



Signaling Elaboration: Combining French Gerund Clauses with Lexical Cohesion Cues

Marianne Vergez-Couret, Clémentine Adam

► To cite this version:

Marianne Vergez-Couret, Clémentine Adam. Signaling Elaboration: Combining French Gerund Clauses with Lexical Cohesion Cues. *Discours - Revue de linguistique, psycholinguistique et informatique*, Laboratoire LATTICE, 2012, 10, <http://discours.revues.org/8631>. <hal-00730360>

HAL Id: hal-00730360

<https://hal.archives-ouvertes.fr/hal-00730360>

Submitted on 9 Sep 2012

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

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

Marianne Vergez-Couret et Clémentine Adam

Signaling *Elaboration*: Combining French Gerund Clauses with Lexical Cohesion Cues

Avertissement

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

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

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

revues.org

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

Référence électronique

Marianne Vergez-Couret et Clémentine Adam, « Signaling *Elaboration*: Combining French Gerund Clauses with Lexical Cohesion Cues », *Discours* [En ligne], 10 | 2012, mis en ligne le 16 juillet 2012, consulté le 17 juillet 2012.

URL : <http://discours.revues.org/8631> ; DOI : 10.4000/discours.8631

Éditeur : Laboratoire LATTICE, UMR 8094 ENS/CNRS

<http://discours.revues.org>

<http://www.revues.org>

Document accessible en ligne sur : <http://discours.revues.org/8631>

Ce document PDF a été généré par la revue.

© Discours - Presses universitaires de Caen



Revue de linguistique, psycholinguistique et informatique

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

Signaling *Elaboration*: Combining French Gerund Clauses with Lexical Cohesion Cues

Marianne Vergez-Couret

LPL

CNRS & Université d'Aix-en-Provence – Université polytechnique de Hong Kong

Clémentine Adam

CLLE-ERSS

CNRS & Université de Toulouse (UTM)

.....
Marianne Vergez-Couret et Clémentine Adam, « Signaling *Elaboration*: Combining French Gerund Clauses with Lexical Cohesion Cues », *Discours* [En ligne], 10 | 2012, mis en ligne le 16 juillet 2012.

.....
URL : <http://discours.revues.org/8631>

.....
Titre du numéro : *Multidisciplinary Perspectives on Signalling Text Organisation*
Coordination : Shirley Carter-Thomas & Frédéric Landragin

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

 discours

 Presses
universitaires
de Caen

Signaling *Elaboration*: Combining French Gerund Clauses with Lexical Cohesion Cues

Marianne Vergez-Couret

LPL

CNRS & Université d'Aix-en-Provence – Université polytechnique de Hong Kong

Clémentine Adam

CLLE-ERSS

CNRS & Université de Toulouse (UTM)

.....

In this paper, we focus on the *Elaboration* relation and on its automatic identification in French, using the theoretical framework of Segmented Discourse Representation Theory (SDRT). One of the information sources identified by the SDRT framework to infer the *Elaboration* relation is based on the existence of a potential subsumption link between the eventualities at stake, depending on lexical semantics and world knowledge. We investigate this claim by combining a weak syntactic marker of the *Elaboration* relation, namely the gerund clause, with lexical cohesion cues. We aim at automatically identifying gerund clauses which are *Elaborations* by finding cohesive links between the host main clause and the gerund clause. This approach makes it possible to accurately detect few cases of intra-sentential *Elaborations* in our corpus, confirming the fact that lexical cohesion cues are relevant for this task.

Keywords: *Elaboration*, lexical cohesion cues, distributional neighbourhood, SDRT, discourse structure

.....

Dans cet article, nous nous focalisons sur la relation d'Élaboration en français, telle qu'elle est décrite dans le modèle théorique de la SDRT (Segmented Discourse Representation Theory), et sur son identification automatique. Selon la SDRT, une des sources d'information permettant d'inférer la relation d'Élaboration est basée sur l'existence d'un lien de subsumption entre les types des éventualités des segments à relier, indiquant que le type de la seconde éventualité est un sous-type de celui de la première dans la sémantique lexicale des éventualités ou grâce à des connaissances du monde. Nous proposons de contribuer à cette question en combinant un indice de la relation d'Élaboration, i. e. la construction syntaxique du gérondif, et des indices de cohésion lexicale. Notre objectif est d'identifier automatiquement des propositions gérondives qui sont des Élaborations en repérant des indices de cohésion lexicale entre la proposition principale et la proposition gérondive. Cette approche permet de détecter avec précision des cas d'Élaboration dans notre corpus, validant le fait que les indices de cohésion lexicale sont pertinents pour cette tâche.

Mots clés: *Élaboration, indices de cohésion lexicale, voisins distributionnels, SDRT, structure du discours*

- 1 The description and detection of discourse structure is a major topic in ongoing research (Moore & Wiemer-Hastings, 2003; Péry-Woodley & Scott, 2006). Many formal and functional approaches attempt to model discourse through relations

between segments (typically clauses) (Asher & Lascarides, 2003; Grosz & Sidner, 1986; Hobbs, 1990; Mann & Thompson, 1987; Wolf & Gibson, 2006). Anaphora resolution, the temporal order of event identification and other empirical problems require knowledge of discourse structure (Grosz & Sidner, 1986; Lascarides & Asher, 1993; Hobbs, 1990). Applied approaches (Baldrige & Lascarides, 2005; Lin, Kan & Ng, 2009; Subba & Di Eugenio, 2009) aim to handle and detect elements of this structure studied by formal and functional approaches in order to develop applications such as automatic generation (McKeown, 1985) and automatic summarization (Marcu, 2000), among other Natural Language Processing (NLP) tasks.

