



Discours

Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics

20 | 2017
Varia

Variation in the Marking of Text Organization in French Research Articles: From Short and Specific to Extended and Vague

Veronika Laippala



Electronic version

URL: <http://journals.openedition.org/discours/9301>

DOI: 10.4000/discours.9301

ISSN: 1963-1723

Publisher:

Laboratoire LATTICE, Presses universitaires de Caen

Electronic reference

Veronika Laippala, « Variation in the Marking of Text Organization in French Research Articles: From Short and Specific to Extended and Vague », *Discours* [Online], 20 | 2017, Online since 22 September 2017, connection on 31 January 2018. URL : <http://journals.openedition.org/discours/9301> ; DOI : 10.4000/discours.9301

Licence CC BY-NC-ND



Revue de linguistique, psycholinguistique et informatique

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

Variation in the Marking of Text Organization in French Research Articles: From Short and Specific to Extended and Vague

Veronika Laippala

Turku Institute for Advanced Studies – School of Languages and Translation Studies
University of Turku, Finland

.....
Veronika Laippala, «Variation in the Marking of Text Organization in French Research Articles: From Short and Specific to Extended and Vague», *Discours* [En ligne], 20 | 2017, mis en ligne le 22 septembre 2017.

.....
URL : <http://discours.revues.org/9301>

.....
Titre du numéro : *Varia*
Coordination : Shirley Carter-Thomas & Laure Sarda

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

 discours

 Presses
universitaires
de Caen

Variation in the Marking of Text Organization in French Research Articles: From Short and Specific to Extended and Vague

Veronika Laippala

Turku Institute for Advanced Studies – School of Languages and Translation Studies
University of Turku, Finland

.....
Adopting a text linguistic, corpus-based approach, this article studies variation in the marking of text organization. Why is text organization sometimes signaled very precisely, while sometimes signaling does not occur at all? The focus is on a particular mode of text organization, taking the form of text sequences, i.e. structures at least partially signaled by markers of addition or order, such as *first*, *the last example*. The material consists of 90 research articles in French with manually XML-annotated text sequences (XML = Extensible Markup Language). The results highlight the variation in the marking and show several factors affecting it. In shorter sequences, the marking is typically explicit and precise, while in longer ones explicit marking is more often omitted; when used, only vague markers, signaling simple addition, are present. In addition, different markers tend to be used in the signaling of sequences of different lengths.

Keywords: marking of text organization, text sequences, enumerative structures, corpus linguistics, discourse annotation

.....
En adoptant une approche textuelle et quantitative, cet article examine la variation du marquage de l'organisation textuelle. Pourquoi l'organisation textuelle est-elle parfois indiquée très précisément, alors que parfois elle n'est pas signalée du tout ? L'article se concentre sur un mode particulier de l'organisation textuelle, les séries linéaires, qui sont des structures dont les items sont au moins partiellement signalés par des marqueurs d'addition, d'ordre, ou de progression, tels que d'abord, le dernier exemple. Le matériel consiste en 90 articles de recherche en français, annotés en XML (Extensible Markup Language). Les résultats soulignent le rôle de la variation dans le marquage et montrent que plusieurs facteurs textuels entrent en jeu. Dans des séries linéaires plutôt courtes, le marquage est typiquement explicite et précis, tandis que dans des séries plutôt longues, le marquage est plus souvent absent ou seulement des marqueurs vagues sont utilisés. De plus, la longueur de la série linéaire a un effet sur le type de marqueur utilisé.

Mots clés: *marquage de l'organisation textuelle, séries linéaires, structures énumératives, linguistique de corpus, annotation discursive de corpus*

.....
Acknowledgements: I would like to thank the two anonymous reviewers for their valuable comments, and the Finnish Cultural Foundation for funding.

1. Introduction

1 A text is often conceived of as a functional unit, in which individual sentences form a coherent entity and are linked by semantic relations (see Halliday & Hasan, 1976 and 1989; De Beaugrande & Dressler, 1981; Charolles, 1997 and 2006). While these semantic relations do not need to be explicitly signaled, such signaling may

occur in the form of various cohesive devices, i.e. linguistic items referring to other parts of the text or indicating to the reader how the sentences are related. These items can also be approached as markers of text organization. When used, they help the reader to interpret the text and its structure (Mauranen, 1993: 161-163; Goutsos, 1996: 505; Charolles, 1997: 2). Empirical tests also indicate that these markers accelerate the reading process (Bestgen & Vonk, 1995; Sanders & Noordman, 2000).

- 2 In previous work, the marking of text organization has often been studied in terms of the semantic relations underlying the markers (see e.g. Mann & Thompson, 1988; Asher & Lascarides, 2003), of the individual markers (e.g. Vergez-Couret, 2009; Hansen, 2005; Bras & Le Draoulec, 2009), or of specific groups of signals, observed either in small corpora or in invented examples (Schneidecker, 2000 and 2006; Martin, 1992). These approaches, however, leave many questions unanswered. For instance, as Halliday (2002: 45) points out, it is only in combination that the markers of text organization contribute to the formation of a text. In addition, restricted corpora or invented examples do not necessarily reflect text organization and its marking in practice, “in their natural context” (Biber et al., 1998: 1 et passim).
- 3 In this article, I discuss variation in the marking of text organization in research articles (RAs) in French. Adopting a holistic, corpus-based approach, my purpose is to identify the factors affecting the presence or absence of marking and its degree of specificity. Why is text organization sometimes signaled very precisely, with markers such as *second* or *finally*, while sometimes signaling does not occur at all, or takes the form only of vague signals such as *in addition* or *moreover*? I focus on a particular mode of text organization, taking the form of *text sequences*, i.e. structures that organize text into linearly ordered items that are at least partially signaled by various markers of addition, progression or order, such as *first*, *the second example*, *finally* (see Jackiewicz, 2005; Afantenos et al., 2012). Most often, these structures organize intraparagraphic text segments, although longer sequences occur as well, the longest ones covering several sections of the RAs.
- 4 In order to focus on language use and bring to bear a global perspective on text organization and its marking, I applied a functional, corpus-based approach, taking into account all kinds of markers used in the signaling of text sequences as well as the alternation of absence and presence of markers. The corpus consisted of 90 original RAs (~ 800,000 words). RAs form an interesting context for the study of text organization. As these texts aim at convincing the reader (Hyland, 1998: 438-439; Markkanen & Schröder, 1997: 9), text organization and its marking are crucial; a badly organized text is not convincing. The articles in my corpus were all published in peer-reviewed journals, in the fields of linguistics, education and history, between 2000 and 2006. Each discipline-specific sub-corpus included articles from 5-6 journals, and the average lengths of the articles varied between the shortest (6,396 words) for the articles in *Cahiers de grammaire* to the longest (11,440 words) for the ones published in *Annales historiques de la Révolution française*.

In the selection process, articles that reported their own research and articles that were not very interdisciplinary were strongly preferred¹.

5 The corpus was manually XML-annotated for all text sequences, items and possible item markers (XML = Extensible Markup Language). To study the alternation of explicit marking and its absence, sequence items without an explicit marker of addition or order were also taken into account. In addition, the analysis included possible *predictions* preceding the actual sequence (e.g. *This sequence consists of two items...*) and *closures* following it (e.g. *These two items...*) (see Tadros, 1994; Francis, 1994).

2. Defining text sequences

6 Early studies on the signaling of text sequences in French concentrated on the use of temporal, spatial or locative adverbial expressions which can function as sequence item markers, such as *premièrement* (“first”), *en deuxième lieu* (“in the second place”) and *enfin* (“finally”) (see Turco & Coltier, 1988; Adam & Revaz, 1989; Schneuwly et al., 1989). These studies give a detailed description of the serial use of these connectives. Since, however, the text sequences included in my study can be introduced by any marker indicating addition, progression or order, not only adverbial expressions, the sequences in my corpus take various forms and may be very different from the canonical text sequences examined in these earlier studies.

7 Many studies have also approached text sequences as structures that consist of equal, ordered items (see also Jackiewicz, 2005: 107; Jackiewicz & Minel, 2003; Adam & Revaz, 1989: 73). In practice, however, this approach has been shown to be very restrictive (Péry-Woodley, 2000: 138; Porhiel, 2007: 107-109) because of the heterogeneity of sequences, also reported e.g. by Hempel and Degand (2008). In my study too, such factors as the variation in item marking and length, as well as the potential absence of marking, call for a more global definition.

