoulec & Péry-Woodley, 2005);
- we do not abstract out lexical expressions, i.e. we do not a priori expect a particular function to be linked to the lexical component of an expression

independently of features to do with its specific inscription in a text, such as textual position; the text organisational potential of expressions lies in their positional and layout properties as well as in the words used²;

- we take into account non-lexical text elements, such as typographical and layout elements (bullets, spaces, indentation...). As our primary interest is in text organisation, our interest in signals of sequencing for instance cannot restrict itself to lexical sequencers leaving out functionally comparable cues such as bullets or numbers³.

8 In such an approach to the signalling of text organisation, text organising structures are first characterised in structural and functional terms. The systematic study of the diverse signals that make them identifiable by readers comes in a second phase, in a dynamic approach whereby they are considered *in situ* and in their interactions, rather than as items in a lexicon. In this perspective, signalling should be thought of in terms of complex markers rather than discrete lexical elements. In the next section, we present enumerative structures, before moving on to the data-intensive method we have set up in order to study these complex markers.

2.2. An object of study: enumerative structures

9 We are interested in enumerating as a basic device in text organisation, which can be described as follows: enumerating is arranging text so that the reader becomes aware of this textual arrangement. The associated semantics is that the reader is led to interpret the enumerated elements as similar in some respect, and therefore as constituting a segment homogeneous in terms of this criterion (which may or may not be explicit). We see enumerating as a generic way of organising text in the sense that it can be resorted to for a wide range of semantic or rhetorical functions.

10 Enumerative structures have been the subject of numerous studies, focusing mostly on lexical markers of successive items. Two main groups of studies can be distinguished:

- those starting from specific lexical markers;
- those starting from document structure and typographical and layout signals.

11 Studies focusing on lexical signalling have mainly been concerned with the semantic-functional categorisation of item introducers; very few have examined

2. An important aspect of Charolles' proposal, and before him Thompson's (1985), is that they link the organisational potential of an expression to its textual position: "Initial and final purpose clauses in English are doing radically different jobs" (Thompson, 1985: 57). See Crompton (2006) for an opposing view, and our discussion in an earlier issue of this journal (Ho-Dac & Péry-Woodley, 2009).

3. The theoretical underpinning for considering "visual" cues on a par with lexical cues can be found in the model of text architecture (Luc et al., 1999; Luc et al., 2000). We also refer to the notion of document structure developed by Power, Scott and Bouayad-Agha (2003) in their work on text generation.

triggers and closures. Tadros (1994), however, in a study of “predictive categories in expository text”, describes Enumeration as “a category of Prediction in which the [predictive] member carries a signal that commits the writer to enumerate” (Tadros, 1994: 71). She looks specifically at the nouns used in triggers, for which she coins the term *enumerable* noun. Indeed a number of studies have focused on nouns characterised by their ability to enter into special constructions (*the fact that...*), or to function as cataphoric or anaphoric pro-forms: *classifiers* (Bouroufi & Vigouroux, 2003; Porhiel, 2007), *abstract, under-specified* or *shell* nouns (Francis, 1994; Legallois, 2006; Schmid, 2000). It is generally noted that when announcing enumerations these undefined meanings tend to be expressed by a plural NP preceded or not by some kind of numeral (exact, such as *two*, or inexact, such as *several*). Other studies confronted with such predictive elements (Schnedecker, 2006; Bras, Prévot & Vergez-Couret, 2008) define them in a more cognitive way as an underspecified meaning updated by the enumeration of items (cf. Bras, Prévot & Vergez-Couret, 2008: 1960). Symmetrically, closures call upon similar nominal expressions in anaphoric encapsulation (Alvarez de Mon y Rego, 2001; Conte, 1996; Francis, 1994).

12 The majority of studies have however been concerned with item introducers. Within the French context, many recent studies revisit earlier work (Turco & Coltier, 1988; Adam & Revaz, 1989) from the point of view of Charolles’ *discourse framing* hypothesis (Charolles, 1997; Charolles et al., 2005; Charolles & Vigier, 2005), viewing item introducers as discourse markers associated with an instructional meaning: they project forward an interpretation criterion, and thus define the initial boundary of a discourse frame i.e. a “package” of clauses clustering around a specific interpretation criterion (Jackiewicz, 2005; Jackiewicz & Minel, 2003, *inter alia*). As well as organisational framing adverbials such as *First, in addition, finally*, etc., some grammatical subjects can also function as item introducers (*The first X, a second Y*, etc.). Schnedecker (2006) and Porhiel (2007) add to this list expressions associated with “two-step enumerations”: adverbials such as *on the one hand, on the other hand*, but also grammatical subjects such as *the one, the other*. Laippala (2008) is one of the few authors to note that items are often not introduced by an explicit lexical marker and that a great diversity of textual elements can play the role of item introducer.

13 A second, less prolific, approach focuses on typographical and layout signals. They are rare, due to linguists’ blindness to the visual properties of written language, as noted by Virbel et al. (2005: 234, our translation): “Linguistics, like —to a lesser extent— information science, has tended to be blind to the role of visual properties in written language”⁴. For these authors, the visual properties of written language are not just formatting options but fully-fledged structural elements. In Virbel’s Model

4. The role of visual properties in text processing has on the contrary long been a preoccupation for language teachers and educational and cognitive psychologists (see Bessonnat, 1988; Stark, 1988; Passerault & Chesnet, 1991).

for Text Architecture, formatting choices signal textual acts and may be paraphrased by meta-sentences such as: “the writer indicates that there is a segmentation” for paragraph break, or “the writer indicates that this is the first item of an ordered list” for an ordered list. According to this view, the number *1*), the first bullet, or a lexical marker such as *First* or *A first step is* express the same textual act: “indicating that what follows constitutes the first item of a list”. Enumerations are the focus of several studies within this framework (Luc et al., 1999; Luc et al., 2000; Maurel, Lemarié & Vigouroux, 2003), where they are viewed as a textual object used for presenting elements as equal with respect to a particular criterion:

The textual act [of Enumerating] consists in transposing textually the co-enumerability of the listed entities into the co-enumerability of the linguistic segments describing them, which thereby become the entities constituting the enumeration (the items). The identity of status of the items in the enumeration expresses the identity of status of the listed entities in the world.
(Luc et al., 2000: 25; our translation)

14 In this definition, the relation between the listed entities or events is created by the textual act of enumerating, and is not dependent on a pre-existing (e.g. ontological) relation. The text architecture view of enumerating is part of the foundation for our definition of enumerative structures. Enumerative Structures (ESs) are segments of text characterised by an internal organisation involving several sub-segments:

- a **trigger** (optional): a segment announcing the enumeration;
- **several items**: segments composing the enumeration (at least two items must be identified for a structure to be present);
- a **closure** (optional): a segment which summarises and/or closes the enumeration.

