

Corpus-based identification of temporal organisation in discourse

Anne Le Draoulec, Marie-Paule Péry-Woodley

ERSS, Université Toulouse Le Mirail

1. Introduction

We report here on a corpus-based study of a particular form of discourse organisation, inspired by the notion of temporal discourse frames. In Charolles' theory of Discourse Framing (Encadrement du Discours, Charolles 1997), a discourse frame is described as the grouping together of a number of sentences which are linked by the fact that they must be interpreted with reference to a specific criterion, realised in a frame-initial introducing expression. For instance, as regards perspective framing, *According to X*, ... provides an essential element for the interpretation of the proposition which follows, and also potentially of several subsequent propositions – as frame-introducing expressions are characterised by their ability to extend their scope beyond the sentence in which they appear (cf. Péry-Woodley 2000).

In the case which concerns us – temporal frames –, the introducing expressions are mostly adverbial expressions such as *today*, subordinate clauses such as *when he left*, and prepositional phrases such as {*in / until / since / about...*} 1989. Temporal expressions can occur in various places in the sentence, indeed they constitute a typical case of these mobile elements whose position (initial, median or final) has interested many generations of linguists (cf. Bally 1944, Firbas 1972, Givón 1979 *inter alia*). The best studied contrast is between initial and final position: a functional difference has been established between autonomous initial adverbials, which as adjuncts have a scene-setting role outside the proposition, and final adverbials which have no autonomy and express a circumstance only modifying the proposition. The distinction is also expressed in terms of syntactic integration in the sentence: initial position is equated with non-integration, final position with integration. Yet position is not the whole story as regards autonomy, or integration: there is a complex interplay between position and punctuational detachment (a comma is a visual indication of detachment which argues for non-integration). Combining position and punctuation yields the following possibilities: initial-detached, initial-non-detached, final-detached, final-non-detached. However, initial position is recognised as having a major cognitive impact (cf. Lambrecht 1994), and in any case, for this position, punctuation seems somewhat erratic. On the other hand, there is a real difference between a detached and a non-detached final adverbial: detachment goes against the integration associated with final position.

What is the relation between these classical theoretical accounts, and discourse framing? Charolles (1997, p.15) asserts that frame introducers are sentence adjuncts, and suggests that they most often appear in initial position. As his focus is on the elaboration of the notion of discourse framing, he does not go into a systematic analysis of the nature of introducers. His intuition about frame introducers seems perfectly well founded, as there is an obvious link between autonomy (or non-integration) of an expression, and its ability to open a frame: the characteristic of a frame introducer is indeed to scope over several sentences. Our objectives in this corpus study are twofold:

- (a) Gather data to document the scoping potential of sentence initial temporal expressions, as well as validate the postulate of a difference in scoping ability linked to integration¹. A first contribution of this study will be to provide a fairly systematic examination, in an extensive corpus, of phenomena which have until now been observed separately, in theoretical mode mostly and with made up examples as concerns integration, and on the basis of limited illustrative data for the issue of scope. At this stage, we choose to concentrate on the most clearly-contrasted cases: initial (detached or non-

¹ An earlier corpus-based study (Thompson 1985), though set in a different theoretical framework, established the same kind of position-linked difference in the case of initial and final purpose clauses.

detached) and final non-detached temporal expressions. Within these, we focus on prepositional phrases (PPs), for methodological reasons which we explain in the next section.

- (b) Examine the extension of frames. The left boundary is established by the introducer. But are there linguistic clues to the final boundary of a temporal frame?

2. Corpus and method

Our initial corpus, the Atlas corpus, is composed of a single text: the book "Atlas de la France scolaire"². This choice was motivated by the fact that, as a geographical and historical account of schooling in France, this book is essentially structured in a spatial and temporal way.