8 A cognitively motivated approach to text sequences has been offered by the model of sequential relations proposed by Goutsos (1996), also adopted in the “Annodis Project” (Afantenos et al., 2012; Ho-Dac et al., 2010). This model is based on the contrast between the writer’s mental representation of the discourse and the linear format of the text and describes text sequences as structures that organize text in successive elements according to a text strategy².

9 In this study, as explained in Section 1, I define text sequences as structures that organize text into ordered items that follow one another linearly and that are at least partially signaled by various markers of addition, progression or order. This definition is broad enough for my purposes and covers all the text sequences included

1. The article is based on my doctoral dissertation. For more detailed information on the corpus collection, see Laippala (2011).

2. *Text strategy*, according to Virtanen (1992: 51), is defined as “the aggregate of all decisions concerning textualization that the text producer makes in view of a communicative goal” (see also Enkvist, 1987).

La volonté du roi de France de conserver les ordonnances anciennes, bonnes et raisonnables ne peut être ignorée tant elle est rappelée dans de nombreux actes .	Prediction
<u>Philippe VI</u> évoque ainsi les <i>plura statuta et arresta pro bono statu regni</i> de ses	1
prédécesseurs dont il s'inspire pour interdire [...]. De même , Jean II confirme	2
le 1 ^{er} juin 1363 des lettres prises par son père [...]. <u>Dans une ordonnance</u> sur le statut des officiers royaux datant du 5 février 1389, Charles VI souligne, lui aussi , sa volonté de [...].	3
The King of France's desire to conserve the good old ordinances cannot be	Prediction
ignored, as it is repeated in many statutes . <u>Philip VI</u> evoked for example the	1
<i>plura statuta et arresta pro bono statu regni</i> of his predecessors on which he based	2
the ban [...]. Similarly , on the first of June 1363 John II confirmed his father's letters [...]. <u>In an ordinance</u> on the status of royal officers dated February 5 th 1389, Charles VI likewise stressed his desire to [...].	3

Table 1 – A text sequence from the history sub-corpus³

in the analysis. As I take into account all kinds of markers used in the signaling, the sequences can be very heterogeneous. The following sequence in Table 1 illustrates this property. The sequence consists of three items, of which only the second and third items are explicitly signaled; the second by the adverb *de même* (“similarly”) and the third by the adverb *aussi* (“too”). In addition, the sequence is preceded by a *prediction* which already envisages the organization of the following text by committing the writer to a discourse act in the following text (Tadros, 1994: 70; also Sinclair, 1996).

- 10 In the text sequence in Table 1, according to my criteria, the first item has no explicit marker, although the repetition of kings' names does contribute to its identification and the item is introduced by *ainsi* (“so”). This is because these devices do not indicate addition or order, which under my definition of text sequences is crucial (see Section 3 for details).

3. Methods: annotating text sequences in practice

- 11 Corpus analysis of text sequences requires their reliable identification in the running text. In this study, I decided to apply linguistic surface criteria based on minimal signaling and length of the text sequences included in the analysis. Criteria based on minimal marking aim at excluding structures where the items are not ordered, such as text segments organized according to Danes' model (1974) of thematic structure, while those based on minimal length aim at excluding single-word lists that operate at the level of syntax rather than at that of text. Without a minimum length criterion, for instance the sentence “There are two cats on the street, a black cat and a white one” could be interpreted as a sequence consisting of a prediction

3. Numbers, bolds and highlighting added. The beginnings of the items are underlined and explicit markers are shown in bold. The translations are my own.

two cats and of two items unmarked for addition or order. In addition, in order to narrow down the data, I did not count as text sequences structures signaled by numbers, dashes or other non-lexical signals, included in the analysis for example in the “Annodis Project” (Afantenos et al., 2012) or in the “Textual Architecture Model” (Luc, 2001).

12 Although especially minimal marking criteria may exclude from the analysis some structures that would seem to be very similar to the sequences included, I consider this approach the only reliable one. In addition, the surface criteria ensure that the results are quantifiable and comparable to other corpora. The criteria are the following:

1. The minimum length is two sentences, starting with a capital letter and ending with a period, exclamation mark or question mark. Possible predictions and closures are included. There is no maximum length.
2. At least one sequence item has to be explicitly signaled by a marker of addition, progression or order. If the sequence includes only one item marker and it is *vague*, such as *de plus* (“in addition”) or *aussi* (“also”) which do not specify the exact position of the item in the list, the sequence must be preceded or followed by a prediction or a closure. If the only item marker is *precise* and specifies the exact position of the item in the list, such as *premièrement* (“first”) or *enfin* (“finally”), predictions and closures may occur but they are not mandatory.
3. If none of the sequence items are signaled by an explicit marker of addition, progression or order, the sequence needs to be preceded or followed by a prediction or closure defining the number of sequence items (see above for an example).

13 It should be noted that in this study I consider as explicitly marked sequence items only those introduced by a linguistic form indicating addition, progression or order; others are referred to as unmarked. Unmarked ones may include a cohesion device helping to identify the sequence item, such as the repetition of the various kings’ names in Table 1. Repetition, however, does not indicate addition or order, which under my definition of text sequences is crucial. It is therefore not counted as an explicit signal.

14 To enable the quantitative study of the sequences and their markings, the structures satisfying the above annotation criteria were manually XML-annotated in the data. This procedure also allowed the quantification of the annotated information, i.e. the sequences, sequence items, their possible markers and marker types, and possible predictions and closures. Without annotation this would not have been possible, as the purpose of the study was to identify all linguistic instances used in signaling and not merely a list of predefined markers that could be searched for in the data similarly to individual words. In addition, only XML-annotation allows the identification of sequence items without an explicit marker of addition or order.

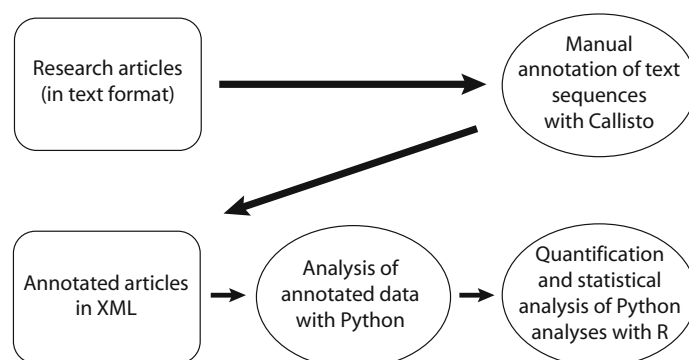


Figure 1 – Description of the method

- 15 The annotation was performed by the author, using the annotation tool Callisto (<http://www.softpedia.com/get/Office-tools/Other-Office-Tools/Mitre-Callisto.shtml>). Difficult cases were discussed with another specialist in the subject. The resulting XML-documents were then analyzed with brief Python scripts, also created by the author, to list and classify the annotated sequences in the six sequence categories presented and analyzed in Section 4. Finally, the quantitative study of these features was performed with R, a software product for statistical computing (<http://www.r-project.org/>). The process is shown schematically in Figure 1.

4. Results

- 16 The final data set consists of a total of 618 text sequences satisfying the criteria given in Section 3. In other words, each RA in the corpus contains on average 6.8 text sequences, approximately one text sequence per every 1,288 words, or 7.8 sequences per 10,000 words. The average length of the sequences is 271.2 words and the median 125 words. The majority of the structures are thus relatively short, around one paragraph. However, longer structures do occur as well, 23 of the sequences consisting of several research article sections with headings. This naturally increases the average length.

- 17 In order to further analyze the marking of the text sequences and the factors affecting it, I divided the 618 text sequences into six categories according to the presence or absence of the markers and the kinds of markers used in their signaling. These categories and their proportions in the entire data set are the following:

- Sequences with both explicitly signaled and unmarked items, 44% of the aggregate of the sequences (see Table 1).
- Sequences with all items explicitly signaled with the same marker type, 31% of the aggregate of the sequences. These sequences resemble the classic examples of text sequence marking, introduced by the same

marker type, such as adverbs (e.g. *d'une part/d'autre part* ["first"/"second"]), anaphoric expressions (e.g. *l'un/l'autre* ["one"/"the other"]) or constructions composed of an ordinal number and a noun (e.g. *le premier exemple* ["the first example"]) functioning as the sentence subject or object (see Table 3).