- 2 In this paper, we focus on the *Elaboration* relation. The *Elaboration* relation is particularly difficult to spot linguistically, since, as pointed out by Knott (1996) *inter alia*, it does not have a prototypical lexical marker. One of the information sources identified by the Segmented Discourse Representation Theory (SDRT) framework to infer the *Elaboration* relation is based on the existence of a potential subsumption link between the eventualities at stake, depending on lexical semantics and world knowledge, but the automatic identification of *Elaboration* has not yet been tested on large corpus data. We investigate this issue by combining a weak syntactic marker of the *Elaboration* relation, namely the French gerund clause, with lexical cohesion cues. This investigation is carried out in the framework of the VOILADIS project¹, which aims to demonstrate the importance of lexical cohesion cues for discourse analysis (Adam, forthcoming).
- 3 Our main goal is to contribute to the automatic identification of the *Elaboration* relation. Our secondary goal, following from the first, is to improve the description and formalization of this rarely studied relation, in order to expand studies on discourse signaling. The aim of the present study is twofold: first, we extend the study of devices used to mark this relation by showing that it can be identified with lexical cohesion cues; second, we collect examples that could be used to evaluate the adequacy of theoretical frameworks, presented below, to real-world data.
- 4 We first introduce the SDRT theoretical framework and the description of *Elaboration* in this framework (section 1). We aim at automatically identifying this relation by means of lexical cohesion cues, which leads us to present and discuss a lexical resource based on distributional similarity (section 2). As we shall show, lexical cohesion cues alone are not sufficient to infer *Elaboration*. We propose to combine these cues with a weak marker of *Elaboration*, the French gerund construction (section 3). A practical experiment of intra-sentential *Elaboration* detection is then presented, combining the gerund clause on the one hand, as a weak cue of *Elaboration*, and lexical cohesion cues on the other hand (section 4). This experiment allows us to take a closer look at the lexical cohesion cues at stake.

1. Financed by the Cluster for Higher Education and Research “Université de Toulouse” (PRES).

1. SDRT, a formal account of discourse structure

1.1. The SDRT framework

5 SDRT (Asher, 1993; Asher & Lascarides, 2003) is a formal theory which assumes that discourse has a hierarchical structure guiding its interpretation via discourse relations between segments (typically clauses). For the purpose of this paper, the features of SDRT described below are relevant (for further details, see Asher, 1993; Busquets, Vieu & Asher, 2001; Asher & Lascarides, 2003).

6 The goal of SDRT is to give representations for discourses via the construction of labeled SDRS (Segmented Discourse Representation Structures, represented here with the symbol π). These structures are constructed recursively, starting from elementary labeled constituents, generally one clause, to complex labeled constituents via discourse relations such as *Elaboration*, *Explanation*, *Result*, and so on. The construction is incremental: a new SDRS is determined from a constituent representing the given context and a new constituent by using a non-monotonic logic which describes the coherence link, via discourse relations, between the new constituent and the given context and by updating this new constituent in the whole discourse structure.

7 Discourse relations are described in two steps: first, they are associated with *triggering rules* to infer them (*glue logic* or logic of information packaging) and second, discourse relations entail *semantic effects* (logic of information content), such as spatio-temporal information about the events described in the constituents. *Triggering rules* to infer relations use both linguistic cues such as discourse markers, syntactic constructions, verb tenses, aspects and moods, argument structure, logical operator, quantifiers, information about lexical semantics, and non-linguistic information about word knowledge and pragmatic principles. SDRT is essentially a theory of the semantics-pragmatics interface.

8 Discourse structure is hierarchical. SDRT clearly distinguishes subordinating relations (such as *Elaboration*) from coordinating ones (such as *Narration*) depending on the contribution of the new constituent: to determine in the former case or to carry on what is ongoing in the given context in the latter case. *Elaboration* is a subordinating relation since the second proposition provides more detail about the eventuality described in the first one. In the following section, we provide a description of the *Elaboration* relation, as it is presented in the framework of SDRT.

1.2. *Elaboration*

9 Two constituents are linked by the *Elaboration* relation only if the second constituent introduces an eventuality (state or event) which provides more detail about the eventuality described in the first constituent. In the SDRT framework, one of the triggering rules to infer *Elaboration* is based on information about lexical semantics and world knowledge. More specifically, *Elaboration* can be non-monotonically inferred if there is a subsumption relation, encoded by the *Subtype_D predicate*, between the types of the eventualities involved. The Subtype_D predicate (*Subtype_D*) means

that the type of the second eventuality is a subtype of the first one according to the lexical semantics of the predicates or some piece of shared knowledge dependent on the given discourse (D). When *Subtype_D* holds, *Elaboration* can be inferred. This is the case for example [1]:

- [1] (a) Martha ate a lovely meal. (b) She devoured lots of salmon.
(Asher & Lascarides, 2003: 282)

10 We can non-monotonically infer that the type of the second event “devour lots of salmon” is a subtype of the first one “eat a lovely meal” thanks to lexical semantics (as developed below). *Non-monotonically* means that this inference can be cancelled if other monotonic inferences are established, as in the following example:

- [2] (a) Martha ate a lovely meal. (b) And then she devoured lots of salmon.