2.1. Preliminaries

Our decision to focus our analysis on PPs reflects in the first instance a methodological choice: the choice to simplify the parameters to be taken into account for the automatic identification of temporal expressions. Yet it is not completely arbitrary, as we observed that the temporal expressions in our corpus were mostly PPs. For instance, a cursory survey came up with an extremely small number of temporal subordinate clauses. Our method can be said to be semi-automatic: the initial automatic detection based on lexical markers was followed by a manual selection of the expressions we viewed as relevant. For the lexical markers, we constituted two lists: one of temporal nouns, based on a study by Berthonneau (1989), and another containing temporal prepositions. We constructed our search filters³ in such a way that we could detect:

- PPs composed of any preposition followed by a date (represented by a regular expression) or a temporal noun. Our list of nouns includes 78 items, among which: *année, jour, février, siècle, moment, période, début*.
- PPs introduced by a temporal preposition, i.e. a preposition which does not require a temporal noun to form a temporal expression: *avant, après, pendant, durant, lors, dès, depuis, jusque*.

We are well aware that these patterns leave out a number of temporal expressions, for instance those composed with a non specialised preposition and an event noun (*à la rentrée des classes*, etc.), or those which are not introduced by a preposition (*l'année précédente, deux ans après*). Our familiarity with the corpus allows us to feel confident that our patterns cover the great majority of the temporal PPs that could be detected. In a study centred on discourse phenomena, we needed extensive context, measured in a discursively relevant unit: we opted for retaining extracts of three paragraphs, one on either side of the one containing the PP.

Among the PPs identified, the following were manually eliminated as non-relevant:

- non-temporal PPs: expressions such as *en même temps* in line 13 of example (1), which does not have a temporal meaning in this context;
- temporal expressions which only bear on one sentence constituent (such as a noun or an adjective), or which are embedded in subordinate or parenthetical clauses. These expressions, which have a status below the sentence, seem to be of limited interest for a study of scoping properties.

This sorting task was mostly concerned with eliminating non-relevant items, but occasionally, as in the analysis of the extended example (1) below, we have also retained expressions which, although not detected by our automatic procedure, are also relevant for our study. For example, we consider *trente ans plus tard*, (line 6), as sharing with the automatically identified surrounding PPs the discourse properties under study⁴.

² Hérin & Rouault 1994. We thank P. Enjalbert et N. Malandain (Greyc, University of Caen) for making this text (56346 words) available to us.

³ This search was executed with Yakwa, a text analysis tool developed by L. Tanguy (an illustrated description is available on www.univ-tlse2.fr/erss/membres/tanguy/Yakwa.html). We are very grateful to L. Tanguy for his readiness to help.

⁴ Such an adjustment is somewhat arbitrary. Yet, in some cases, it is perhaps less so than allowing technical considerations to strictly determine the selection of expressions.

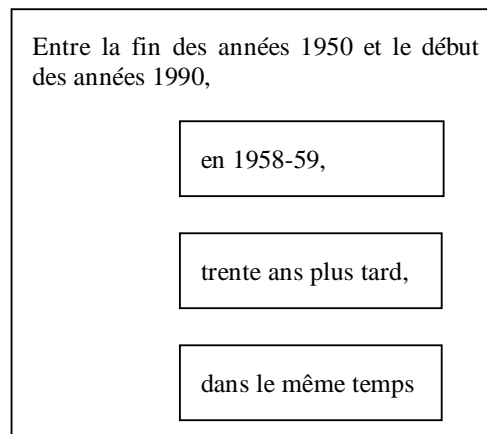
2.2. Analysis of an example

We begin by presenting an extract from our corpus, which will serve as a basis for explaining our analytical approach (temporal PPs appear in italics):

(1)