- **Sequences with all items explicitly signaled but with different marker types, 5% of the aggregate of the sequences.** In these sequences, the items are signaled by a combination of different marker types, such as adverbs, anaphoric expressions and constructions consisting of an ordinal number and a noun functioning as the sentence subject or object (see Table 6).
- **Sequences with no explicit markers introducing the actual sequence items, 8% of the aggregate of the sequences.** These sequences do not include any actual markers of addition or order; the number or order of sequence items is only indicated before or after the actual sequence by a prediction or closure (see Table 4).
- **Sequences where the only explicit marker is a coordinating conjunction, 8% of the aggregate of the sequences.** These sequences do not include any actual markers of addition or order except for coordinating conjunctions; the number or order of sequence items is only indicated before or after the actual sequence by a prediction or closure.
- **Sequences where all items correspond to text sections with headings, 4% of the aggregate of the sequences** (see Table 5).

18 Hempel and Degand (2008: 690) note that the signaling of text sequences is generally speaking relatively diverse. The proportions of the different text sequence categories in my data support this, as the sequences with both explicitly signaled and unmarked items cover 44% of the aggregate of the material, and yet another 5% of the sequences are signaled by a combination of different marker types. Sequences with all items explicitly marked with the same marker type, which resemble the classic examples of text sequences, cover only 31% of my material.

19 The analysis of these six text sequence categories reveals that their lengths vary. The lengths are presented in Table 2, showing that three sequence categories are clearly shorter than the others: the sequences with only coordinating conjunctions as explicit markers, the sequences with no explicit markers, and the sequences with all items signaled by the same marker type. The sequences with both explicitly signaled and unmarked items and the sequences with different marker types are longer. Quite naturally, the sequences composed of entire sections with headings are noticeably the longest.

20 The variation in the text sequence lengths across the six sequence categories suggests that the length of the text sequence affects its marking. In addition to this, two other factors affecting the marking stand out from the analysis: the degree of precision with which the whole sequence is signaled, and the morpho-syntactic role of the markers used in the signaling. In the following sections, I will focus

	Median length (words)	Average length (words)
CC only explicit marker (N = 49), 8%	74	76.8
No explicit markers (N = 47), 8%	98	144.9
All items signaled by the same marker type (N = 192), 31%	105	179.2
Explicit and unmarked items (N = 272), 44%	149	233
All items explicitly signaled but with different marker types (N = 35), 5%	175	255
Sequences composed of sections (N = 23), 4%	1,458	2,187

Table 2 – The lengths of the six text sequence categories in the data
(CC = coordinating conjunction)

on analyzing these factors. As the basis of the analysis, the six sequence categories were divided into *short* and *long*, the former including the three shortest sequence categories and the latter the three longest ones⁴.

4.1. Short and precise

21 The short sequences include three text sequence categories, presented above in Section 4:

- Sequences with all items explicitly signaled with the same marker type (see Table 3).
- Sequences with no explicit markers introducing the actual sequence items (see Table 4).
- Sequences where the only explicit marker is a coordinating conjunction.

22 The signaling of these short sequences is very explicit and precise. The sequence in Table 3 illustrates this. In addition to having all the items explicitly introduced, the markers used are precise, indicating the position of the items in the list; the reader knows that *d'une part* (“first”) initiates the first item and *d'autre part* (“second”) the second one. Moreover, the sequence structure is repeated before the actual sequence by the prediction *ce pour deux raisons* (“for two reasons”). It announces that the following text will be composed of two items, each giving one reason for the disappearance of missions intended to supply Paris during the autumn of 1793. Finally, the end of the sequence is also indicated explicitly by the connective *donc* (“therefore”), declaring the consequences of these events.

4. The difference between the lengths of these two groups of sequences is statistically significant with Kruskal-Wallis, $df = 1$, p -value $< 2.2e-16$.

En revanche, au cours de l'automne 1793, se produit progressivement une quasi-disparition des missions spécifiques destinées au ravitaillement de Paris, ce pour deux raisons . D'une part , un décret du 1 ^{er} brumaire an II (22 octobre 1793) crée une Commission des subsistances, chargée initialement de veiller à l'exécution du Maximum des prix des grains voté le 11 septembre 1793, renforcé par le Maximum général des prix et salaires le 29 septembre. Celle-ci élargit très vite ses attributions [...]. D'autre part , les représentants du peuple envoyés dans les départements pour des missions à «généralistes» se substituent alors souvent à leurs prédécesseurs pour veiller aux approvisionnements, [...] y compris, le cas échéant, le ravitaillement de Paris. Il faut donc attendre les lendemains de Thermidor pour voir réapparaître les missions spécifiquement réservées à cet approvisionnement de la capitale [...].	Prediction
	1
	2
	Closure
However, during the autumn of 1793, the missions intended to supply Paris gradually almost disappeared. This happened for two reasons . On the one hand , the decree passed on 1 st Brumaire year II (October 22, 1793) appointed a Subsistence Commission, whose task was initially to supervise the execution of The Law of the Maximum on the price of grain. The law was passed on 11 September 1793 and reinforced by the General Maximum on prices and salaries on 29 September. The Commission, however, quickly widened its functions [...]. On the other hand , the representatives of the people sent to the departments for "all-inclusive" missions often took their predecessors' place and supervised the provisioning [...], including the supplying of Paris. Therefore , it was only after Thermidor that the missions specifically reserved for supplying the Capital reappeared [...].	Prediction
	1
	2
	Closure

Table 3 – Text sequence from the history sub-corpus

23 The structure of the text sequence in Table 3 is very explicitly signaled: both items start with a connective. In addition to this, the structure is repeated by the prediction *pour deux raisons* ("for two reasons") preceding the actual items. In fact, this repetition is very frequent within sequences where all items are explicitly signaled, 39% of them being preceded **and/or followed** by a prediction or closure indicating the number of sequence items. Instead of replacing the actual sequence marking, these components thus seem to highlight the organization that is also indicated within the sequence itself⁵.

24 Also the use of *precise* markers indicating the position of the item in the list, such as *d'une part* ("first") and *d'autre part* ("second") in Table 3, are in fact very frequent in this sequence category, as illustrated in Figure 2. The diagram shows the respective frequencies of *precise* markers and *vague* markers which signal simple addition and do not specify the position of the item in the list, such as *in addition*. The contrasted text sequence categories are sequences with all items signaled by the same marker type, which is a sequence category with short sequences, and the three longer sequence categories (see Section 4). The other sequence categories belonging to the short sequences, sequences with no explicit markers, and those where the only explicit marker is a coordinating conjunction, are not included in the figure because they do not include any markers to be analyzed.

5. The exact percentages for all text sequence categories are given in Figure 3 in the next section.

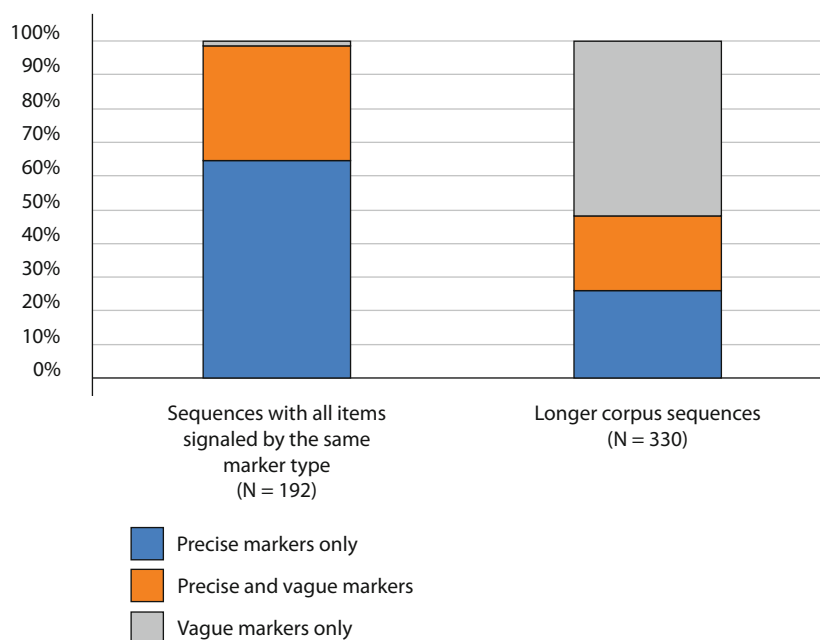


Figure 2 – Precise and vague markers in sequences with all items explicitly signaled, which are also short, and in longer corpus sequences (the latter includes the sequences with both explicitly signaled and unmarked items, sequences with different marker types and sequences composed of sections; the lengths are described in Table 2)

25 As Figure 2 indicates, the usage of precise markers in sequence marking alone is clearly more frequent in the sequences with all items signaled by the same marker type, which are also short, than in longer corpus sequences. While the majority of the short ones are signaled by precise markers only, the majority of the longer ones are signaled by vague markers. The difference between the frequencies of these markers is also statistically significant⁶.