11 The discourse markers “And then” monotonically indicates that π_b (constituent introduced by (b)) is attached to π_a (constituent introduced by (a)) by *Narration*.

12 In the framework of SDRT, the lexicon is an important (but not exclusive) information source for inferring *Subtype_D* predicate. The lexicon includes information about the semantic type of objects that are denoted by common nouns, verbs, and so on. A subtype is related to a supertype by some notion of substitutability: the subtype inherits many supertype characteristics and has some specific differences; the subtype can be substituted by the supertype but the reverse is not necessarily true. The concept of subtype is closely related to the linguistic notion of hypernymy.

13 In [1], sentences (a) and (b) include words that are semantically linked. First, the type of the event described in e_b “devour” is a subtype of the type of the event described in e_a “eat”. Second, the word “meal” must be lexically specified to be of the type “food” and “salmon” is also of the type “food” but this lexical information is not directly coded in the type hierarchy. More lexical information is needed, for instance, that the property of the event “meal” is that it is eaten; all words of the type “food” have this property; finally, “salmon” is food derived from the animal salmon. This information at the lexical level between predicates (“eat” and “devour”) and arguments (“meal” and “salmon”) sharing the same θ -role (here patient) allows us to infer *Subtype_D* between the labeled constituents π_a and π_b at the discourse level.

14 SDRT claims that while discourse structure is sensitive to non-linguistic information such as world knowledge, it is conceptually and computationally more efficient to take into account linguistic knowledge to which we have direct access. However, non-ambiguous discourse markers are rarely available for *Elaboration*.

1.3. Signaling *Elaboration*

15 It is worth remembering that our goal is to automatically identify *Elaborations*. This is a very challenging task: Scott and De Souza (1990), Knott (1996), Knott et al. (2001) observed that *Elaboration* is a relation for which there are no obvious surface signals.

Automatic identification using prototypical discourse markers is therefore impractical. A different approach (going beyond traditional discourse markers) is required in order to automatically detect this relation.

16 Marcu (2000) illustrates these two points. His algorithms are based on discourse markers and word co-occurrences. Even if he uses the discourse marker *specifically* to identify *Elaboration*, that marker does not occur frequently and covers few cases of *Elaboration*. Thus, he also reports on a non-linguistic marker, based on the number of sentences in a paragraph or the number of paragraphs in a section: if this number is small and no discourse markers are used, the relation between the sentences or paragraphs is generally *Elaboration*. This rule is actually a second-best option. *Elaboration* appears to be considered as a catch-all relation.

17 Even if no frequent obvious surface signal exists for *Elaboration*, recent work has emphasized ambiguous cues of *Elaboration*. For example Bras (2007) shows that the French adverbial *d'abord* (*first*) requires subordination with a constituent above it in the discourse structure via *Elaboration*, *Explanation*, *Result* or *Flash-back*. Vergez-Couret (2010) shows that French focus adverbs, such as *notamment*, *particulièrement*, *précisément*, etc., and syntactic structures such as gerund clauses, play a role in signaling *Elaboration*.

18 In this paper, we would like first to put aside discourse markers in order to focus on the source of information encoded by the Subtype_D predicate, i.e. on semantic relationships between types of objects and events at stake in *Elaborations*. We set aside the theoretical point of view and adopt a descriptive approach to the data² in order to find strategies to automatically detect *Elaboration*. The main idea is first to describe the lexical relations involved between words in two segments linked by *Elaboration*. And secondly, we discuss what kind of lexical resource is needed in order to highlight these relations automatically.

2. *Elaboration* and lexical cohesion

2.1. From *Subtype* to lexical similarity

19 At first glance, it may seem that a resource providing information about hypernymy could be the appropriate resource in order to automatically detect *Elaboration*. However *Elaboration* exhibits a wider range of lexical relations.

20 The *Elaboration* relation, at the discourse level, may rely on relations at the lexical level; however, these relations are diverse and not restricted to hypernymy. Since these relations emerge in discourse, the lexical phenomena involved can be different from those found in traditional lexical resources, such as WordNet which identifies synonymy, hypernymy, meronymy, antonymy, and troponymy (Fellbaum,

2. The data we will discuss in the next section are extracted from the ANNODIS corpus, a corpus of French written texts annotated with discourse relations (Péry-Woodley et al., 2009).