15 The example in Figure 2 provides an illustration. The text segment delimited by the ES (a whole subsection) is internally organised as follows: first, the heading together with the opening paragraph announce that the following text will list four directions for action (re. the relationship between France and the Middle East); next, four bulleted items detail each of the directions, which are thereby presented as co-enumerable, i.e. identical in status with regard to the co-enumerability criterion; finally, the last paragraph of the subsection closes the enumeration with a conclusion.

16 The example in Figure 2 also serves to illustrate the diversity of cues contributing to signal an ES. Here we can count in the heading (“Des orientations d’actions”/*Directions for action*), the quantified expression at the end of the first paragraph (“autour de quatre thèmes”/*around four themes*), the bullets and syntactic parallelism at the start of each item, and the adverbial “En conclusion” (*In conclusion*) in the closure. Studying this diversity and examining the interaction between types of cues is a major aim of this study, based on the ANNODIS resource (Péry-Woodley et al., 2009), which is described next.

II) Des orientations d'action	ES	TRIGGER
Les orientations proposées peuvent être regroupées autour de quatre thèmes.		
- Mieux organiser notre politique étrangère dans la région ce qui passe, notamment, par la mise en place de structures permettant [...].		ITEM 1
- Accentuer notre coopération avec des partenaires d'influence, notamment en établissant une coopération renforcée avec certains [...].		ITEM 2
- Manifester notre souci de voir émerger des systèmes démocratiques dans la région en développant une politique d'influence auprès des [...].		ITEM 3
- Contribuer plus efficacement à la solution des principales crises régionales, ce qui comporterait les actions suivantes : [...].		ITEM 4
En conclusion, les turbulences qui affectent le moyen orient ont atteint un niveau de haute intensité qui représente, pour les pays occidentaux et, plus spécialement, pour l'Europe, de grands risques, notamment [...].		CLOSURE

Figure 2. An example of enumerative structure (ES) from the ANNODIS resource

3. Enumerative structures in the ANNODIS resource

17 The ANNODIS resource⁵ is a diversified corpus of written French texts enriched with several discourse-level markups. It is designed to provide richly annotated data for the study of discourse structures. The existence of such a resource is a necessary condition for the adoption of a data-intensive approach: starting from attested structures, and calling upon corpus-linguistics methods to address their diverse realisations and functions. Though the resource also comprises annotations of coherence relations and topical chains, we focus here solely on enumerative structures. The next section outlines the general characteristics of the resource in relation to these structures: corpus, annotation procedures and results.

3.1. A diversified corpus of written French texts

18 The composition of the corpus reflects our objectives and starting hypotheses. In particular, our focus on multi-level discourse structures interacting with document structure led us to favour long expository texts, which “in contrast to narratives, [...] do not structure themselves around a given organising schema. Rather, each such text constructs its own hierarchical ‘super-structure’, as a prerequisite for text coherence and wellformedness” (Katzenberger, 2005: 1). Length was also a criterion since longer texts cannot rely solely on thematic continuity but require the interaction of several forms of organisation. Third, the corpus had to be composed of texts in which crucial elements of discourse signalling such as subdivisions and layout are available and can be easily recovered. Finally, the corpus had to be diversified, including various types within the category of expository texts in order to allow the

5. The ANNODIS project was financed by the French National Research Agency's Humanities and Social Sciences Programme. Information on the ANNODIS project, including a number of related presentations and publications, can be found on w3.erss.univ-tlse2.fr/annodis. The ANNODIS resource, along with the other annotated corpora produced as part of the project, will be available by mid-2012 from REDAC (<http://redac.univ-tlse2.fr/corpus/>) and CNRTL (<http://www.cnrtl.fr/>). More detailed accounts of the principles underlying the annotation procedure and interface can be found in Péry-Woodley et al. (2011).

study of variation in the use and signalling of our structures. The composition of the ANNODIS corpus annotated with enumerative structures is described in Table 1.

Id	Source (S), genre (G), main type (T)	Document structure	Number of
WIK ₂	S = <i>Wikipedia.fr</i> G = encyclopedia T = expository	strong	texts: 30 words: 231,000 words/text: 7,700
LING	S = <i>CMLF, linguistics conference proceedings</i> G = academic T = expository	medium	texts: 25 words: 169,000 words/text: 6,760
GEOP	S = <i>IFRI, institut de géopolitique</i> G = reports and articles T = argumentative	medium	texts: 32 words: 266,000 words/text: 8,325
TOTAL: 156 texts; 687,000 words			

Table 1. Quantitative overview of the ANNODIS corpus annotated with Enumerative Structures

3.2. Annotating enumerative structures

19 A detailed annotation manual was produced to guide annotators⁶ in the task of identifying and marking both the discourse structures and the cues perceived as signalling them. In an approach based on Biber’s methodology for an emergent text-typology (Biber, 1988), the manual annotation process was preceded by the automatic pre-marking of a wide range of features. The typographical, layout and lexico-syntactic features pre-marked in this phase include:

- visual devices (headings, bulleted/numbered items);
- punctuation: paragraphs ending with [:], punctuational motifs such as [: ...; ...; and/or...];
- lexico-syntactic features based on the previous studies mentioned in 2.2: item introducers, predictive elements, anaphoric encapsulation, sentence-initial circumstantial adverbials (as potential frame introducers), other sentence-initial elements (e.g. connectives, coreferential expressions).

20 The pre-marking was carried out automatically by way of local grammars applied to POS-tagged and syntactically-analysed⁷ text, making use of specifically designed lexicons. Table 2 lists the categories of features automatically pre-marked in this way.

6. The annotations were performed by three 3rd-year linguistics students, nevertheless considered “naïve” insofar as they had no specific prior knowledge of the background of the annotation model.

7. TreeTagger was used for the POS-tagging and Syntex (Bourigault, 2007) for the syntactic parsing.

Feature type	Feature description	Examples
Item introducers	sequencers, connectives (sentence initial)	<i>first, at the same time, finally, the first X</i>
Circumstantial adverbials	PPs including spatial, temporal or notional lexemes	<i>Since 1956, In linguistics</i>
Typography, layout	indentation, line space, paragraph breaks, bullets, heading, punctuation	[: ...; ...; and/or...]
Prospective elements	NPs including specific lexemes	<i>around four themes, the following elements</i>
Encapsulations and concluding elements	pre-verbal demonstrative NPs with a numeral as determiner	<i>These three steps, In conclusion</i>