26 Our results thus suggest that the marking of the text sequences with all items explicitly signaled by the same marker type is very precise: in addition to having all the items explicitly introduced, their structure is often repeated by a prediction or closure, and the markers used are typically precise and indicate the exact position of the item in the sequence. Also the other short text sequence categories, i.e. the sequences with no explicit markers and the sequences where the only explicit marker is a coordinating conjunction, are easily identified in the text, despite the lack of explicit item markers. This is illustrated by the sequence in Table 4 which consists of two unmarked sequence items and a prediction (see Tadros, 1994).

6. χ^2 , X^2 -squared = 259.49, $df = 3$, p -value < 2.2e-16.

À ce stade, deux hypothèses peuvent être avancées : Est-ce le côté inhabituel ou «exotique» du jeu «mât de cocagne» qui a retenu toute l'attention des élèves en difficultés: Est-ce le fait d'y avoir réellement joué? On peut au moins affirmer une chose: c'est bien parce que ce jeu n'était pas connu des élèves de maternelle, qu'après l'indispensable explicitation orale puis écrite (élaboration d'une fiche technique), la maîtresse leur a proposé de l'appréhender en y jouant réellement.	Prediction, 1
	2, closure
At this stage, two hypotheses can be proposed : Is it the exoticism or the rarity of greasy-pole climbing that captured the attention of disadvantaged pupils? Is it the fact of having gotten to actually play it? At least we can claim one thing: it is because the kindergarten pupils were not familiar with the game before the teacher's explanations that the teacher suggested that they should actually play it.	Prediction, 1
	2
	Closure

Table 4 – A text sequence from the education sub-corpus with a prediction indicating the number of sequence items (the actual items are not explicitly signaled)

27 As defined by the annotation criteria in Section 3, the sequences with no explicit markers and the sequences with coordinating conjunctions as the only explicit markers are always preceded and/or followed by an expression indicating the number of sequence items. This may also explain the lack of explicit item markers: as these sequences are short and their structure is already indicated by the prediction/closure, explicit item marking could be redundant (see Grice, 1975). In addition, a manual analysis of the predictions and closures attached to these sequences reveals that one of them always expresses the items' semantic category, telling the reader what kind of items to look for. For instance, in Table 4, the prediction announces that items correspond to *deux hypothèses* ("two hypotheses"). This specification obviously increases the predictions' cohesive effect and further renders explicit item marking redundant (see Francis, 1994, for an extensive analysis). Moreover, the repetition of lexical or syntactic patterns is very frequent within the actual items of these sequences. For instance, the two items in Table 4 begin by repeating the interrogative clause *est-ce X* ("is it X"), which also creates strong syntactic and lexical cohesion between the items (see e.g. Halliday & Hasan, 1976; Hoey, 1991; Tanskanen, 2006). According to the manual analysis, this kind of repetition of entire phrases in each item is found in 29% of the sequences either with no explicit markers or with only coordinating conjunctions. This repetition may also contribute to the absence of the actual item markers.

4.2. Long and vague

28 The longer corpus sequences belong to three sequence categories, already presented above in Section 4:

- Sequences with both explicitly signaled and unmarked items, 44% of the aggregate of the sequences (see Table 1).
- Sequences with all items explicitly signaled but with different marker types, 5% of the aggregate of the sequences (see Table 6).

- Sequences where all items correspond to text sections with headings, 4% of the aggregate of the sequences (see Table 5).

29 The lengths of these sequences are shown in Table 2. Obviously, sequences consisting of entire sections with headings are the longest. Those with different marker types are longer than those with both marked and unmarked items. Other corpus sequences described in the previous section are clearly shorter. As already mentioned, the difference between the lengths of these and other corpus sequences is statistically significant.

30 The sequence in Table 1 already illustrates sequences with both explicitly signaled and unmarked items. In addition to the presence of both marked and unmarked items, it also features other characteristics typical of all longer sequences. As noted in the previous section, the number of sequence items is specified by a prediction, by a closure or by both, in 39% of the sequences with all items marked. In these three longer sequence categories this is less frequent, nor does it occur in the example in Table 1, although the prediction does indicate the beginning of a list. Figure 3 shows the percentages of sequences with the sequence item number specified before and/or after the actual sequence by a prediction or a closure, in the entire corpus. In sequences with both marked and unmarked items, the item number is specified in only 17% of sequences, in those with different marker types in 29%, and in sequences with sections in 17%. The difference in frequencies between these three sequence categories and the shorter sequence category with all items explicitly signaled is also statistically significant⁷.

31 In addition to the rarity of predictions or closures indicating the sequence item number, *vague* markers indicating simple addition, such as *in addition*, are also characteristic of the marking of the long text sequences. In the text sequence in Table 1, both of the explicit markers are vague: *de même* (“similarly”) and *aussi* (“too”). As opposed to short sequences with all items explicitly marked, in the longer sequences the use of precise markers, such as *d’abord* (“to begin with”), is relatively infrequent. As shown in Figure 4, these vague markers are highly dominant in sequences with marked and unmarked items, while the majority of sequences with different marker types are signaled with both precise and vague markers. Sequences composed of sections form an exception to the longer sequences, being mostly introduced by precise markers.

32 Longer sequences thus seem to be signaled less explicitly than shorter ones; they include unmarked items as well as frequent vague markers indicating simple addition. Furthermore, the number of items in the sequence is rarely specified before or after the actual sequence, via a prediction or a closure, giving the reader the task of interpreting the text structure.

33 One potential reason for the absence of marking of items in the longer sequences might be the use of strong cohesive devices. For instance, the repetition of entire phrases in each item, such as that of the interrogative *est-ce X* (“is it X”) in the items

7. χ^2 , X^2 -squared = 108.2192, $df = 1$, p -value < 2.2e-16.

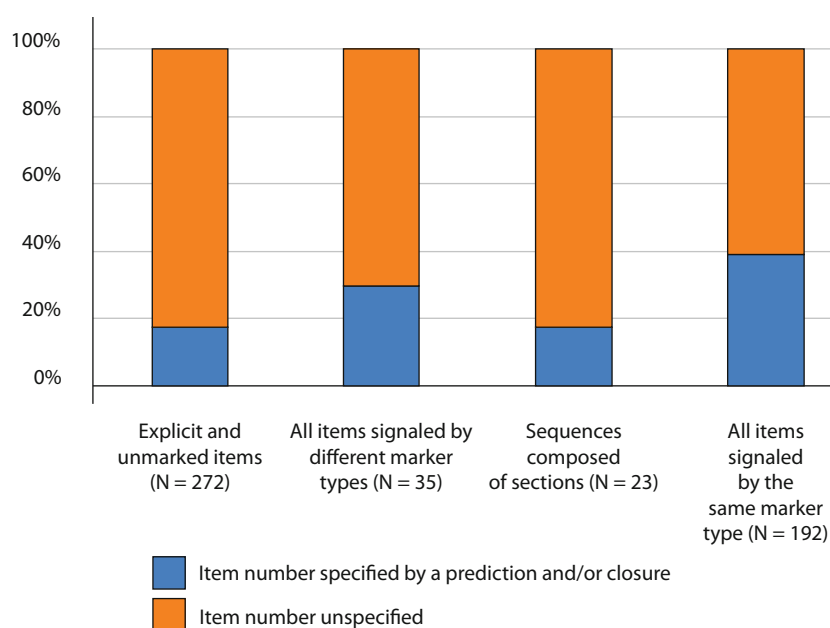


Figure 3 – Distribution of predictions and/or closures specifying the number of sequence items by text sequence category (the blue bar refers to the sequences where the item number is specified by a prediction, by a closure or by both, while the red bar refers to the sequences where the item number is not specified at all)

in Table 4, might render the use of explicit markers unnecessary. This, however, is not the case: the manual analysis of 80 sequences with marked and unmarked items indicates that these structures do not include repetition or other strong cohesive devices that could explain the absence of explicit markers. Rather, these sequences seem to be characterized by the absence of such devices: for instance repetition or structural similarity of the items is less frequent than in other structures with explicitly signaled items.