1998). Such relations can be established by discourse, and may be tightly related to a specific enunciation (Mortureux, 1993). We illustrate this issue in the following attested examples of extra-sentential *Elaboration*:

- [3] (a) Un véhicule a effectué une spectaculaire sortie de route, hier vers 18 h 15, sur l'A36. (b) La voiture circulait dans le sens Mulhouse-Montbéliard (c) lorsqu'après être passée à hauteur du 35^e RI, (d) elle a quitté la chaussée sur sa droite.
 (a) *A vehicle left the road in a spectacular fashion yesterday around 6.15 on the A 36.*
 (b) *The car was travelling from Mulhouse to Montbéliard (c) when after reaching the 35th RI, (d) it left the road on the right-hand side.*
- [4] [...] (a) qui rappelle la vocation des bénévoles de l'association: (b) être un soutien pour la paroisse, (c) apporter une petite contribution financière aux travaux grâce aux manifestations et aux dons, (d) accomplir de multiples tâches et démarches touchant aux bâtiments paroissiaux, (e) contribuer à la convivialité entre les paroissiens.
 [...] (a) *which brings to mind the role of the Association's volunteers: (b) to be a support to the parish, (c) to financially contribute a small amount to projects through activities and donations, (d) to complete many tasks and procedures dealing with the parish buildings (e) and to foster friendly relations between parishioners.*

21 In example [3], the type of event in segment (d) “quitter la chaussée” (*leave the road*) is a subtype of the type of event in segment (a) “effectuer une spectaculaire sortie de route” (*leave the road*). *Elaboration* (π_a, π_d) holds. Three lexical links play a role in inferring *Elaboration*: “véhicule” (*vehicle*)/“voiture” (*car*), “sortie” (*exit*)/“quitter” (*leave*) and “route” (*road*)/“chaussée” (*roadway*). The first link, “véhicule” (*vehicle*)/“voiture” (*car*), is clearly classified as hypernymy and can be brought closer to the lexical subtype. However the “route” (*road*)/“chaussée” (*roadway*) link is in fact meronymy, and the “sortie” (*exit*)/“quitter” (*leave*) link, which plays a crucial role here, is more subtle to categorize, since cross-category relations are generally not listed in typologies. These cross-category relations are nevertheless very frequent in *Elaborations*.

22 In example [4], the types of states in segments (b) to (e) are subtypes of the type of state “vocation des bénévoles de l'association” (*role of the Association's volunteers*). *Elaboration* ($\pi_a, [\pi_b, -\pi_c]$) holds. At the word level, this inference relies on links between “vocation” (*vocation*) and words such as “soutien” (*support*), “accomplir” (*to complete*), “tâche” (*task*) or “contribuer” (*to foster*). These links are established in discourse, and will most probably not appear in a generic resource, since they do not match traditional lexical relations. Such links are more accurately referred to as lexical similarity relations.

23 First of all, these observations lead us to the conclusion that lexical relations involved in *Elaborations* are more diverse than hypernymy. For example, in [3], the relation of discursive subtype between “effectuer une spectaculaire sortie de route” (*leave the road in a spectacular fashion*) and “quitter la chaussée sur la droite”

(*leave the road on the right-hand side*) does not rely on relations of lexical subtype (in contrast with what we have seen in example [1] between “devour lots of salmon” and “eat a lovely meal”). Lexical subtype can be considered between “effectuer une sortie spectaculaire” and “quitter”, but in this case, it is necessary to take into account a level higher than the word level. Unfortunately, such higher levels are difficult to consider in an automatic detection application. Staying at the word level, it becomes beneficial to consider loose associative relations between words, thus making it possible to take into account links such as meronymy on the one hand (“route” and “chaussée”) and cross-category relations on the other hand (“quitter” and “sortie”).

24 Secondly, we observe that lexical relations involve different syntactic positions, but mostly link words that appear as subjects, verbs or head nouns of objects.

25 Thus, the lexical relations involved in the interpretation of *Elaboration* are relatively loosely constrained: they are not limited to hypernymy and may appear at different syntactic positions. However, we note that not all lexical cohesion links are necessarily involved in the identification of *Elaboration*. Lexical cohesion covers the whole text as it participates in its general coherence.

26 With these points in mind, we discuss in the next section the requirements for a resource to be used in our task of identification of *Elaboration*.

2.2. Selecting the appropriate lexical resource: distributional neighbours

27 We have seen that recognition of the *Elaboration* relation seems indeed to involve lexical cohesion cues. In order to automatically detect this relation, the chosen resource is crucial: it should comprise these links and allow for their automated usage. As stressed in the previous section, a generic resource seems poorly fitted to this task. We have focused on a resource built from corpora, taking into account semantic proximity links, possibly across parts of speech. In particular, we have chosen the *Voisins de Wikipédia* database, a resource built by distributional analysis. The principle of distributional analysis is to pair words based on their shared contexts, following Harris’ (1968) hypothesis. The paired words share second-order affinities: they do not need to appear together in the corpus, but their environments are similar (Grefenstette, 1994). The lexical relations evidenced are then paradigmatic.

28 The *Voisins de Wikipédia* database was built from a full archive of the online encyclopaedia Wikipedia, which contains more than 194 million words. The archive was processed through the Syntex-Upéry chain developed by Bourigault (2002). First, a syntax analysis is performed. Then, all <governor, relation, dependent> triplets are listed; an example triplet is:

<circuler, à bord de, voiture> (<travel, in, car>)

29 The triplets are then transformed into <predicate, argument> couples, where the predicate is a combination of two components, i.e. the governor and the relation:

<circuler_à bord de, voiture> (<travel_in, car>)

30 The similarity between distributions is computed for each predicate couple and each argument couple using Lin’s score: predicates are paired based on their shared arguments; reciprocally, the same pairing is performed on arguments, based on their shared predicates. Thus, arguments “véhicule” (*vehicle*) and “voiture” (*car*) are paired through shared predicates such as “circuler_à bord de” (*travel_in*), “capot_de” (*hood_of*), “conduire_obj” (*to drive_obj*), etc. Predicates “sortie_de” (*exit_from*) and “quitter” (*leave*) are paired through shared arguments such as “territoire” (*territory*), “hôpital” (*hospital*), “autoroute” (*motorway*), etc.