Table 2. Features automatically pre-marked for assisting ESs annotation

<p>En revanche, le régime a patronné trois formations importantes. Bien qu'il ait réduit de moitié les effectifs de la Garde Républicaine, passée de 150 000 à 70 000 hommes, il a veillé à en reconstituer les précieuses unités mécanisées et blindées. Pour ce faire il a eu recours, outre quelques importations illégales, à la cannibalisation des matériels rescapés du pilonnage, souvent au détriment de l'armée. Le régime s'est aussi détourné de son aviation au profit d'un Corps aérien plus opérationnel. Il en a consolidé les escadrons habitués à opérer en coordination étroite avec la Garde républicaine. L'importation de pièces de rechange s'est d'ailleurs révélée plus facile pour les hélicoptères, qui bénéficient d'un double statut civil et militaire. Enfin, les incursions quasi quotidiennes des avions anglo-saxons dans les zones d'exclusion aérienne et les "frappes" régulières de missiles de croisière ont stimulé l'intérêt porté par Saddam Hussein à la Défense aérienne, renouée et amadouée par des privilèges semblables à ceux dont bénéficie la Garde républicaine. On ne saurait souligner assez que c'est là la principale disposition militaire classique prise par l'Irak contre un adversaire étranger.</p>	ES	TRIGGER	
			ITEM 1
			ITEM 2
			ITEM 3
<p>En somme, le régime a remodelé et réorienté ses forces armées pour aller vers un système plus sûr et plus compact, au caractère répressif et défensif.</p>		CLOSURE	

Figure 3. An annotated enumerative structure in ANNODIS⁸

21 The manual annotation phase made use of the specially-designed Glozz interface (Mathet & Widlöcher, 2009). In order to assist the annotators in their task, this interface displays text with visual layout (like a real-life document) and highlighting of pre-marked features. The annotation procedure consisted in three tasks: 1) detecting the structures, 2) delimiting and labelling their components, and 3) marking the

8. The presentation of analysed examples obeys the following conventions: horizontal lines in the left column indicate paragraph breaks in the original (in Figure 3, the trigger and all three items are in one paragraph, the closure comes after a paragraph break); all boxed segments in the left column correspond to a complete paragraph, except when an excessively long paragraph had to be cut, which is signalled by "[...]" (as in Figures 8 and 10).

cues signalling the structure, i.e. the features that the coder saw as having helped her/him detect it. This last task included the validation of any pre-marked feature seen as relevant, as well as the identification of additional features that had not been pre-marked. The annotated additional features are grouped in three new cue types, concerning mostly the signalling of items: syntactic parallelism, coreferential expressions, and trigger reiteration. As well as the three sub-segments described in 2.2 (trigger, items and closure), the annotation model for enumerative structures comprised another optional segment called **enumeratheme**. The enumeratheme is an expression specifying the co-enumerability criterion. Figure 3 gives an example of a complete annotated ES including a trigger, three items, a closure, two enumerathemes (framed), and four features annotated as cues (in bold): a prospective element in the trigger, two connectives indicating item breaks and a concluding connector in the closure.

- 22 More detailed accounts of the principles underlying the annotation procedure and interface can be found in Péry-Woodley et al. (2011). The next section moves on to a presentation of the resulting annotated data.

3.3. Overview and “visual” typology of the enumerative structures annotated

- 23 In order to give an idea of the scale of the resource constituted, here is a quantitative summary of the results of the annotation campaign, in terms of the different objects presented in section 3.2:

- 968 enumerative structures (ESs),
- 725 triggers,
- 3295 items,
- 129 closures,
- 449 enumerathemes,
- 3779 features annotated as signalling ESs.

- 24 ES frequency varies across the three sub-corpora (over 40% of the annotated ESs are found in WIK2 and around 30% in each of the other two genres). Though the data in the ANNODIS resource make it possible to perform detailed contrastive analyses of ESs and their composition (see 3.1.), this paper focuses solely on the features signalling ESs without taking into account genre variations. Table 3 gives figures for the different types of features annotated as ES cues⁹.

- 25 The first analysis of the ANNODIS data concerns the interaction between ESs and document structure. As presented in Ho-Dac, Péry-Woodley and Tanguy

9. 37% of the cues annotated were added by the coders to the pre-marked features (1413 out of 3779).

(2010), a first visual typology can be proposed to characterise enumerative structures according to their interaction with document structure at different granularity levels:

- Type 1: multisection (items=sections), as in Figure 9;
- Type 2: formatted lists, as in Figures 2, 4 and 6;
- Type 3: extending over several paragraphs, as in Figures 3 and 8;
- Type 4: contained within one paragraph, as in Figures 5 and 7.

Feature type	Number	Feature subtype	Number
Item introducer	1,262	sequencer	333
		circumstantial adverbial	495
		parallelism	422
		co-referential expression	12
Prospective element	278		
Encapsulation	36		
Visual device and punctuation	1,792	bulleted/numbered text	923
		(sub)heading	318
		punctuation*	551
Other feature*	199		
Connective*	212		

Table 3. Features annotated as ES cues

(* indicates that these features may occur in different components: trigger, item or closure)

26

Table 4 gives an overview of these four types in terms of frequency and main characteristics (number of words, of items, of paragraphs).

Visual type of ES		Number of ESs		Number of words			Number of items		Number of paragraphs		
		number	%	mean	min	max	mean	max	mean	min	max
1	multisection	93	13.1	1,671	231	7,831	3.4	12	16.2	3	72
2	formatted list	186	26.3	182	8	1,541	4	48	NA	NA	NA
3	multiparagraph	164	23.2	380	29	2,953	3.6	20	4	2	25
4	(intra)paragraph	265	37.4	109	13	730	2.7	8	1	1	1
Total		708	100	397			3.4		4		

Table 4. Frequency and main characteristics of the four visual ES types¹⁰

10. The minimum number of items is not given as the minimum required is two items.

27 Type 1 corresponds to ESs interacting with the highest level of visual segmentation, i.e. segmentation into sections. Each item of a Type 1 ES corresponds to a section (or subsection). As expected, this level shows the highest values for length and number of paragraphs (cf. the rightmost columns in Table 4).

28 At the other end of the scale, on the most local level, Type 4 depicts ESs inserted inside a paragraph (156 ESs) or corresponding exactly to a paragraph (109 ESs)¹¹. Type 4 ESs are the most frequent in the corpus. According to the annotators' feedback, they are also the most difficult to annotate. Type 4 ESs have the smallest cardinality (number of items) with a mean below 3 (2.7) and, as can be expected, they are the shortest.

29 In between these two extremes, Type 2 and 3 ESs match layout to a greater or lesser extent, and cover different lengths and granularity levels. Type 2 ESs are the only ones to be defined solely in terms of specific typographical and layout features (bullet points or numbers). Perhaps as a consequence, this type shows the widest variety in length and number of items, from very local formatted lists composed of only two items to large-scale lists of up to 48 items covering an entire section¹². In terms of average length, Type 2 ESs are fairly close to Type 4 ESs, which suggests that formatted lists can, in some configurations, be as local as (intra)paragraph ESs.

30 Type 3 ESs also show variations in length and cardinality but to a lesser degree than Type 2 ESs. Multiparagraph ESs are longer than formatted lists, in other words Type 3 ESs structure larger spans of text and may have a higher cardinality than Type 1 ESs. In terms of granularity level, Type 3 ESs seem to be more global than Type 2 ESs, as shown by their high number of words and paragraphs.

31 Concerning the main characteristics of these four visual types of ESs, some simple statistical measures provide the following interesting significant correlations:

- cardinality is higher in Types 1 and 2 (3.8 items on average against 3);
- Types 1 and 2 are also characterised by a significantly higher presence of triggers;
- enumerathemes are more often present in Type 2 ESs and less often in Type 1 ESs;
- closures are significantly less frequent in Type 1 ESs.