34 In the sequences consisting of entire sections, the headings already segment and organize the text (see Sullet-Nylander, 1998; Virbel, 2002; Ho-Dac et al., 2004; Jacques & Rebeyrolle, 2006; Rebeyrolle et al., 2009) thereby creating cohesion. In this sense, this sequence category differs from the other five categories, and this might also render the explicit marking of the sequence items unnecessary. Figure 5 presents the kinds of markers used in the signaling of sequences composed of sections. It shows that sequences composed of sections are very similar to other long text sequences; most of them (70%) include both marked and unmarked items (for a similar result see Ho-Dac et al., 2012). As such, they are thus not necessarily different from ones without headings. This, however, requires further analyses that are outside the scope of this article.

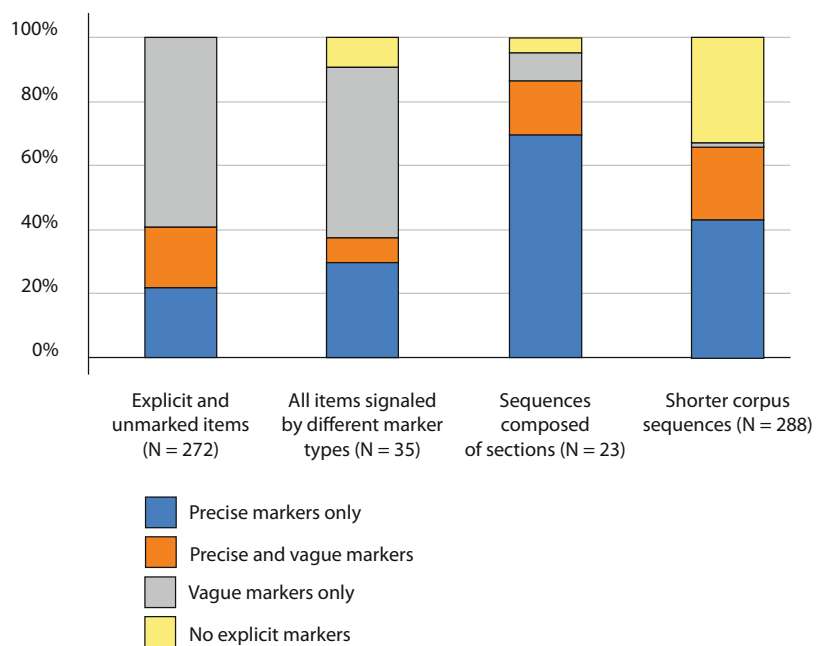


Figure 4 – Precise and vague markers in the three longer sequence categories and in the shorter corpus sequences (the latter includes the three shorter sequence categories presented in Section 4)

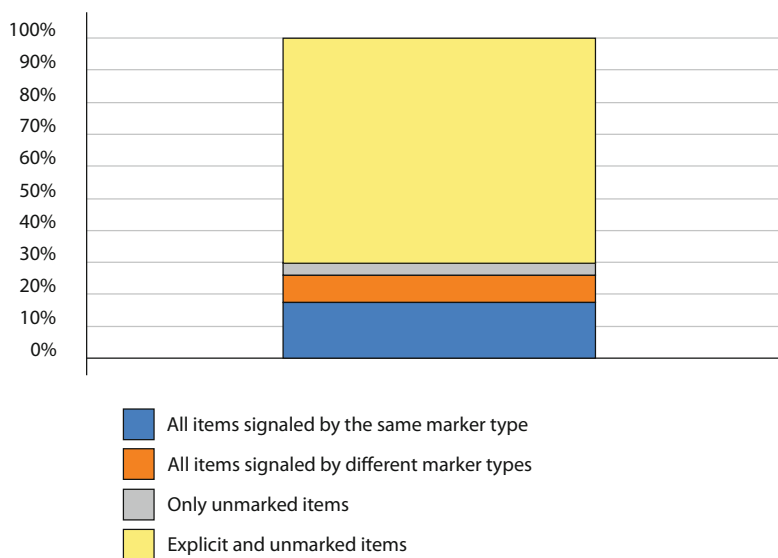


Figure 5 – Markings of sequences composed of sections (N = 23)

<p>Dans la section suivante, nous proposons une nouvelle batterie de tests, plus élaborés cette fois, visant à donner quelques raisons suffisantes pour traiter les exemples ci-dessus comme des cas d'ellipses.</p> <p>3.3. Seconds diagnostics: VPE <i>vs</i> anaphore nulle Déterminer si on est en présence d'une ellipse, dans le sens d'un matériel syntaxique/sémantique qui a été supprimée, ou d'une anaphore (nulle) n'est pas chose facile. Néanmoins, il semble exister au moins trois tests permettant de mettre au jour cette distinction. De manière générale, ces tests tendent tous à souligner la présence d'un constituant syntaxique dans la proposition cible. [...]</p>	Prediction
<p>3.3.1. Recyclage d'un antécédent Une différence fondamentale entre ellipse et anaphore (nulle) concerne la capacité de «recycler» un matériel absent dans l'antécédent. Comme le signalent Grinder & Postal (1971), seule l'ellipse verbale possède cette propriété [...].</p>	1
<p>3.3.2. Les constructions ACD (Antecedent Contained Deletion) Comme nous l'avons déjà souligné plus haut à § 2 (cf. Abeillé, 1991), l'ellipse modale partage avec la VPE anglaise, le fait d'être licite dans les constructions relatives; en particulier, dans les constructions où le SV élidé est contenu dans le SV antécédent. [...]</p>	2
<p>3.3.3. Le Pseudo-Gapping Le dernier test est un peu moins direct, puisqu'il se fonde sur un exemple de Pseudo-Gapping: [...]. Les résultats des tests ci-dessus sont résumés dans le tableau suivant [...].</p>	3
<p>Closure</p>	Closure
<p>In the following section, we propose a new, more advanced test battery designed to give sufficient cause for analyzing the above examples as ellipsis.</p> <p>3.3. Second diagnosis: VPE vs. null anaphora Distinguishing between null anaphora and ellipsis in the sense of deleted syntactic/semantic material is not easy. Still, there seem to exist at least three tests that help in making this distinction. Generally, all these tests aim at highlighting the presence of a syntactic element in the target clause. [...]</p>	Prediction
<p>3.3.1. Antecedent recycling A fundamental difference between ellipsis and (null) anaphora is the capacity of "recycling" material that is absent in the antecedent. As noted by Grinder & Postal (1971), only verbal ellipsis has this property [...].</p>	1
<p>3.3.2. ACD constructions (Antecedent Contained Deletion) As we have noted earlier in Section 2 (cf. Abeillé, 1991), modal ellipsis shares with VPE [<i>English verb-phrase ellipsis</i>] the capacity of being legitimate in relative constructions; in particular in constructions where the elided verbal subject is included in the verbal subject's antecedent. [...]</p>	2
<p>3.3.3. Pseudo-Gapping The last test is somewhat less direct since it is based on an example of Pseudo-Gapping: [...]. The results of these tests [...].</p>	3
<p>Closure</p>	Closure

Table 5 – Text sequence composed of whole sections with headings, from the linguistics sub-corpus

35 The sequence in Table 5 illustrates sequences composed of whole sections with headings. The sequence starts with a prediction, which first indicates to the reader that the following sections will enumerate tests enabling the distinction between null anaphora and verbal ellipsis. Later, the prediction also specifies the number of the tests. The first two items, i.e. subsections, do not include explicit markers of order or addition. Except for the headings, there are few other strong cohesive devices contributing to the identification of the items. Although for instance the words *antécédent* (“antecedent”) and *ellipse* (“ellipsis”) are repeated, they do not help to identify the items or distinguish the text structure. The last item, in contrast, starts with a noun phrase combining the noun *test*, already repeated in the prediction, and the adjective *dernier* (“last”), indicating that the item completes the list of tests as projected by the prediction.