Neighbours	Shared context
véhicule/voiture	circuler_à bord de pare-brise_de portière_de percuter_suj immatriculation_de ...
sortie_de/quitter	territoire système éducatif conservatoire hôpital autoroute ...

Table 1. Instances of distributional neighbours with their shared context

31 The resource thus obtained contains 4 million pairs, covering a large panel of relations. An example of distributional neighbourhood links projected on the text sample [3] is provided below. Only links between the two sentences appear.

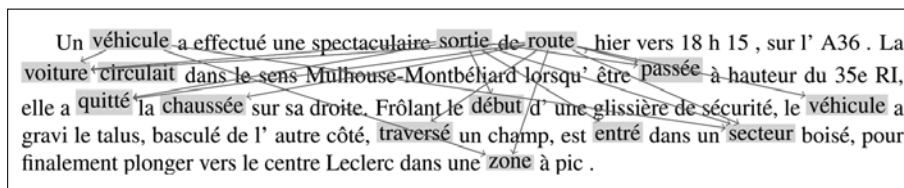


Figure 1. Neighbourhood links projected on [3]

32 Here, the aforementioned links relevant for identifying the *Elaboration* relation are observed: “véhicule” (*vehicle*)/“voiture” (*car*), “sortie” (*exit*)/“quitter” (*to leave*), and “route” (*road*)/“chaussée” (*roadway*). Other links participating in the global lexical cohesion of the text are also observed, but these links are not involved in *Elaboration*, for instance “route” (*road*) and “véhicule” (*vehicle*). Naturally, cohesive links appear between two consecutive segments linked by a relation. Finally, some links are not relevant at all in this context, for example “route” (*road*) and “traverser” (*to cross*).

33 Even though lexical relations relevant to our task are revealed by projecting the *Voisins de Wikipédia* in the text, many other irrelevant lexical links will interfere, making a direct inference to the discourse level impossible. It is therefore necessary to define a more restrictive marker, taking into account more elaborate criteria than the simple presence of neighbourhood links.

34 In order to address this issue, two strategies were considered:

- the first strategy restricts neighbourhood links to specific syntactic positions;
- the second strategy combines neighbourhood links to other cues for *Elaboration*.

35 We chose to perform an experiment that merges the two strategies, with a detection based on targeted neighbourhood links in specific syntactic positions combined with the presence of another cue of *Elaboration*, the gerund clause. Our hypothesis is that the combination of these two cues will build a stronger cue. As we shall see in the next section, the gerund construction is a cue of *Elaboration*. Gerund constructions are particularly suitable for automatic identification since they are easily picked up in text and since discourse segments are determined by syntax (the segment to attach is the gerund clause and the target segment is the main clause). Thus, we can focus on the contribution of lexical cohesion cues.

3. Gerund clauses, *Elaboration* and lexical cohesion cues

36 Gerund clauses have an adverbial or circumstantial value. The gerund clause establishes a syntactic subordination to its host main clause: two processes are linked in this way and different semantic values can be expressed (Halmøy, 1982). This section first lists these different semantic values potentially expressed by gerund clauses and then explains how distributional neighbourhood links may highlight them.

3.1. Semantic relations in gerund clauses

37 Halmøy (1982) proposes a typology of French gerund clauses based on semantic relationships between the events described by the main clause and the gerund clause. In this section, we detail this classification, adopting the perspective of discourse relations. Indeed, we have seen that inferring discourse relations may rely on the presence of semantic relationships between the types of events involved (given *Subtype_D*). As elaborating constituents normally come after the elaborated ones, the present analysis is restricted to gerund clauses that appear after the main clause.

- Temporal relationship

38 Halmøy classifies gerund clauses for which only a temporal interpretation is possible as illustrated in the following example:

[5] Une femme a toujours un paquet de linge à déposer en partant au cinéma, le pain à ne pas oublier en rentrant du travail.

(a) *A woman always has a bundle of laundry to drop off when she goes to the cinema, always has to remember to get the bread when she comes back from work.*

39 The role of the gerund clause is to specify a time for the event described in the main clause. In that case, *Elaboration* will not be inferred.

– Relation of dependence

40 Halmøy highlights three different kinds of dependence relations between the gerund clause and the main clause: <consequence, condition> illustrated in [6], <effect, cause> illustrated in [7], and <result, means> illustrated in [8] and [9]. The starting point of the event described in the gerund clause is logically prior to the event denoted by the main clause.

[6] (a) Il se porterait mieux (b) en mangeant moins.

(a) *He would feel better (b) if he ate less.*

41 The gerund clause indicates a hypothesis “if he ate less” and the main clause the consequence, which will become true if the hypothesis is true “he would feel better”.

[7] (a) Marion a pleuré (b) en m’entendant crier.

(a) *Marion cried (b) when she heard me scream.*

42 A causal interpretation is, here, superimposed on the temporal interpretation.

[8] (a) Paul a éteint le feu (b) en pissant dessus.