32 The first three correlations can be seen as suggesting that ESs which interact explicitly with layout tend to be more semantically specified: in Type 1 ESs, headings give the interpretation criterion of each item, while Type 2 ESs are usually associated with an enumeratheme. Type 2 ESs may therefore be seen as a local counterpart of

11. No statistical test has shown any significant differences between these two subtypes of ES.

12. This particular ES appears in the Wikipedia entry describing the Watergate scandal and lists all the chronological events of this affair (see http://fr.wikipedia.org/w/index.php?title=Scandale_du_Watergate&oldid=45530224).

Type 1 ESs: bullet points or numbers in Type 2 introduce items in a manner comparable to what section headings do in Type 1. But, as bullet points say nothing about the co-enumerability criterion, an explicit expression of it —the enumeratheme— becomes necessary. The counterpart of this hypothesis may be that Types 3 and 4 ESs call upon another form of semantic specification by using lexical cues in items. Such lexical cues can be seen as semantically specifying every item in a way similar to headings for Type 1. The next section proposes a representation of the cues associated to each type in order to describe and compare their signalling.

4. Analysing the annotated data: how are enumerative structures really signalled?

- 33 This is where we can reap the benefits of our method, whereby the identification of signalling devices comes as a result of the analysis of the annotated data. In this section, we use the visual typology just introduced as the basis for the detailed account of our analysis of annotator-identified ES cues (validated or added cues). This account starts with an illustrated overview of the signalling of the different ES components (trigger, items and closures). It goes on (section 4.2) to focus on the data concerning item cues only, first from a quantitative viewpoint, then with a qualitative analysis of specific aspects relevant to the questions raised in section 2.1 on the metadiscursive nature of signalling.

4.1. Overview of ES composition and signalling

- 34 The overview of ES composition¹³ and signalling summarised in Table 5 is organised according to the four ES types described in the previous section. Table 5 indicates for each type the percentage of structures containing optional elements (triggers and closures), and focuses on the presence of annotated cues in each ES element. At this stage, an element is considered “cued” if it contains at least one annotated cue of any type.

Type	ESs w. trigger		Cued trigger		Cued items		ESs w. closure		Cued closure	
1	82	88%	73	89%	314	100%	2	2%	2	100%
2	176	95%	175	99%	741	100%	27	15%	22	82%
3	101	62%	91	90%	446	75%	31	19%	24	77%
4	172	65%	145	84%	478	66%	26	10%	23	89%
Total	531	75%	484	91%	1979	83%	86	12%	71	83%

Table 5. Presence of cues per ES element and visual type

13. See Table 4 for figures concerning numbers of items.

35 Looking at the detailed composition of each type, we see in Table 5 that whereas most ESs are introduced by a trigger (75% overall) few have a closure (from 2% for Type 1 to 19% for Type 3). As noted earlier, the presence of a trigger in an ES is related to its visual type: triggers appear significantly more often in ESs of Types 1 or 2.

36 Looking now at percentages of cued elements, the high level of signalling is striking, with over 80% of triggers, items and closures containing at least one annotated cue. This is not surprising given that ESs are textual patterns whose function is entirely tied up with their identifiability. Signalling is particularly high in triggers (91% of triggers are cued), where it often takes the form of a combination of a lexico-grammatical and a punctuational cue, as illustrated in the Type 2 ES in Figure 4.

[...] Comme H. Paul, il ramène le concept au calcul de l'équation de la quatrième proportionnelle. Il lui assigne deux rôles majeurs :		
ES	TRIGGER	
	ITEM 1	
	ITEM 2	

Figure 4. Example of a Type 2 ES (formatted list)

37 The next example is a Type 4 ES (Figure 5) where the absence of signalling in the trigger is compensated in the items, signalled by two distinct and internally coherent sequences of cues —syntactic parallelisms (*Par*+ definite NP) and sequencers (*tout d'abord, ensuite, enfin*):

ES	TRIGGER	
	ITEM 1	
	ITEM 2	
	ITEM 3	

Figure 5. Example of a Type 4 ES (intraparagraph)

38 As far as closures are concerned, various types of lexico-grammatical cues are taken into account, such as encapsulations (*ses forces armées* in Figure 3; *les positions a. et b.* in Figure 6 below), connectives (*En conclusion* in Figure 2; *En somme* in Figure 3).

39 The aim of this rapid illustrated tour of ES signalling is to give an impression of the diversity of cues found in the three elements making up the ESs annotated in the ANNODIS resource. Already, these examples show how difficult it would be to neatly separate propositional content from metadiscursive elements. In Figure 4, an enumeratheme is expressed in the trigger: its characteristic form contributes to the identification of the trigger (*rôles majeurs*), but it also contributes to content by specifying the co-enumerability criterion. The items in this same example are signalled by a combination of visual cues (bullets and formatted list layout) and syntactic parallelism (initial infinitive: *passer, permettre*). Typography and layout can indeed be seen as primarily textual, but they work together with the parallel infinitives in Figure 4, so that it is the combination of cues that signals the items, intimately meshing the textual and ideational metafunctions. In a similar way, though with a different combination of cues, the items in Figure 5 are signalled by a combination of syntactic parallelism and sequencers, with both the textual and the ideational metafunctions present here as well. The next two sub-sections will examine these interactions in more detail by focusing on the signalling of items.

4.2. The signalling of items

40 Along the lines of the analysis sketched above, several facets of the signalling of items will be developed here: first, the annotated cues will be grouped in major types and the distribution of these types in the data presented and discussed; we then propose an illustrated zoom on some types of cues (circumstantial adverbials and headings) in order to pursue our discussion of the interaction between the ideational and the textual metafunctions.

4.2.1. Annotated cues in items

41 To introduce this section, Figure 6 presents a “belt and braces” Type 2 ES, in which the items are signalled by a combination of cues: numbered by lower case letters, they start with correlative adverbials (*D’une part, /D’autre part,*), and each one makes up a separate paragraph. As trigger and closure are also present and clearly signalled, this ES seems really meant to be noticed.

42 On the basis of our analysis of observed combinations such as this in the ANNODIS data, we propose to first identify cue usage and common associations. In other words, whereas section 4.1 looked at cued ES elements whatever the type of cue used and without regard for the use of multiple cues, we will now delve into the different types of cues annotated, and examine their combinations. The cues annotated alongside items have been grouped into six categories, with the following labels:

- **Seq.:** sequencers and connectives,
- **Circ.:** circumstantial adverbials¹⁴,
- **Bullets:** bullets,
- **Punct.:** punctuation,
- **Head.:** headings,
- **Paral.:** syntactic parallelism between items.