1. Si l'on définit la démocratisation de l'enseignement par l'accès d'un nombre croissant de jeunes aux formations et
2. aux diplômes de l'enseignement secondaire et de l'enseignement supérieur, il y a bien eu, en une génération,
3. démocratisation du système éducatif français. En effet, *entre la fin des années 1950 et le début des années 1990*,
4. le nombre des collégiens a triplé, celui des lycéens quadruplé et celui des étudiants sextuplé. Les taux de
5. scolarisation des adolescents et jeunes adultes ont considérablement augmenté: *en 1958-59*, à peine 20% des
6. jeunes de dix-huit ans étaient scolarisés; *trente ans plus tard*, la proportion dépasse 80% (apprentis compris) et
7. *dans le même temps* le nombre des admis au baccalauréat est passé d'à peine 60 000 à plus de 400 000.
8. Toutes les catégories sociales ont bénéficié de l'allongement des scolarités et de l'accès de plus en plus ouvert aux
9. formations des cycles secondaires et supérieurs: le nombre des étudiants français d'origine ouvrière était de
10. 30 000. *au début des années 1960*, il est de 130 000 *au début des années 1990*; celui des étudiants provenant
11. des catégories sociales aisées, patrons, professions libérales, cadres supérieurs, est passé *dans le même temps* de
12. 100 000 à 350 000 environ. Mais le rapprochement de ces derniers chiffres, s'il confirme l'ouverture des niveaux
13. supérieurs de formation aux jeunes des milieux économiquement modestes, indique en même temps les limites
14. de la démocratisation du système éducatif. Certes, la proportion d'étudiants d'origine ouvrière (salariés agricoles
15. et personnels de service compris) dans les universités a presque doublé *au cours des vingt dernières années*;
16. mais elle n'est encore que de 12%, alors que plus de 40% des jeunes d'âge scolaire sont d'origine ouvrière. Et
17. près du tiers des étudiants des universités proviennent des milieux aisés, alors que les enfants des familles de
18. cadres supérieurs et des professions libérales ne forment guère que 10% de la population scolarisable.

We will concentrate here on the analysis of two sets of tokens which lend themselves particularly well to exhibiting the functional difference between sentence initial and sentence final temporal expressions (from now on ITEs and FTEs). The first set consists of the ITEs: *entre la fin des années 1950 et le début des années 1990*, *en 1958-59*, *trente ans plus tard*, *dans le même temps*. The second one consists of the FTEs: *au début des années 1960*, *au début des années 1990*⁵. The first set exemplifies the framing characteristics described in the introduction. *Entre la fin des années 1950 et le début des années 1990* opens a frame which extends beyond the sentence where it appears: the interpretation criterion thus provided scopes over a number of sentences which constitute a segment, within which the next three ITEs (*en 1958-59*, *trente ans plus tard*, *dans le même temps*) open sub-frames⁶. This can be visually represented in the following manner:



There is no such temporal framing with the expressions in the second set. The organisation is provided by the anaphoric link between the clauses, and the FTEs are only temporal modifiers within the clause (in other words, they only provide a specification).

⁵ *dans le même temps* in the same paragraph is not dealt with in our study as it is in a median position, a position which will be examined in further work. Though it appears to function here in the same way as an FTE, we do not want to prejudge the equivalence of these two positions.

⁶ The interpretation whether an ITE opens a new frame or a sub-frame depends crucially on world knowledge: it is world knowledge that tells us that the periods *1958-59* and *trente ans plus tard* are included in *entre la fin des années 1950 et le début des années 1990*.

3. The scope of ITEs

As one of our main objectives is to investigate the scoping properties of temporal expressions, our first task is to examine the way in which ITEs structure the text. And insofar as they open segments which can extend over several sentences, the first question is to determine what may signal the final boundary of such segments. In other words, what indicates to the reader that the interpretation criterion provided by the ITE no longer applies? This preoccupation leads us to establish a distinction between ITEs opening a frame, and ITEs opening a sub-frame: the above example (line 3 to 7) shows that only the former can have an extended scope, whereas the scope of the latter is clearly limited to individual items of the enumeration.

In our corpus, we have identified a number of parameters associated with the end of a frame (shown in our examples by ♦). Two major types of clues are directly connected to time: the first is occurrence of another temporal expression referring to a time not included in the period denoted by the frame introducer; the second is change of tense. Another clue of a more general nature is change of paragraph.

Further work will be needed to determine precisely the interaction between these clues. We can however already present some observations. It might seem that the presence of an ITE introducing a new frame should be a sufficient clue to a closing frame boundary. And yet, in our corpus, an overwhelming number of these cases display a change of tense as well, as in the example below which combines the two parameters: the opening of a new frame is marked both by another ITE (*en 1989*) and by a change from a perfect to a present tense.

(2)

En juin 1992, 747 500 candidats se sont présentés à l'examen, dont 35 000 candidats individuels ; près des trois quarts ont été reçus ; mais pour les candidats individuels le taux de réussite a été à peine de 50 %. Pour la série collège (85 % de l'ensemble des candidats), 76 % des candidats des établissements scolaires ont obtenu le brevet, ceux des collèges privés sous contrat réussissant mieux que ceux des collèges publics (85 % de taux moyen de succès dans les premiers, 74 % dans les seconds). ♦

En 1989, tant les collégiens du privé que ceux du public ont de meilleurs résultats dans les départements des académies de l'Ouest où les élèves du privé sont nombreux, d'Orléans -Tours, Reims et Grenoble, ainsi que dans les Midis aquitain et méditerranéen.