(a) *Paul stopped the fire (b) by peeing on it.*

43 Gerund clauses expressing means are often considered the prototypical case of the construction. These gerund clauses answer the question *How?* In [8], the gerund clause expresses the means used “pisser” (*to pee*) in order to achieve a precise goal “éteindre le feu” (*to stop the fire*). Thus, there is a causal relationship between the two events. In these cases, *Elaboration* will not be inferred.

44 However, the gerund clause may also express the means used in accomplishing an event as illustrated in [9]. There is a kind of inclusion relation: the event “essayer d’accomplir quelque chose” (*to try to do something*) includes the event “frotter” (*to rub*). This interpretation is facilitated by certain kinds of verb in the main clause such as an effort verb like “essayer” (*to try*) in [9].

[9] (a) Il essayait de se réchauffer les pieds (b) en les frottant contre ses mains.

(a) *He was trying to warm his feet (b) by rubbing them with his hands.*

45 We keep this issue for further work and only keep in mind that a tenuous distinction between cases ([8] and [9]) exists. In our opinion, in cases such as [9], an *Elaboration* relation is inferred since (b) explicitly describes a manner specification of

the attempt to warm his feet. Example [9] is close to Halmøy's examples involving a relation of inclusion.

– Relation of inclusion

46 According to Halmøy, the relation of inclusion holds between an event described in the main clause which is abstract or metaphoric and an event described by the gerund clause which gives a concrete realization specifying the main clause's event, as illustrated in [10] and [11].

[10] (a) Ségolène Royal avait volé dans les plumes de ses collègues (b) en annonçant une réforme de l'accouchement sous X.

(a) *Ségolène Royal discomfited her colleagues (b) by announcing a reform to anonymous birth giving.*

[11] (a) Aucun hôpital n'accepterait de violer la loi, (b) en mettant au monde l'enfant d'un couple non marié.

(a) *No hospital would agree to break the law, (b) by helping give birth to the child of an unmarried couple.*

47 "Voler dans les plumes de ses collègues" (*to discomfit her colleagues*) and "violier la loi" (*to break the law*) are instances of the so-called *criterion predicates* (Kearns, 2003; Sæbø, 2006). Criterion predicates require conventional criteria and are unspecific: "The key notion here is that there is some conventional criterion an action must meet in order to qualify as being an event of the criterion-matching kind" (Kearns, 2003: 599). There always is an action which matches or satisfies the criterion: for example in [16], in order to "voler dans les plumes de ses collègues" (*to discomfit her colleagues*), Royal must do something; in this context, "annoncer une réforme de l'accouchement sous X" (*to announce a reform to anonymous birth giving*) matches or satisfies the criterion. Criterion-matching is also characteristic of "violier la loi" (*to break the law*). Given that the law forbids to "mettre au monde l'enfant d'un couple non marié" (*helping give birth to the child of an unmarried couple*), that constitutes *breaking the law*. Usually with criterion predicates, one needs more information of how the action is executed to evaluate its truth. If one says that he broke the law, one will want to know what he actually did. One way of specifying more concrete criteria is to modify the main clause with a gerund clause as in [10] and [11].

48 To describe these cases, Halmøy uses the term *hypernym*. However, it appears odd to consider hypernymy here. The type of event described in the gerund clause is not a hyponym of the one described by the main clause but rather a specific way of performing an unspecified action (described by the criterion predicate). In these cases, the interpretation of the *Elaboration* relation does not rely on the inference of the *Subtype_D* predicate but relies on the ability to consider the event described by the gerund clause as a reasonable criterion for satisfying the kind of event described by the host. This match is determined on the basis of world knowledge and on discourse context.

– Relation of hyponymy

49 Halmøy describes the cases below as hyponymy.

[12] (a) Elle répondit (b) en bafouillant...
(a) *She answered (b) babbling...* [*She stammered an answer...*]

[13] (a) Ludi vient à leur rencontre (b) en galopant.
(a) *Ludi comes their way (b) galloping.* [*Ludi galloped towards them.*]

50 Verbs in the main clause generally involve communication: “répondre” (*to answer*) or movements: “venir” (*to come*) and verbs in the gerund clauses describe the manner of speaking or moving. One could say that the relation here relies more on a sort of *double hyponymy*. This point is illustrated by examples [12] and [13]: in [12], “bafouiller” (*to babble*) is not a hyponym of “répondre” but “bafouiller” and “répondre” are both different hyponyms of “dire” (*to say*); in [13], “galoper” (*to gallop*) is not a hyponym of “venir à leur rencontre” (*to come*) but “galoper” and “venir à leur rencontre” are both different hyponyms of “se déplacer” (*to move*). Thus, “bafouiller” can be considered a hyponym of “répondre” and “galoper” as a hyponym of “venir à leur rencontre”. Finally, in the SDRT framework, the *Subtype_D* predicate holds in these cases and *Elaboration* is inferred.

– Relation of simultaneity

51 The actions described in the main clause and the gerund clause happen at the same time and are not linked by any logical relation.