<p>2.2 Deux manières de nier la polysémie Une réponse possible est de refuser d’attribuer à toute unité lexicale du sens préétabli (voir Kleiber, 1999). Le problème de la pluralité de sens se trouve alors réglé, puisque l’exclusion de toute monosémie entraîne en même temps celle de toute pluralité de sens. Nous n’envisagerons pas ici une telle position, parce qu’elle oblige à prendre position sur le statut et la nature du sens en général (Larsson, 1997 et Kleiber, 1999). Si l’on met de côté une telle position extrême, mais tout à fait défendable, la polysémie en tant qu’association de plusieurs sens à une même forme lexicale se trouve niée de deux manières apparemment paradoxales :</p>	ES	TRIGGER
<p>-a- D’une part, les vocables donnés comme polysémiques se voient en quelque sorte “monosémisés” par la mise en avant d’un invariant supérieur, de quelque nature qu’il soit, qui unifie les différentes interprétations. L’idée, implicite ou explicite, est que ce sens subsumant, explicatif de la diversité interprétative, ne correspond pas aux interprétations observables en discours et [...]</p>		ITEM 1
<p>-b- D’autre part, de façon tout à fait inverse aux tentatives de monosémisation, on fait proliférer les sens ou variations de sens. Autrement dit, on procède à la multiplication des cas de polysémie. Soit en postulant qu’il s’agit d’une variation [...]</p>		ITEM 2
<p>Les positions -a- et -b- ne sont qu’apparemment paradoxales : il n’y a aucune contradiction, d’un côté, à faire disparaître la pluralité de sens sous un sens amont ou invariant et, de l’autre, à procéder à la multiplication des petits pains polysémiques, dans la mesure où le point commun des deux positions est que la multiplicité de sens n’est qu’un phénomène discursif et non un phénomène de l’unité lexicale elle-même, puisque dans les deux cas, les sens relevés ne sont pas associés à l’unité lexicale elle-même, mais sont le résultat d’une construction discursive.</p>		CLOSURE

Figure 6. “Belt and braces” example of Type 2 ES

43 Table 6 summarises the analysis of cue usage in the ANNODIS corpus. It shows the proportion of items signalled by each type of cue for the four ES types. All values are percentages, and are relative to the existence of the corresponding element: a 100% cue means that every item in ESs of this type has such a cue. Percentages do not add up to 100 as items may each have zero or several cues (of different kinds).

14. Under the label “Circ.” we have provisionally lumped together not only temporal and spatial adverbials, but also those described in the discourse framing literature as “more abstract localizing adverbials” (Charolles, 2005: 16), such as the ones introducing the items of Figure 8. The rationale for this categorisation is that they all contribute to propositional content, unlike organisational adverbials. It is however unsatisfactory and requires further work.

ES Type	ITEM CUES					
	Head.	Bullets	Punct.	Seq.	Circ.	Paral.
1	100	0	0.7	0.7	0.7	6.1
2	0	100	NA	2.2	1.6	14.9
3	0	0	14.6	27.8	37.2	15
4	0	0	13.2	46.7	14.2	20.1
Total	15.9	31.2	6.7	18.7	12.7	14.8

Table 6. Cue usage: distribution of item cues in the four ES types in the ANNODIS data¹⁵

44 Headings and bullets are 100% present in ESs of Types 1 and 2 as they feature in the definition of these ES types. Other kinds of cues are extremely rare for these two types, except for parallelisms in Type 2 (14.8%). Types 3 and 4 ESs make use of a greater variety of cues, and we see lexico-syntactic cues (Seq., Circ., Paral.) gaining in importance as the visual dimension recedes. There are however noteworthy differences in lexical cue usage between multiparagraph and (intra)paragraph ESs. Type 3 ESs make roughly equal use of circumstantial adverbials and sequencers to signal items (27.8% and 37.2% respectively), whereas in Type 4, with visual marks totally absent, sequencers (46.7%) clearly prevail over circumstantial adverbials (14.2%). We will return to the use of circumstantial adverbials, with illustrations and analyses.

45 Interesting as these trends may be, one needs to be extremely cautious in drawing any conclusions because Table 6 says nothing about the frequency of multiple cue usage, as illustrated in Figures 4, 5 and 6, nor about actual cue combinations. We are currently developing a method to compute and represent these combinations so as to capture major trends in our annotated data. This work will be the topic of a separate publication.

46 We now propose to examine examples of structures containing two non-archetypical cue types: circumstantial adverbials, in order to further the observations made in previous sections, and headings, characteristic of multisection ESs.

4.2.2. *Circumstantial adverbials as item cues*

47 The following examples are meant first to illustrate how sequences of adverbials other than sequencers function as item cues in non-equivocal enumerative structures, and second to point out some specificities of this cue type.

15. Punct.=NA for Type 2 ESs because final item punctuation marks were not annotated as cues for this type of ESs, as they were considered redundant with bullets.

<p>Aux États-Unis, la seule personne à avoir été jugée jusqu'à présent pour son implication directe avec les attentats du 11 Septembre est le Français Zacarias Moussaoui. Arrêté moins d'un mois avant les attaques, il a été accusé par les autorités fédérales américaines d'avoir eu connaissance des attentats à venir mais de n'avoir pas communiqué ses informations. Le 3 mai 2006, au terme de deux mois de procès, il a été reconnu coupable par le jury du tribunal fédéral d'Alexandria en Virginie de six chefs d'accusation de complot en liaison avec les attentats terroristes du 11 Septembre et condamné à la prison à perpétuité, sans possibilité de remise de peine.</p>	<p>ES</p>	<p>ITEM 1</p>
<p>En Allemagne, le marocain Mounir al-Motassadeq arrêté le 28 novembre 2001, est condamné une première fois à quinze ans de prison en 2003 pour complicité dans ces attaques. Remis en liberté en février 2006 après que sa condamnation a été cassée, il voit sa première peine confirmée par le tribunal de Hambourg le 8 janvier 2007.</p>		<p>ITEM 2</p>
<p>En Espagne, le Syrien Imad Eddin Barakat Yarkas, chef de la cellule locale d'Al-Qaida est arrêté le 13 novembre 2001, inculpé de conspiration en vue des attentats de septembre 2001. Il est condamné le 26 septembre 2005 à vingt-sept ans de prison.</p>		<p>ITEM 3</p>

Figure 7. A Type 3 ES (several paragraphs) with adverbials as item cues

48 The triggerless multiparagraph ES in Figure 7 was selected to illustrate how content-bearing adverbials can function as clear item cues. In this Type 3 ES, the paragraph-initial spatial adverbials present in all items clearly signal the organisation of the enumeration (of court cases) in terms of the countries where the events took place.

49 The example in Figure 8 shows a Type 4 ES with circumstantial adverbials¹⁶ as item cues, which is somewhat atypical for (intra)paragraph ESs (cf. 4.2.1 and below). The “inconvenients” (*disadvantages*) announced in the trigger are spelled out in an enumeration which could just as well be introduced by sequencers (*Firstly...*; *Secondly...*; cf. section 2.1). Instead, it is organised in terms of different points of view (*Au plan logique* vs. *Au plan pratique*), with sentence-initial adverbials contributing along two dimensions: they organise text (textual metafunction) and they specify each point of view (ideational metafunction).