4.3. Markers at local and global levels of text

36 The analysis has shown that the degree of precision with which the sequence is signaled and the frequency of explicit signaling vary according to the length of the text elements: in shorter text sequences the marking is often precise and explicit, while in longer texts the marking tends to be vague and implicit. Previous studies of variation in the marking of text organization at different levels of the text, however, have focused on the types of markers used. The relationship between different marker types and the length of the text sequence is discussed in this section as well.

37 The different types of markers analyzed are the following:

- anaphoric expressions not combined with a noun (*le premier, le second, l’un, l’autre*, etc. [“first”, “second”, “one”, “the other”, etc.]);
- adverbial connectives (*d’abord, ensuite, enfin, premièrement, de même*, etc. [“first”, “then”, “finally”, “in the first place”, “similarly”, etc.]);
- coordinating conjunctions;
- noun phrases consisting of an ordinal and a noun, NOT functioning as a syntactic adverbial in the sentence (*le premier + nom, l’autre + nom*, etc. [“the first + noun”, “the other + noun”, etc.]).

38 The marker types were manually annotated according to these four categories. The main difference between anaphoric expressions and noun phrases consisting of an ordinal and a noun is that the former never include a noun. Neither marker type can function as a syntactic adverbial; they are always syntactic subjects or objects of the sentence. Anaphoric expressions are by definition referential, but the noun phrases too are in practice often referential expressions, although this was not analyzed during the annotation. Adverbial connectives mainly consist of fixed connectives, such as the ones listed above, but also include some temporal, spatial and ordinal adverbials, such as *dans le premier exemple* (“in the first example”), *dans la troisième section* (“in the third section”).

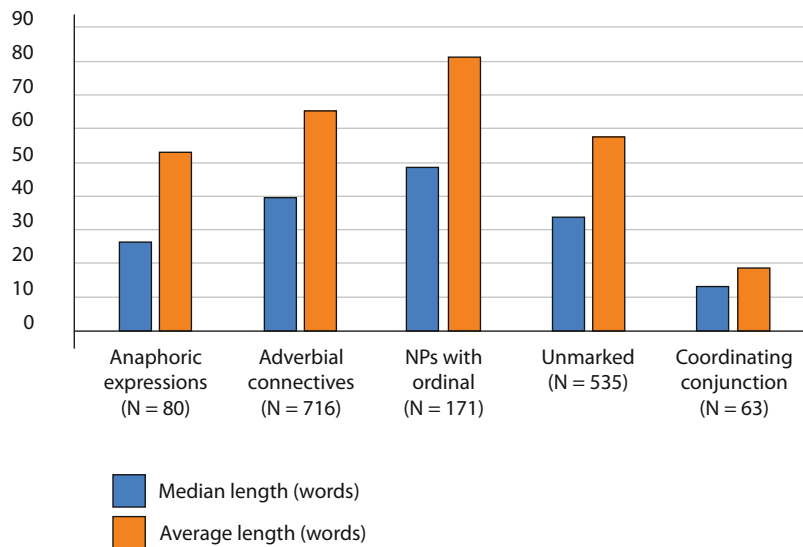


Figure 6 – Lengths of items according to marker type
(sequences consisting of entire sections with headings are excluded; NP = noun phrase)

39 Figure 6 shows the lengths of items signaled with different marker types. Items signaled by noun phrases consisting of an ordinal and a noun are the longest, those introduced by connectives the second longest. The shortest items start with a coordinating conjunction and the second shortest with an anaphoric expression. The differences between the lengths of these groups are statistically significant⁸.

40 The results for connectives and anaphoric expressions confirm those reported in previous studies. For instance Schnedecker (2000: 22) states that the antecedents of *le premier* (“first”) and *le second* (“second”) can usually be found no farther than two sentences away from their sources (see also Ariel, 1990, and Goutsos, 1996). Although the perspective of my study is not exactly the same, my findings support this result: the median length of items introduced by anaphoric expressions is only 26 words in my data. For adverbial connectives, the median length of items in my corpus is 39 words. Similar results are reported in previous studies, as many of them emphasize that connectives organize medium-length, most often intraparagraphic text segments (Ho-Dac et al., 2010).

8. With Kruskal-Wallis, between items marked with connectives/NPs with order: Kruskal-Wallis chi-squared = 7.2885, df = 1, p-value = 0.00694; between anaphoric expressions/connectives: Kruskal-Wallis chi-squared = 17.9901, df = 1, p-value = 2.221e-05; between coordinated conjunctions/anaphoric expressions: Kruskal-Wallis chi-squared = 22.9666, df = 1, p-value = 1.648e-06.

41 The longest sequences in my corpus, in addition to items corresponding to entire sections, are introduced by noun phrases consisting of an ordinal and a noun, functioning as the syntactic subject or object of the sentence (see e.g. the last item in Table 5). Unlike anaphoric expressions or connectives, these markers do not form a coherent, pre-defined marker category. Nevertheless, the results indicate that they have a clearly specified function in the marking of longer text sequences, and probably in the marking of global-level text organization in general. This finding is also in line with previous studies. The power of sentence-initial adverbs in text organization has been highlighted in many studies (e.g. Charolles, 1997 and 2005; Virtanen, 1992), and the present findings stress the fact that marking becomes less explicit and vaguer at global text levels; it is thus possible that sentence-initial adverbs are too strong and too explicit for this text function. In the light of the present results, it may be that noun phrases are less distinguishing and thus better adapted to the general style of marking used in longer text sequences. Naturally, however, these markers deserve a detailed study, not possible within the scope of this article.

5. Discussion

42 At least two points stand out from the analysis. First of all, the marking of text sequences varies significantly, especially according to the length of the organized elements. In shorter sequences the marking is typically explicit and precise: explicit signals are usually present, and they often indicate the precise position of the item in the list. In addition, the sequence structure is frequently repeated before or after the sequence by a prediction or a closure. In longer sequences, explicit marking is more often omitted; when used, only vague markers, signaling simple addition, are present. In the shortest sequences, however, explicit item markers are not used; instead, the sequence structure is revealed by a prediction or closure stating not only the number but also the semantic category of sequence items. The use of different marker types varies as well: anaphoric expressions are used in the shortest sequences, adverbial connectives in intermediate structures, and noun phrases attached to an ordinal in the longest sequences. This variation is shown in Figure 7.

43 The absence of explicit markers in the shortest sequences can easily be explained by the Gricean maxim of quantity (1975) and the risk of excessive repetition. The less explicit and vague marking of the longer ones, on the other hand, is more difficult to explain, especially as intuitively the reverse might be expected (this hypothesis is also proposed by Jackiewicz, 2005: 98). Besides, as the manual qualitative analysis demonstrated the exact opposite, the absence of explicit markers cannot be explained either by repetition or by other forms of lexical cohesion, which in the longer sequences would render explicit marking unnecessary. What the findings indicate is that, for whatever reason, writers simply choose to organize their texts less explicitly and precisely at a more global than at a local level.