[14] (a) Le chef faisait les cents pas (b) en fumant sa pipe en bambou.
(a) *The boss was pacing up and down, (b) smoking his bamboo pipe.*

[15] (a) À ses côtés, Jacques Chirac envoyait des baisers à la foule (b) tout en aplatissant une mèche folle, dérangée par la brise.
(a) *Standing at his side, Jacques Chirac blew kisses to the crowd (b) while flattening a wisp of hair disturbed by the breeze.*

52 These cases are puzzling in the SDRT model since none of the discourse relations in Asher and Lascarides’ (2003) study seem appropriate to analyze them. Behrens and Fabricius-Hansen (2010) propose to introduce a new discourse relation, *Accompanying circumstance*. Vergez-Couret (2010) proposed solutions and justifications for introducing this relation in the SDRT framework. But, for the purpose of our experiment, we only keep in mind that the *Elaboration* relation does not hold since it could not be inferred between two types of event in a fortuitous relation.

53 Gerund clauses link two processes. The different semantic values expressed are not conveyed by the gerund itself but depend on the combination of the two linked verbs. The interpretation is done a posteriori and determined by the semantic relationship between the verbs and other elements given by context, as Halmøy (1982) pointed out.

54 In some cases, as illustrated above, the semantic value expressed by the gerund (i.e. some cases of relation of dependence <result, means>, relation of inclusion and relation of hyponymy) leads to inferring the *Elaboration* relation. This inference constitutes our baseline for the annotation of gerund clauses which are *Elaborations* (see section 4).

55 More importantly, we note that the semantic value carried by the gerund clause, leading to the inference of *Elaboration*, is subtle and larger than the semantics of *Subtype_D*. Our hypothesis relies on the usefulness and adequacy of lexical similarity relations in which lexical items share similar environments (i.e. the neighbours) in order to reach our goal, i.e. the automatic identification of gerund clauses that are *Elaborations*.

3.2. Combining the gerund clause with distributional neighbours

56 We propose to combine the gerund clause, an ambiguous cue of *Elaboration*, with lexical cohesion cues. More specifically, we propose to detect gerund clauses that carry *Elaboration*, by assuming that such clauses involve links of lexical cohesion. Such clauses should involve lexical cohesion links. Such a combination should be more reliable than the separate cues.

57 This is illustrated in examples [16] and [17] where an *Elaboration* relation is inferred:

[16] (a) Puis on irrigua les alentours (b) en creusant un canal dérivé du Zab Supérieur.
(a) *Then, they irrigated the surrounding areas (b) by digging a canal flowing from the River Zab Supérieur.*

58 [16] is a case of <result, means> dependence. The *digging* is the means used to achieve the goal of *irrigating*. We infer *Elaboration* between (a) and (b) since the activity denoted by the type of event “irriguer” (*to irrigate*) may involve the activity denoted by the event “creuser” (*to dig*). It also seems to us that “irriguer” and “creuser” could be found as distributional neighbours since they can share similar environments such as: “irriguer les terres” (*to irrigate the soil*); “creuser la terre” (*to dig the soil*). Example [17] is also of type <result, means> dependence.

[17] (a) United Fruit Company investit dans le pays, (b) en achetant des parts dans le chemin de fer, l’électricité et le télégraphe.
(a) *United Fruit Company invests in the country, (b) buying shares in railway, electricity and telegraph.*

59 The main clause introduces the event “investir” (*to invest*) and the gerund clause introduces the event “acheter des parts” (*to buy shares*). An SDRT analysis would exploit lexical and world knowledge to infer *Subtype_D* and then *Elaboration*. One would also expect “investir” and “acheter” to be found as neighbours, since they may share similar environments: “investir dans un ordinateur et des équipements de bureau/l’immobilier/une nouvelle usine” (*to invest in computer and office equipment/*

real estate/a new company); “acheter un ordinateur et des équipements de bureau/dans l’immobilier/une nouvelle usine” (*to buy computer and office equipment/real estate/a new company*).

60 Conversely, gerund clauses which are not *Elaborations* should not involve lexical similarity relations; see examples [18] and [19].

[18] (a) Kokopelli ouvrit le piège (b) en hurlant.
(a) *Kokopelli opened the trap, (b) screaming.*

[19] (a) Environ 200 colons français se dirigèrent vers le nord (b) en transportant leurs biens dans des chariots.
(a) *Approximately 200 French colonists moved north, (b) carrying all their belongings on carts.*

61 [18] and [19] are examples involving a relationship of simultaneity. In [18], one could expect a causal relation between the *opening* and the *screaming* but the gerund construction mostly conveys that the two events happen at the same time. As the events denoted by “ouvrir” (*to open*) and “crier” (*to scream*) are in a fortuitous relation, no lexical similarity is expected between them.

62 In [19], the events described by “se diriger” (*to move*) and “transporter” (*to carry*) are also in a fortuitous relation. We assume that the two verbs are not linked by any relation of lexical similarity.

63 Our main idea is that verb phrases in the main clause and the gerund clause will generally contain words that are neighbours in *Elaboration* cases but not in other cases. For example, it seems to us that “irriguer” and “creuser” in [16] and “investir” and “acheter” in [17] could be found as neighbours but not “ouvrir” and “hurler” in [18] and “se diriger” and “transporter” in [19].

64 With this hypothesis in mind, we set up the experimentation presented in the next section.

4. Experimental validation

4.1. Motivations and strategy

65 The goal of the presented experiment is to reliably identify gerund clauses which are *Elaborations*: we aim for the highest precision. This task is challenging: it is sparsely attempted in the literature and the reliability attained is low. Nevertheless, such attempts are required for a better understanding of the *Elaboration* relation.