<p>[...] La multiplication de ces dates fut sans doute à Doha un moyen d'obtenir un compromis entre tenants d'un cycle étroit et court et partisans d'un cycle large et long, au prix toutefois de plusieurs inconvenients. Au plan logique, l'idée de butoir s'oppose à la notion même de cycle, où les arbitrages se font à la fin entre tous les sujets. Au plan pratique, ces dates butoirs ne sont guère respectées et provoquent, comme c'est le cas actuellement, des commentaires critiques sur l'enlisement, l'échec et l'absence de perspective des discussions. [...]</p>	<table border="1"> <tr> <td data-bbox="1072 1429 1125 1668" rowspan="3"> <p>ES</p> </td> <td data-bbox="1125 1429 1248 1529"> <p>TRIGGER</p> </td> </tr> <tr> <td data-bbox="1125 1529 1248 1579"> <p>ITEM 1</p> </td> </tr> <tr> <td data-bbox="1125 1579 1248 1668"> <p>ITEM 2</p> </td> </tr> </table>	<p>ES</p>	<p>TRIGGER</p>	<p>ITEM 1</p>	<p>ITEM 2</p>
<p>ES</p>	<p>TRIGGER</p>				
	<p>ITEM 1</p>				
	<p>ITEM 2</p>				

Figure 8. A Type 4 ES (one paragraph) with adverbials as item cues

16. Or more precisely “abstract localizing adverbials” (cf. 4.2.1).

50 Returning to the commentary on Table 6 (4.2.1), content-bearing adverbials were seen to be a favourite cue in multiparagraph ESs (present in 37.2% of these ESs), as against sequencers in (intra)paragraph ESs (46.7%, against 14.2% for circumstantial adverbials). An explanation for this difference can be suggested on the basis of the constraints on the organising role of time adverbials proposed in (Ho-Dac & Péry-Woodley, 2009). The authors argue that in order to function as discourse segmentation markers, time adverbials need to be paragraph initial. Such a constraint is incompatible, by definition, with (intra)paragraph Type 4 ESs. Sequencers on the other hand, which are specialised in the signalling of ESs, may be more independent from positional constraints. Though infrequent, (intra)paragraph ESs with adverbially cued items do occur, as shown in Figure 8.

4.2.3. Headings, multi-level structures and constraints on Type 1 ESs

51 Type 1 ESs can be described as a sequence of same-level headed sub-sections with an upper-level heading acting as a trigger. Two examples of these high-level textual patterns are given below. Both have been drastically abridged, but in Figure 10 some of the text contained in the two items of the top-level ES has been left to show an instance of nested ESs (two Type 3 within one Type 1).

6. Les conquêtes amoureuses de César	ES	TRIGGER
<p data-bbox="352 1055 751 1084">6.1. Les femmes de la haute société romaine</p> <p data-bbox="384 1084 1054 1126">D'après l'historien latin Suétone, César séduit de nombreuses femmes tout au long de sa vie et plus particulièrement celles issues de la haute société romaine.</p> <p data-bbox="384 1126 1054 1200">Il aurait ainsi séduit Postumia, la femme de Servius Sulpicius, Lollia, la femme d'Aulus Gabinius et Tertulla, la femme de Marcus Crassus. Il semble avoir également fréquenté Mucia la femme de Pompée.</p> <p data-bbox="384 1200 1054 1323">César entretient des relations particulières avec Servilia Caepionis, mère de Brutus, qu'il semblait particulièrement apprécier. Ainsi, Suétone rapporte les divers présents et avantages qu'il offrit à sa bien-aimée, dont notamment une magnifique perle d'une valeur de six millions de sesterces. L'amour de Servilia pour César est publiquement connu à Rome.</p> <p data-bbox="384 1323 1054 1397">Le penchant de César pour les plaisirs de l'amour semble également attesté par ces quelques vers chantés par ses soldats, lors de son triomphe à Rome au retour de ses campagnes en Gaule, et rapportés par Suétone :</p> <p data-bbox="384 1397 1054 1447">"Citoyens, surveillez vos femmes : nous amenons un adultère chauve Tu as fornicqué en Gaule avec l'or emprunté à Rome."</p>		ITEM 1
<p data-bbox="352 1447 488 1476">6.2. Les reines</p> <p data-bbox="384 1476 1054 1525">César a des relations amoureuses avec Eunoé, femme de Bogud, roi de Mauritanie.</p> <p data-bbox="384 1525 1054 1688">Cependant, sa relation avec Cléopâtre VII est restée plus célèbre. Suétone rapporte que César a remonté le Nil avec la reine égyptienne et l'a fait venir à Rome en la comblant d'honneurs et de présents. C'est aussi un bon moyen pour lui de tenir l'Égypte, où trois légions sont présentes, et dont le rôle dans l'approvisionnement en céréales de l'Italie commence à devenir prépondérant. Toujours est-il que Cléopâtre est présente à Rome au moment de l'assassinat de César et qu'elle rentre rapidement dans son pays après le meurtre.</p>		ITEM 2

Figure 9. Example of a Type 1 ES (multisection)

52 Of the four types, Type 1 ESs are the most closely linked to document structure, i.e. the combination of segmentation (sections, sub-sections) and "labelling" of

segments via headings (and sub-headings). Indeed, document structure could be seen generally as realising a process similar to enumerating: a number of sub-sections are presented as equal and forming a set by virtue of being sub-sections of section X. The relation between a section heading and the sub-headings below it could arguably be seen as an inclusion relation in the same way as for the co-items in an enumerative structure. Yet we insist that all headed sections including headed sub-sections cannot be seen as enumerative structures, and they clearly were not identified as such by our annotators. What distinguishes the Type 1 ESs identified by the annotators is that the first level heading provides a semantic criterion, the co-enumerability criterion which is at the root of the act of enumerating. In Figure 9 for instance, the rationale for placing the groups of women referred to by the successive sub-headings on an equal footing is that they are presented as having a common property, they are amongst Cesar’s amorous conquests. All the items listed in the two-level ESs in Figure 10 are grouped in this paper as types of linguistic tests, organised in two sub-groups, syntagmatic and paradigmatic.