We may wonder why we do not find in our corpus any example of a date change without a tense change, even when this would be possible (as in (2) where the present could easily be replaced by a perfect). A hypothesis might be that tense should reflect the temporal difference, and indeed this is what happens in many cases. However, our above example intriguingly contradicts this hypothesis: the change from perfect to present does not follow a temporal logic, as the second frame is temporally anterior. What is at work here is definitely a discursive logic: change of tense is a way to make very clear to the reader that she is moving to a new segment.

Change of tense does not inherently guarantee a frame boundary. But in cases when, in the absence of a new temporal expression, the tense changes from non-present to present, this can suffice to indicate such a boundary, as the use of the present is then generally an implicit reference to the time of utterance. The following example is a case in point, with a transition from perfect to present, made even more explicit by the temporal adjective *actuelle* in the NP *la tendance actuelle*.

(3)

Depuis le début des années 1960, la composition du corps enseignant a été diversifiée : les disciplines, catégories indiciaires et grades ont été multipliés, avec l'apparition de CAPES et de CAPET artistiques et techniques et la création de certains corps enseignants (PEGC et plus récemment, PLP1 et PLP2). ♦ Même si la *tendance actuelle* est à la simplification, les grades restent nombreux (près d'une quinzaine) de même que les statuts (titulaire, titulaire académique, titulaire remplaçant, stagiaire, non -titulaire). Le corps professoral demeure hétérogène.

Rather than change of tense, the relevant clue is therefore the transition from non-present to present⁷.

⁷ This is a somewhat simplified account: with *depuis*, things are much more complex, as the period referred to includes the present time. And in French, as opposed to English, this property may lead to a present tense being used to express a situation which extends from past to present: *Depuis 1993, j'habite à Toulouse* / *Since 1993, I have been living in Toulouse* (English requires a perfect). In example (3), the present tense in *la tendance actuelle est* is a "true" present which only refers to the time of utterance, and implies that we

Another example is given in (4), where two different past tenses (perfect and imperfect) coexist within the same frame opened by *De la fin du siècle dernier jusqu' aux années 1950*, and contrast with the present tense which is necessarily outside this frame.

(4)

De la fin du siècle dernier jusqu' aux années 1950, l'école primaire a été le pilier du système scolaire français. Elle inculquait les connaissances de base, lire, écrire et compter, qui serviraient toute la vie. Elle avait aussi pour mission de former les citoyens de la République. Elle délivrait le certificat d'études qui, pour le plus grand nombre, attestait de la réussite des études et marquait l'entrée dans le monde du travail. ♦

Les sessions du certificat d'études *n'ont plus lieu*.

The third frame boundary clue mentioned above is change of paragraph, which is often regarded as linked to topic change (cf. Longacre 1979, Bessonat 1988). In our data, it appears to be indeed a relevant clue, as in examples (2) and (4). It is however not necessary, as shown in example (3), where a change of frame occurs without a change of paragraph. Is it sufficient? We have no case when a new paragraph is still to be interpreted as part of an earlier introduced frame: so a new paragraph seems to be a sufficient clue to indicate a frame boundary. Is it sufficient by itself? In most cases, the new paragraph is associated to some of the above clues. But we also find instances where a paragraph change, in the absence of any temporal clue, implies a frame boundary. In example 5, the paragraph change is accompanied neither by a new temporal expression, nor by a tense change:

(5)

Malgré cette progression rapide et générale, les écarts restent sensibles entre les académies et les départements. *En 1990*, la proportion de bacheliers par classe d'âge n'atteint pas 40 % dans les académies de la grande couronne parisienne. Elle se situe entre 40 et 45 % dans le Nord, en Lorraine et en Alsace, dans les académies de Toulouse, Limoges, Grenoble, Rennes ainsi qu'en Corse. En Île -de -France les écarts sont écrasants entre Paris (67 % de bacheliers dans la classe d'âge) et la Seine-Saint-Denis (29 %) Dans l'ensemble, la moitié sud du pays continue à avoir de plus fortes proportions de bacheliers que la moitié nord, Bretagne et, à présent, Lorraine exceptées. Mais ces différences s'atténuent. ♦

L'allongement des scolarités, dont l'augmentation continue du nombre de bacheliers donne la mesure, se traduit par la fréquence sans cesse croissante des scolarités longues, poursuivies jusqu'au terme des études secondaires, sans préjuger des entrées de plus en plus nombreuses dans les formations supérieures.