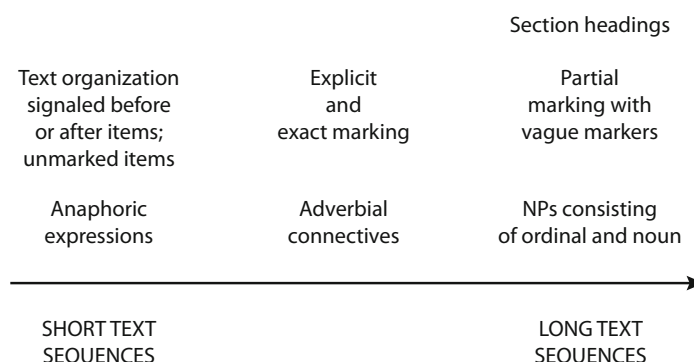


Figure 7 – Marking of text sequences at local and global levels of text

44 Second, the analysis highlights the high frequency of heterogeneous sequence features, often absent in traditional studies of the marking of text organization in French. In particular, these features seem to be frequent in longer text sequences. According to the analysis, different marker types are used in 5% of the corpus sequences, while another 44% include one or more unmarked items. In contrast, early studies on text sequences in French often focus on sequences consisting only of items signaled explicitly by the same kind of markers (see e.g. Adam & Revaz, 1989; Turco & Coltier, 1988). In consequence, structures with unmarked items and those signaled by different marker types are excluded from the analyses; they often do not correspond to classic definitions, or are treated as “heterogeneous”.

45 According to previous studies (Luc, 2001; Hempel & Degand, 2008; Péry-Woodley, 2000), other features associated with heterogeneous sequences include the absence of predictions and the inclusion of semantically or functionally different items or items of different lengths. A manual, qualitative analysis of the sequences with different marker types in my data indicates that, in addition to the mixing of marker types, these sequences also include further characteristics related to heterogeneity. There is variation in particular in the length and syntactic structures of the items, and the semantic categories of the items listed may change as well. The sequence in Table 6 offers a good illustration of these features. It starts by a prediction declaring the semantic category of the listed items: *différentes propriétés linguistiques* (“different linguistic properties”). This, however, is not repeated in the items. Instead, the first item is introduced by the adverbial connective *d’abord* (“to begin with”), followed by the noun *cas* (“cases”). The noun is also repeated in the second item combined with the adjective *autre* (“other”), indicating addition and the beginning of a new item. The third item is signaled by the adverbial connective *de même* (“similarly”). Items four, five and six, in contrast, repeat the adjective *autre* (“other”) but combine it with different nouns. The closure of the sequence refers to the items with yet another noun *exemples* (“examples”).

Je voudrais maintenant parcourir les différentes propriétés linguistiques que j'ai pu relever dans mon échantillon d'occurrences de la séquence « , <i>et</i> ». Il y a d'abord des cas où la virgule permet de distinguer des niveaux de parathésie, lorsque <i>et</i> est répété: [...]. Un autre cas de figure susceptible de motiver la présence d'une virgule devant <i>et</i> correspond à un décalage syntaxique entre les éléments conjoints. Par exemple dans: [...]. De même , les grammaires citées suggèrent que la virgule apparaît devant <i>et</i> quand le sujet change. Par exemple, dans Grevisse (1969, p. 1146): [...]. Un autre phénomène syntaxique semble jouer souvent, c'est la complexité ou la longueur du membre conjoint introduit par <i>et</i> . À ce propos, Drillon (1991, p. 180) cite Tassis (1859) qui, selon lui, «ne recommande la virgule que lorsque les propositions sont longues» [...]. Un autre critère semble jouer un rôle, c'est le statut prédicatif de l'élément introduit par <i>et</i> . Par exemple, en cas d'ellipse, la virgule est très courante: [...]. Ces exemples sont à la fois importants et intéressants parce qu'ils montrent, à mon sens, que la question n'est pas directement syntaxique: la virgule n'étant pas une marque linguistique, mais une marque du code graphique, ne peut pas en fait avoir une fonction syntaxique.	Prediction
	1
	2
	3
	4
I would now like to survey the different linguistic properties that I have been able to analyze in my sample of the occurrences of the sequence “ , <i>et</i> [<i>and</i>]”. There are first the cases where the comma enables the distinction of levels of bracketing when <i>et</i> is repeated: [...]. Another case likely to motivate the presence of the comma before <i>et</i> corresponds to a syntactic discrepancy between the coordinated elements. For instance in: [...]. Similarly , the grammar books studied suggest that a comma appears before <i>et</i> when the subject changes. For instance, in Grevisse (1969, p. 1146): [...]. Another syntactic phenomenon often having an effect is the complexity or the length of the coordinated element introduced by <i>et</i> . On this, Drillon (1991, p. 180) cites Tassis (1859), who, according to him, “recommends using a comma only when the clauses are long” [...]. Another criterion that seems to have an effect is the predicative status of the element introduced by <i>et</i> . For instance, with ellipsis the comma is very frequent: [...]. These examples are both important and interesting because they show, in my opinion, that the question is not directly syntactic: as the comma is not a linguistic but a graphic marker, it cannot have a syntactic function.	Prediction
	1
	2
	3
	4
5	
Closure	

Table 6 – Text sequence from the linguistics sub-corpus with items signaled with different marker types

46 The semantic categories of the nouns used in the sequence in Table 6 thus vary. This difference is further highlighted by the use of connectives for some items; as these are adverbials, and the *other + noun* constructions in the other items function as sentence subjects, the beginnings of the items are also structurally different. In addition, likewise typical of this sequence category, the item lengths vary. While the others cover approximately 150 words each, item four is 659 words long. This too works against the homogeneity of the items.

47 The sequence in Table 6 thus illustrates many of the features associated with heterogeneous sequences. According to the manual analysis, these are also frequent in other sequences with items signaled with different marker types. Despite this, the analysis also reveals that the structures are not necessarily difficult to read or

understand. The text in Table 6, for instance, is coherent and readable. This can perhaps be explained by the capacity of the reader to interpret the similarity between the items beyond the surface of the lexical item markers. Although the lexical items focus on different aspects of the listed elements (properties, case, syntactic phenomenon, criterion, example), while processing the text, the reader understands that they all refer to the same entity and the heterogeneous marking of the items does not matter (see Francis, 1994).

6. Conclusion

48 Focusing on RAs in French, the article has highlighted variation in the marking of text organization. Three factors have emerged which correlate with the marking. First of all, the use of explicit markers seems to co-occur with the use of precise markers and the repetition of the marking before and after the actual sequence as well. In contrast, the markers used in sequences with unmarked items tend to be vaguer and the structure is less often indicated elsewhere in the text. Second, while the structure of shorter sequences tends to be signaled explicitly with precise markers, that of longer ones is often vaguer and more implicit, with more responsibility assigned to the reader in the reading process. Third, different markers tend to be used in the signaling of sequences of different lengths: the shortest sequences are signaled by anaphoric expressions, medium-length ones by adverbials, and the longest ones by

49 While these findings as to the syntactic roles of marking support those of previous studies, those on the degree of precision and absence of explicit marking in longer sequences are more surprising and even counter-intuitive. Why is the marking more precise in shorter sequences and more implicit in longer ones? The question calls for further studies, both corpus-based analyses and studies on the cognitive aspects of language processing, in order to better understand how the marking system works. Also the relationship between non-lexical signals, such as numbers and dashes as item markers, and the linguistic markers examined in this study calls for further analyses. In the current study, non-lexical signals were present only in section headings in sequences composed of entire sections, and the analysis did not reveal significant differences in the marking of these sequences and the other long sequences. The number of sequences with headings was, however, very small and restricted, and a more extensive analysis on the relationship between lexical signals and linguistic markers as indicators of sequence items would be justified.

50 In addition, variation in the marking of text sequences across genres requires further work. Laippala (2011) compared the use of text sequences in the three disciplines included also in the corpus of the current article. She found differences in particular between the RAs in history and linguistics, while education falls between these two. In the history articles, text sequences were less frequent than in linguistics, and also the marking of the sequences was less explicit in history.

To complete these results, it would be interesting to compare also other disciplines, other languages and other marker types, signaling different semantic relations. Naturally, corpus studies of this type need to rely on manual annotation, a very time-consuming task, making their implementation more complex. The growing number of discourse-annotated corpora, however, may perhaps contribute to this end by providing data that can be adapted to answer these questions more easily than creating annotated corpora for this purpose alone.