3.2.2 Tests syntagmatiques et paradigmatiques	ES1	TRIGGER	
3.2.2.1 Syntagmatiques		ITEM 1	
Il y a des correspondants formels syntagmatiques et paradigmatiques à la non-unifiabilité que [...]		ES2	ITEM 1
En premier lieu , on mentionnera un critère guère utilisé, car pas facilement manipulable et d’application limitée aux SN, mais pourtant intéressant : la difficulté [...]			ITEM 2
En deuxième lieu , on dispose du critère de cooccurrence compatible (Mel’Cuk, Clas et Polguère, 1995), appelé aussi critère de Green-Apresjan par [...]			ITEM 3
Troisièmement, le critère de distribution différente (ou de cooccurrence différentielle selon Mel’Cuk, Clas et Polguère, 1995) est une autre manifestation de la non-unifiabilité. L’idée sous-jacente est que [...]		ITEM 2	
3.2.2.2 Paradigmatiques		ES3	ITEM 1
La non-unifiabilité se laisse aussi appréhender au niveau paradigmatique : l’hypothèse est qu[...].			ITEM 2
En premier lieu , l’absence d’hyperonyme pour couvrir lexicalement les deux lectures est un écho direct de la non-unifiabilité sémantique. S’il y a un hyperonyme disponible [...]			ITEM 3
En deuxième lieu , la synonymie (intra- et interlinguistique) s’avère être fort utile : le test consiste à montrer que deux lectures non unifiables ont [...]			ITEM 4
La synonymie appelle tout naturellement l’antonymie : la dissociation de deux lectures peut se manifester également dans l’existence d’antonymes différents, [...]			
[...]			

Figure 10. Nested Type 3 ESs in a high-level Type 1 ES

53 In these examples, unlike in non enumerative sets of headings, there is a specific semantic relation holding between higher-level and lower-level headings. Headings

in multisection ESs are concerned, like any headings, with segmenting the text and labelling the headed segments, but they also provide the semantic criterion which is at the basis of the textual act of enumerating, in our terms they express the enumeratheme. Both the ideational and the textual metafunctions are concerned here, and overlap in a way that would be impossible to disentangle.

5. Conclusion

54 In order to study the signalling of discourse organisation, this paper draws upon a corpus of discourse-level annotated texts. The existence of this corpus means that we have been able to broaden the scope and escape from the danger of the circularity which threatens semasiological approaches, and achieve—at least in part—our objective of discovering complex discourse markers. This open-ended approach requires new conceptual and methodological tools, and we are strongly aware of the fact that we are a long way from having made optimal use of the data available. We can however claim to have presented some results on the signalling of enumerative structures, results which lead to some observations on the signalling of text organisation in general and the notion of metadiscourse discussed at the beginning.

55 We show enumerating to be an extremely frequent textual pattern, prevalent in all three sub-corpora. We also show the diversity of structures that come under the label “enumerative structure”, which are diverse in size, textual granularity level, semantico-pragmatic function, and forms of signalling. The cues signalling enumerative structures have been identified: besides classical cues such as sequencers and typical formatted list layouts, we have pointed to the role of headings, circumstantial adverbials in initial position, and syntactic parallelism.

56 As regards the opening question of the separateness of textual signalling or metadiscourse from textual content, the analysis of enumerative structures leads to viewing the textual and ideational metafunctions, together with the interpersonal metafunction which is particularly obvious in titles and headings, as intimately interlocked. Looking at specific cue-types, it is possible to see them as specialised text organisation signals, but when one takes in the signalling of the structure as a whole, it becomes clear that enumerative structures fundamentally require a conjunction of two functions: segmenting/sequencing and expressing the enumeratheme (or co-enumerability criterion). In some cases, a single cue-type can fulfil this dual function: headings for instance, or paragraph-initial circumstantial adverbials, can have the property of both sequencing and labelling. In other cases, the two functions are distributed over different cues: bullets and numbers may ensure the segmenting/sequencing, associated with syntactic parallelism or circumstantial adverbials which make the enumeratheme explicit. What we see with this complex interplay of cues severely questions the notion that textual metadiscourse can be systematically identified as distinct from textual content.

References

- ADAM, J.-M. & REVAZ, F. 1989. Aspects de la structuration du texte descriptif: les marqueurs d'énumération et de reformulation. *Langue française* 81: 59-98.
- AGUILAR, M. 2008. *Metadiscourse in Academic Speech. A Relevance-Theoretic Approach*. Bern: Peter Lang.
- ALVAREZ DE MON Y REGO, I. 2001. Encapsulations and Prospection in Written Scientific English. *Estudios ingleses de la Universidad Complutense* 9: 81-102.
- BESSONNAT, B. 1988. Le découpage en paragraphes et ses fonctions. *Pratiques* 57: 81-105.
- BIBER, D. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
- BOURAOU, J.-L. & VIGOUROUX, N. 2003. Les marqueurs des structures énumératives sur le web: analyse pour la transmodalité. In J. MADELAINE & C. FAURE (eds), *Document électronique dynamique: actes du sixième colloque international sur le document électronique (24-26 novembre 2003, Caen)*. Paris: Europa: 199-217.
- BOURIGAUT, D. 2007. *Un analyseur syntaxique opérationnel: Syntax*. "Habilitation à diriger les recherches" report. Université de Toulouse.
- BRAS, M., PRÉVOT, L. & VERGEZ-COURET, M. 2008. Quelle(s) relation(s) de discours pour les structures énumératives? In J. DURAND, B. HABERT & B. LAKS (eds), *Actes du congrès mondial de Linguistique française (CMLF 2008)*. Paris: Institut de linguistique française: 1945-1964. Available online: <http://w3.erss.univ-tlse2.fr:8080/index.jsp?perso=vergez&subURL=publis/CMLFo8.pdf>.
- CHAROLLES, M. 1997. L'encadrement du discours – univers, champs, domaines et espaces. *Cahiers de recherche linguistique* 6: 1-73.
- CHAROLLES, M. 2005. Framing Adverbials and their Role in Discourse Cohesion: From Connection to Forward Labelling. In *Papers of the Symposium on the Exploration and Modelling of Meaning (SEM-05)*. 13-30. Available online: <http://w3.erss.univ-tlse2.fr/sem05/proceedings.html>.
- CHAROLLES, M. et al. (eds) 2005. *Journal of French Language Studies: Temporal and Spatial Dimensions of Discourse Organisation* 15 (2). Cambridge: Cambridge University Press.
- CHAROLLES, M. & VIGIER, D. 2005. Les adverbiaux cadratifs en position préverbale: portée cadrative et organisation des discours. *Langue française* 148: 9-30.
- CONTE, M.-E. 1996. Anaphoric encapsulation. *Belgian Journal of Linguistics* 10: 1-10.
- CRISMORE, A., MARKKANEN, R. & STEFFENSEN, M.S. 1993. Metadiscourse in Persuasive Writing: A Study of Texts Written by American and Finnish University Students. *Written Communication* 10 (1): 39-71.
- CROMPTON, P. 2006. The Effect of Position on the Discourse Scope of Adverbials. *Text and Talk* 26 (3): 245-279.
- FRANCIS, G. 1994. Labelling Discourse: An Aspect of Nominal-Group Lexical Cohesion. In M. COULTHARD (ed.), *Advances in Written Text Analysis*. London – New York: Routledge: 83-101.
- HALLIDAY, M.A.K. 1977. Text as Semantic Choice in Social Contexts. In T. VAN DIJK & J. PETÖFI (eds), *Grammars and Descriptions*. Berlin: Walter de Gruyter: 176-226.