At this point in our study, we may feel that we have satisfactory confirmation of the specific functional role of ITEs. By extending their scope over several sentences, in conditions that we have gone some way towards exploring, they play a part in discourse organisation and more specifically discourse segmentation. We have focused on ITEs, as we started from the accepted view that FTEs can only give a time specification, and not open a frame. In this perspective, it should be out of the question for FTEs to share the scoping properties of ITEs. And yet, reconsideration of the behaviour of FTEs in the corpus questions this initial dichotomy. We find examples of FTEs which appear to have extended scope:

(6)

Le nombre d'élèves par classe stagne dans les collèges *depuis plusieurs années*: la baisse sensible du nombre d'élèves s'accompagne en effet de nombreuses suppressions de postes d'enseignants, récupérés pour les lycées.

(7)

En moyenne, 600 à 700 écoles ont été fermées chaque année *depuis 1980*: quelques -unes par fusion entre l'ancienne école de filles et l'ancienne école de garçons; d'autres dans le cadre d'un regroupement pédagogique intercommunal; les plus nombreuses par baisse des effectifs au -dessous du seuil de fermeture d'une école réduite au fil des années à une seule classe [...]

(8)

Plus des trois quarts des 800 000 élèves qui étaient en 5e en 1991-92 sont passés en 4e. La proportion est en augmentation continue *depuis quatre ou cinq ans*. En revanche, les orientations vers les préparations aux CAP de l'enseignement technique (2 % des élèves) sont de moins en moins fréquentes - près de 10 % des élèves sont maintenant orientés vers les 4e technologiques.

In each of these examples, one might think that the FTE provides an interpretation criterion which applies to the next sentence(s). Section 4 below proposes a closer look at the data in order to establish whether one

are outside the frame introduced by *depuis* (which would be confirmed by the translation with a present tense in English: *the present tendency is...*).

can legitimately consider that FTEs have scoping properties, thus questioning the classical view of the discourse role linked to sentence position.

4. Do FTEs have scoping properties?

On closer examination, it appears that the sentences which might appear to be within the scope of an FTE always enter into specific relations with the sentence specified by this FTE. We approach these discourse relations within the framework of RST (Mann & Thompson 1988), without presenting the model, but only providing the elements which are necessary for our purposes. In examples (6) and (7), we identify respectively relations of Non-Volitional Cause and Elaboration: in the sentence(s) after the colon, elements are given which provide a cause for the situation described in the proposition specified by the FTE, or develop this proposition. As these relations do not in any way involve temporal considerations, it seems logical that the temporal specification continues to apply. The same is true of the Contrast relation in (8): a Contrast relation clearly extracts the elements in opposition, while all others are assumed to remain constant. Here, the contrast is established between two school orientations, and by default the time period is assumed to remain the same. We would therefore argue that what is at stake here is not a scoping property of FTEs, but a default continuation of a temporal specification. Nevertheless, we are aware that this point remains excessively based on the theoretical postulate that we precisely wanted to validate with our data. It would undoubtedly be more satisfactory to find empirical evidence to support our argument.

In order to eliminate the possibility of temporal specification by default, we must envisage relations with temporal implications. It is in such cases that the scoping property may be fully thrown into light. Sequence is such a relation: by definition, it involves progress in time. If we find a temporal expression which extends over several sentences in a Sequence relation, we can be sure that this extension corresponds to a frame scope, and not to a default continuation: it cannot be a default if the relation implies a moving forward of the temporal reference point. We re-examined our data in this perspective, looking for this kind of configuration. In accordance with our hypothesis, we found no FTE exhibiting what would constitute evidence of a scoping property. Unfortunately, we did not find positive evidence either, i.e. ITEs extending their scope over sentences in a Sequence relation. It became clear to us that this could be due to the expository nature of our corpus. So with the same method of identification of temporal PPs, we investigated a corpus made up of film summaries and film reviews⁸, more likely to contain narrative elements. Indeed, in this new corpus, the configuration we were looking for was frequent, and, in accordance with our expectations, only with ITEs. We give below two relevant examples:

(9)

Flash-back : *pendant l'été 1945*, les bombardiers américains B-29 déversent des tonnes de bombes incendiaires sur la ville de Kobe. D'innombrables victimes périssent dans un gigantesque incendie. Parmi les rares survivants, dans un paysage d'Apocalypse, Seita un jeune garçon de quatorze ans protège sa petite sœur de quatre ans, Setsuko. [...]

(10)

Au début des années 50, le poète Pablo Neruda, sous le coup d'un mandat d'arrêt au Chili parce que communiste, arrive en exil en Italie avec sa femme et s'installe dans une petite île du Sud. Mario Ruoppolo, fils de pêcheur en chômage, est recruté par le chef du bureau de poste local comme facteur auxiliaire dont Neruda sera l'unique client. Après une prise de contact assez froide, les deux hommes sympathisent [...]

In both cases, we have a sequence of events which must be interpreted as taking place in the period referred to by the ITE.

The existence of such examples, together with the absence of similar patterns with FTEs, at last provides clear evidence of what we wanted to show: only for ITEs is it appropriate to use the notion of scope.

Another dimension is opened up by some work in progress seeking to compare the potential of ITEs for extending their scope and delimiting segments of text, with what happens with certain kinds of titles. Our Cinema corpus provides us with several examples of interest to this question. They contain temporal expressions which, although their status is unclear and needs investigating, share some characteristics with titles (in particular they constitute an autonomous punctuational unit): what is worth noting is that these temporal expressions behave very much like ITEs in their ability to scope over a – potentially large –

⁸ The Cinema corpus (81397 words), extracted from two Internet sites, the television channel Canal+ and the daily newspaper Libération.

number of sentences. Thus, *Juillet 1914* in (11) works in a very similar way to *pendant l'été 1945* in (9) or *Au début des années 50* in (10).

(11)

Juillet 1914. Le château du vieux comte Pascal de Sainteville dans la France profonde entre "campagne bucolique" et "montagne magique". Pierre Mercadier, riche rentier et professeur dilettante d'histoire, Paulette, sa femme, qui "n'aime rien ni la musique ni l'amour" et leurs enfants, dont Gabriel, 12 ans, débarquent pour l'été chez leur oncle Pascal. Ils sont déçus de constater que le comte, ruiné, a dû louer une partie du château aux Pailleron : Ernest, ouvrier puis contremaître qui a épousé Blanche, la fille du patron, avant de devenir patron lui-même, Suzanne, leur fille et Yvonne, son amie pianiste. Le charme de Blanche séduit Pascal, Pierre et Gabriel, mais Pierre a la préférence. Gabriel entreprend son initiation aux choses de l'amour avec Suzanne et Yvonne sans craindre de les faire souffrir alternativement. [...]

5. Conclusion

The corpus analysis has enabled us to throw light on several aspects of temporal discourse framing. As concerns the distinction between initial and final temporal expressions, the homogeneous (expository) nature of the Atlas corpus was at once a disadvantage and an advantage. It did not provide data capable of clearly establishing the specificity of ITEs as regards scoping. But it revealed that all the contexts that could be interpreted as pointing to a similar scoping behaviour of ITEs and FTEs displayed the same kind of discourse relation. In turn, this observation led us to formulate the hypothesis that texts with a narrative component would be more likely to provide suitable evidence. And indeed this was the case. This little detour urged us to look into the relationship between discourse framing and another form of text organisation, i.e. discourse relations, as regards temporal structure. Furthermore, it clearly points to the necessity to take genre into account in any corpus-based study.

As a preliminary analysis, the work presented here simplifies the issue in dealing exclusively with the two most distinct cases. To be exhaustive, it will need to be completed by a study of median and detached final temporal expressions. Our intuition is that their scoping behaviour is similar to that of FTEs, but this needs to be supported by a corpus analysis such as this one.