References

- ADAM, J.-M. & REVAZ, F. 1989. Aspects de la structuration du texte descriptif: les marques d'énumération et de reformulation. *Langue française* 81: 59-98.
- AFANTENOS, S., ASHER, N., BENAMARA, F., BRAS, M., FABRE, C., HO-DAC, L.-M., LE DRAOULEC, A., MULLER, P., PÉRY-WOODLEY, M.-P., PRÉVOT, L., REBEYROLLE, J., TANGUY, L., VERGEZ-COURET, M. & VIEU, L. 2012. An Empirical Resource for Discovering Cognitive Principles of Discourse Organization: The ANNODIS Corpus. In *Proceedings of the 8th International Conference on Language Resources and Evaluation – LREC 2012*. Stroudsburg: Association for Computational Linguistics: 2727-2734. Available online: http://www.lrec-conf.org/proceedings/lrec2012/pdf/836_Paper.pdf.
- ARIEL, M. 1990. *Accessing Noun-Phrase Antecedents*. London – New York: Routledge.
- ASHER, N. & LASCARIDES, A. 2003. *Logics of Conversation*. Studies in Natural Language Processing. Cambridge – New York: Cambridge University Press.
- BESTGEN, Y. & VONK, W. 1995. The Role of Temporal Segmentation Markers in Discourse Processing. *Discourse Processes* 19 (3): 385-406.
- BIBER, D., CONRAD, S. & REPPEN, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge – New York: Cambridge University Press.
- BRAS, M. & LE DRAOULEC, A. 2009. *D'abord* marqueur de structuration du discours. *Journal of French Language Studies* 19 (2): 229-248.
- 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 M. AURNAGUE, M. BRAS, A. LE DRAOULEC & L. VIEU (eds.), *Proceedings of the Symposium on the Exploration and Modelling of Meaning – SEM-05 (Biarritz, November 14-15, 2005)*. 1-18. Available online: <http://www.lattice.cnrs.fr/IMG/pdf/FramingAdvDiscCohesionBiarritz.pdf>.
- CHAROLLES, M. 2006. Cohérence, pertinence et intégration conceptuelle. In P. LANE (ed.), *Des discours aux textes: modèles et analyses*. Mont-Saint-Aignan: Presses universitaires de Rouen et du Havre: 39-74.
- DANES, F. 1974. Functional Sentence Perspective and the Organization of the Text. In F. DANES (ed.), *Papers on Functional Sentence Perspective*. Prague – The Hague: Academia – Mouton: 106-128.
- DE BEAUGRANDE, R.-A. & DRESSLER, W.U. 1981. *Introduction to Text Linguistics*. London – New York: Longman.

- ENKVIST, N.E. 1987. A Note towards the Definition of Text Strategy. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 40 (1): 19-27.
- 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: 83-103.
- GOUTSOS, D. 1996. A Model of Sequential Relations in Expository Text. *Text* 16 (4): 501-533.
- GRICE, H.P. 1975. Logic and Conversation. In P. COLE et J.L. MORGAN (eds.), *Syntax and Semantics*. New York – London: Academic Press. Vol. 3: *Speech Acts*: 41-58.
- HALLIDAY, M.A.K. 2002. Text as Semantic Choice in Social Contexts. In J. WEBSTER (ed.), *Linguistic Studies of Text and Discourse*. The Collected Works of M.A.K. Halliday 2. London – New York: Continuum: 23-81. [Reprinted from: T.A. VAN DIJK & J.S. PETÖFI (eds.) 1977. *Grammars and Descriptions*. Berlin – New York: W. de Gruyter: 176-226.]
- HALLIDAY, M.A.K. & HASAN, R. 1976. *Cohesion in English*. English Language Series 9. London: Longman.
- HALLIDAY, M.A.K. & HASAN, R. 1989. *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*. Oxford: Oxford University Press.
- HANSEN, M.-B.M. 2005. A Comparative Study of the Semantics of *Enfin* and *Finement*, in Synchrony and Diachrony. *Journal of French Language Studies* 15 (2): 153-171.
- HEMPFEL, 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., FABRE, C., PÉRY-WOODLEY, M.-P., REBEYROLLE, J. & TANGUY, L. 2012. An Empirical Approach to the Signalling of Enumerative Structures. *Discours* 10. 1-27. Available online: <http://discours.revues.org/8611>.
- 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. & 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.
- HOEY, M. 1991. *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- HYLAND, K. 1998. Persuasion and Context: The Pragmatics of Academic Metadiscourse. *Journal of Pragmatics* 30 (4): 437-455.
- JACKIEWICZ, A. 2005. Les séries linéaires dans le discours. *Langue française* 148: 95-110.
- JACKIEWICZ, A. & MINEL, J.-C. 2003. L'identification des structures discursives engendrées par les cadres organisationnels. In *Actes de la 10^e conférence sur le Traitement automatique des langues naturelles – TALN 2003 (Batz-sur-Mer, 11-14 juin 2003)*. Available online: https://www.atala.org/doc/actes_taln/AC_0090.pdf.
- JACQUES, M.-P. & REBEYROLLE, J. 2006. Titres et structuration des documents. *Schedae*, preprint 1, vol. 1: 1-12. Available online: <http://www.unicaen.fr/services/puc/images/preprint0012006.pdf>.
- LAIPPALA, V. 2011. D'abord, ensuite, enfin et o, De plus – *Organisation textuelle par des séries linéaires dans les articles de recherche*. Thèse de doctorat. University of Turku, Finland. Available online: <https://www.doria.fi/handle/10024/72452>.

- LUC, C. 2001. Une typologie des énumérations basée sur les structures rhétoriques et architecturales du texte. *Verbum* 23 (1):103-123.
- MANN, W.C. & THOMPSON, S.A. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8 (3): 243-281.
- MARKKANEN, R. & SCHRÖDER, H. (eds.) 1997. *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts*. Research in Text Theory 24. Berlin: W. de Gruyter.
- MARTIN, J.R. 1992. *English Text: System and Structure*. Philadelphia: J. Benjamins.
- MAURANEN, A. 1993. *Cultural Differences in Academic Rhetoric: A Textlinguistic Study*. Frankfurt am Main – New York: P. Lang.
- PÉRY-WOODLEY, M.-P. 2000. *Une pragmatique à fleur de texte: approche en corpus de l'organisation textuelle*. "Habilitation à diriger les recherches" report. Université de Toulouse – Le Mirail.
- PORHIEL, S. 2007. Les structures énumératives à deux temps. *Revue romane* 42 (1): 103-135.
- 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.
- SANDERS, T.J.M. & NOORDMAN, L.G.M. 2000. The Role of Coherence Relations and Their Linguistic Markers in Text Processing. *Discourse Processes* 29 (1): 37-60.
- SCHNEDECKER, C. 2000. Ordres des ordinaux pronominaux. *Travaux de linguistique* 41: 7-34.
- 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: Duculot.
- SCHNEUWLY, B., ROSAT, M.-C. & DOLZ, J. 1989. Organismes textuels dans quatre types de textes écrits. Étude chez des élèves de dix, douze et quatorze ans. *Langue française* 81: 40-58.
- SINCLAIR, J. 1996. *Preliminary Recommendations on Corpus Typology*. Available online: <http://www.ilc.cnr.it/EAGLES/corpusyp/corpusyp.html>.
- SULLET-NYLANDER, F. 1998. *Le titre de presse. Analyses syntaxique, pragmatique et rhétorique*. Stockholm: université de Stockholm.
- TADROS, A. 1994. Predictive Categories in Expository Text. In M. COULTHARD (ed.), *Advances in Written Text Analysis*. London – New York: Routledge: 69-82.
- TANSKANEN, S.-K. 2006. *Collaborating towards Coherence: Lexical Cohesion in English Discourse*. Amsterdam: J. Benjamins.
- TURCO, G. & COLTIER, D. 1988. Des agents doubles de l'organisation textuelle, les marqueurs d'intégration linéaire. *Pratiques* 57: 57-79.
- VERGEZ-COURET, M. 2009. Le rôle de l'adverbe *notamment* dans la mise en œuvre des relations de discours. *Journal of French Language Studies* 19 (2): 249-268.
- VIRBEL, J. 2002. Éléments d'analyse du titre. In *ISLsp 2002 – Inscription spatiale du langage: structures et processus (29-30 janvier 2002)*. Toulouse: Institut de recherche en informatique de Toulouse (IRIT): 123-132.
- VIRTANEN, T. 1992. *Discourse Functions of Adverbial Placement in English: Clause-Initial Adverbials of Time and Place in Narratives and Procedural Place Descriptions*. Åbo: Åbo Akademi University Press.