- HEMPEL, S. & DEGAND, L. 2008. Sequencers in Different Text Genres: Academic Writing, Journalese and Fiction. *Journal of Pragmatics* 40 (4): 676-693.
- HO-DAC, L.-M., JACQUES, M.-P. & REBEYROLLE, J. 2004. Sur la fonction discursive des titres. In S. PORHIEL & D. KLINGLER (eds), *L'unité texte*. Pleyben: Perspectives: 125-152.
- HO-DAC, L.-M. & PÉRY-WOODLEY, M.-P. 2009. A Data-Driven Study of Temporal Adverbials as Discourse Segmentation Markers. *Discours* 4. Available online: <http://discours.revues.org/5952>.
- HO-DAC, L.-M., PÉRY-WOODLEY, M.-P. & TANGUY, L. 2010. Anatomie des structures énumératives. In *Actes de la 17^e conférence sur le Traitement automatique des langues naturelles (TALN 2010 – Montréal, 19-23 juillet 2010)*. Available online: http://www.iro.umontreal.ca/~felipe/TALN2010/Xml/Papers/all/taln2010_submission_26.pdf.
- HYLAND, K. 1998. Persuasion and Context: The Pragmatics of Academic Metadiscourse. *Journal of Pragmatics* 30 (4): 437-455.
- HYLAND, K. & TSE, P. 2004. Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics* 25 (2): 156-177.
- IFANTIDOU, E. 2005. The Semantics and Pragmatics of Metadiscourse. *Journal of Pragmatics* 37 (9): 1325-1353.
- JACKIEWICZ, A. 2005. Les séries linéaires dans le discours. *Langue française* 148: 95-110.
- JACKIEWICZ, A. & MINEL, J.-L. 2003. L'identification des structures discursives engendrées par les cadres organisationnels. In *TALN 2003 (10^e conférence annuelle sur le Traitement automatique des langues naturelles, Batz-sur-Mer, 11-14 juin 2003)*. Rocquencourt: INRIA: 155-164.
- KATZENBERGER, I. 2005. The Super-Structure of Written Expository Texts – A Developmental Perspective. In D.D. RAVID & H.B.-Z. SHYLDKROT (eds), *Perspectives on Language and Language Development: Essays in Honor of Ruth A. Berman*. Berlin – New York: Springer: 327-336.
- LAIPPALA, V. 2008. Nature des marqueurs des séries linéaires dans des articles scientifiques. In J. DURAND, B. HABERT & B. LAKS (eds), *Actes du congrès mondial de Linguistique française (CMLF 2008)*. Paris: Institut de linguistique française: 1365-1378. Available online: <http://www.linguistiquefrancaise.org/articles/cmlf/pdf/2008/01/cmlf08122.pdf>.
- LE DRAOULEC, A. & PÉRY-WOODLEY, M.-P. 2005. Encadrement temporel et relations de discours. *Langue française* 148: 45-60.
- LEGALLOIS, D. 2006. Quand le texte signale sa structure: la fonction textuelle des noms sous-spécifiés. *Corela: numéro thématique Organisation des textes et cohérence des discours*. Available online: <http://corela.edel.univ-poitiers.fr/index.php?id=1465>.
- LUC, C. et al. 1999. A Linguistic Approach to Some Parameters of Layout: A Study of Enumerations. In R. POWER & D. SCOTT (eds), *AAAI Fall Symposium Technical Report "Using Layout for the Generation, Understanding or Retrieval of documents"*. 35-45. Available online: <http://www.aaai.org/Papers/Symposia/Fall/1999/FS-99-04/FS99-04-005.pdf>.
- LUC, C. et al. 2000. Les énumérations: structures visuelles, syntaxiques et rhétoriques. In M. GAIO & É. TRUPIN (eds), *Document électronique dynamique: actes du troisième colloque international sur le document électronique (4-6 juillet 2000, Lyon)*. Paris: Europa: 21-40.

- MAO, L.R. 1993. I Conclude Not: Towards a Pragmatic Account of Metadiscourse. *Rhetoric Review* 11 (2): 265-289.
- MATHET, Y. & WIDLÖCHER, A. 2009. La plate-forme GLOZZ: environnement d'annotation et d'exploration de corpus. In *Actes de la 16^e conférence sur le Traitement automatique des langues naturelles (TALN 2009 – Senlis, 24-26 juin 2009)*. Available online: http://www-lipn.univ-paris13.fr/taln09/paper/paper_TALN_120.html.
- MAUREL, F., LEMARIÉ, J. & VIGOUROUX, N. 2003. Oralisation de structures visuelles: de la lexico-syntaxe à la prosodie. In A. METTOUCHI & G. FERRÉ (eds), *Actes de « Interfaces Prosodiques 2003 » (IP2003 – Nantes, 27-29 mars 2003)*. 137-142.
- PASSERAULT, J.-M. & CHESNET, D. 1991. Le marquage des paragraphes: son rôle dans la gestion des traitements pendant la lecture. *Psychologie française* 36 (2): 159-165.
- PÉRY-WOODLEY, M.-P. et al. 2009. ANNODIS: une approche outillée de l'annotation de structures discursives. In *Actes de la 16^e conférence sur le Traitement automatique des langues naturelles (TALN 2009 – Senlis, 24-26 juin 2009)*. Available online: http://www-lipn.univ-paris13.fr/taln09/paper/paper_TALN_52.html.
- PÉRY-WOODLEY, M.-P. et al. 2011. La ressource ANNODIS, un corpus enrichi d'annotations discursives. *TAL* 52 (3).
- PORHIEL, S. 2007. Les structures énumératives à deux temps. *Revue romane* 42 (1): 103-135.
- POWER, R., SCOTT, D. & BOUAYAD-AGHA, N. 2003. Document structure. *Computational Linguistics* 29 (2): 211-260.
- REBEYROLLE, J., JACQUES, M.-P. & PÉRY-WOODLEY, M.-P. 2009. Titres et intertitres dans l'organisation du discours. *Journal of French Language Studies* 19 (2): 269-290.
- SCHMID, H.-J. 2000. *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin – New York: Mouton de Gruyter.
- 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 – Paris: De Boeck-Duculot.
- STARK, H. 1988. What Do Paragraph Markers Do? *Discourse Processes* 11 (3): 275-304.
- TADROS, A. 1994. Predictive Categories in Expository Texts. In M. COULTHARD (ed.), *Advances in Written Text Analysis*. London – New York: Routledge: 69-82.
- THOMPSON, S.A. 1985. Grammar and Written Discourse: Initial vs. Final Purpose Clauses in English. *Text* 5 (1-2): 55-84.
- TURCO, G. & COLTIER, D. 1988. Des agents doubles de l'organisation textuelle, les marqueurs d'intégration linéaire. *Pratiques* 57: 57-79.
- VIRBEL, J. et al. 2005. Approche cognitive de la spatialisation du langage. De la modélisation de structures spatio-linguistiques des textes à l'expérimentation psycholinguistique: le cas d'un objet textuel, l'énumération. In J. BULLIER & C. THINUS-BLANC (eds), *Agir dans l'espace*. Paris: Éditions de la Maison des sciences de l'homme: 233-254.