Our second objective was to investigate the linguistic signalling of frame boundaries. We have started exploring three clues which seem to be commonly involved in the identification of the end of a frame. But we are very conscious that this exploration must be pursued much further, if it is to account for the complexity of the data.

We remained somewhat evasive as to the first clue: occurrence of a temporal expression referring to a new time. In our examples the frame boundary coincided with the opening of a new frame by a second ITE. Yet we found examples (which we did not mention for the sake of simplicity) where what appeared as a new segment only contained a temporal specification: that this should be enough to indicate the end of the preceding frame shows that the end of a frame does not necessarily signify the opening of a new one. This has to do with the fact that the theory of discourse framing constitutes a partial model of text organisation: as not all segments are frames, some other inter-segment organising principles must intervene, such as those (mentioned above) that call upon discourse relations. Furthermore, we came across some examples where the discourse relation alone seemed to imply the end of a frame, in particular when a comment is made on information presented within a frame⁹.

As regards the second clue (change of tense from non-present to present), the observations we made are indeed representative of the data in our corpus. However, given that in the French tense system, the perfect for instance is related to both present and non-present, the interaction between tenses is likely to lead to a much more complex situation

As we are interested in text segmentation, the third clue (change of paragraph) should be looked at in relation to other forms of visual segmentation in the light of studies such as those of Virbel (1989, 2001) and Nunberg (1990).

⁹ As in the following example, where the comment on what has happened *depuis quelques années* ("for the last few years") must clearly be interpreted outside this same temporal frame:

Depuis quelques années, en prenant pour base de référence l'annonce ministérielle de l'objectif d'amener les trois quarts des jeunes de chaque classe d'âge aux niveaux de formation correspondant à la classe Terminale de lycée (1985), les caractéristiques principales du système éducatif français sont en cours de rapide mutation. Quelques chiffres donnent la mesure de ces changements : [...]

A final comment: given our objective of finding clues to final boundaries, we acted as though these boundaries could always be unambiguously established. But we realise that there are also many cases when the boundary is left unspecified, without it hampering the reading process. There is nothing surprising about this, as underspecification is a general linguistic principle. However it would be worth quantifying this tendency with regard to temporal framing, and extending the quantitative study to different types of framing, such as the perspective framing mentioned in the introduction.

References

- Bally C 1944 *Linguistique générale et linguistique française*, 2ème édition. Berne, Francke.
- Berthonneau A-M 1989 *Composantes linguistiques de la référence temporelle - Les compléments de temps, du lexique à l'énoncé*. PhD thesis, Université de Paris VII.
- Bessonat B 1988 Le découpage en paragraphes et ses fonctions. *Pratiques* 57: 81-105.
- Charolles M 1997 *L'encadrement du discours : Univers, champs, domaines et espaces*. Nancy, Université de Nancy2.
- Firbas J 1972 On the Interplay of prosodic and Non-prosodic Means of Functional Sentence Perspective. In Fried U (ed), *The Prague School of Linguistics and Language Teaching*. London, Oxford University Press.
- Givón T 1979 *On Understanding Grammar*. New-York / San Francisco / London, Academic Press.
- Lambrecht K 1994 *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. Cambridge, CUP.
- Longacre RE 1979 The paragraph as a grammatical unit. In Givón T (ed) *Syntax and semantics* vol 12. *Discourse and syntax*. New-York / London, Academic Press, pp 115-134.
- Luc C, Virbel J to appear 2001 Le modèle d'architecture textuelle - Fondements et expérimentation. *Verbum* 23 (1).
- Mann WC, Thompson SA 1988 Rhetorical Structure Theory: toward a functional theory of text organization. *Text* 8 (3): 243-281.
- Nunberg G 1990 *The Linguistics of Punctuation*. Menlo Park, CSLI.
- Péry-Woodley M-P 2000 Cadrer ou centrer son discours ? Introduteurs de cadres et centrage. *Verbum* 22 (1): 59-78.
- Thompson S A 1985 Grammar and Written Discourse: Initial vs. Final Purpose Clauses in English. *Text* 5 (1-2): 55-84.
- Virbel J 1989 The Contribution of Linguistic Knowledge to the Interpretation of Text Structures. In J. André J, Quint V, Furuta RK (eds), *Structured Documents*. Cambridge, CUP.