66 While this task is interesting in itself, our experiment will also illustrate the improvement provided by taking into account lexical cohesion phenomena for discourse analysis, and show the relevance of lexical neighbourhood for detecting

these phenomena. If using distributional neighbours brings a significant performance improvement, this will indicate that *Elaboration* is a lexically marked relation.

67 In order to reach these goals, we chose to use the gerund clause as a weak cue of *Elaboration* (hereafter noted **G**) in combination with the distributional neighbours. Two combination strategies are tested.

68 The first strategy consists in extracting gerund constructions in which verbs in the gerund clause and the main clause are neighbours, as in [20]:

[20] [...] (a) et les villages **contribuaient** également à ce grand projet religieux (b) **en envoyant** des vivres.
 [...] (a) and villages also contributed to this great religious project (b) by sending supplies.

69 The semantic proximity between “contribuer” (*to contribute*) and “envoyer” (*to send*) captured by the neighbourhood link plays a role in signaling an *Elaboration* relation between the gerund clause (b) and the main clause (a). This pattern is noted **G + NV** (Gerund construction with Neighbour Verbs).

70 The second strategy consists in adding a new constraint. Along with the presence of verbs connected by a neighbourhood link, this strategy imposes that a neighbourhood link also exist between the head noun of the verbs’ objects, as in [21]:

[21] (a) Les Skrulls [...] **élargissent** leur empire (b) **en englobant** dans celui-ci les mondes moins avancés qu’ils rencontrent.
 (a) The Skrulls [...] expand their empire (b) by incorporating the less evolved worlds they discover.

71 Here, two neighbourhood links support the inference of *Elaboration*: the one linking the verbs “élargir” (*to expand*) and “englober” (*to incorporate*) and the one linking “empire” (*empire*) and “monde” (*world*). This pattern is noted **G + NV + NO** (Gerund construction with Neighbour Verbs and Neighbour head noun of the Objects).

72 The three patterns, **G**, **G + NV** and **G + NV + NO** are increasingly restrictive. The **G + NV + NO** cases are included in **G + NV** cases which are included in **G** cases. If our hypothesis is true, each of these patterns should extract *Elaborations* with increasing precision. In the next section, we present the extraction of these three patterns in greater detail.

4.2. Extraction of *Elaboration* candidates

73 In this experiment, the corpus used is a fraction of the French Wikipedia: 45,823,899 words from 5,106,831 sentences, which amounts to roughly one fifth of the online encyclopaedia. This corpus was pre-processed with Syntex for the syntactic analysis. All sentences featuring a [verb clause, gerund clause] pair were extracted (**G**). Two subsets of these candidates were then formed, by taking into account lexical neighbourhood constraints on the verb pairs (**G + NV**) and on the verb object pairs

(**G+NV+NO**), as explained in the previous section. Table 2 gives the number of candidates obtained depending on which markers were used; remember that the set **G+NV+NO** is included in the set **G+NV**, which is included in **G**.

G	G+NV	G+NV+NO
18571	375	193

Table 2. Number of candidates

74 The number of *Elaboration* candidates (**G+NV** and **G+NV+NO**) is small considering the total number of gerund clauses, especially since we are only considering inter-sentential realizations of *Elaboration*. Nevertheless, in the current state of research on this barely studied relation (notably because of the lack of obvious markers), defining a reliable marker is a significant improvement, even if the number of matches is small.

4.3. Annotation of extracted candidates

75 Each text was independently annotated by two experts in discourse relations³ with the specifications defined in section 3.1. 314 examples were annotated, approximately 100 for each case (**G**, **G+NV**, **G+NV+NO**). The examples were presented in a random order to the annotators with the question: *Is the gerund clause an elaborating segment of the main clause?*

76 The inter-annotator agreement rate was 89% (280 cases of agreement vs. 34 cases of disagreement). The kappa score (Cohen, 1960) is 0.70, which indicates a moderate to good inter-annotator agreement. This reveals the difficulty of the task. The kappa score is, however, good enough to consider automation of this task.

77 In a second run, the 34 examples on which the experts disagreed were explored, in order to ensure that the reference annotation was as reliable as possible, and also to analyze the types of inter-annotator variation. The discussion allowed us to refine the annotation for the vast majority of disagreement cases. Finally, only 9 cases resulted in the experts disagreeing; such cases include texts for which two interpretations are possible. To ensure meaningful results, these 9 marginal cases were discarded from the reference which was subsequently used to evaluate the results of automated *Elaboration* detection.

4.4. Results and perspectives

78 The table below summarizes the results obtained when testing the three strategies for *Elaboration* detection. For each pattern the following information is given:

3. The authors of this paper.

- number of candidates extracted from the corpus,
- number of candidates annotated in the reference,
- number of candidates annotated with *Elaboration*,
- number of candidates NOT annotated with *Elaboration*,
- precision of the pattern as marker of *Elaboration*,
- confidence interval.

	Extracted	Annotated	Elab.	Not Elab.	Precision	Confidence interval
G	18571	102	62	40	60.8%	9.45%
G + NV	375	100	81	19	81.0%	6.59%
G + NV + NO	193	104	99	5	95.2%	2.8%

Table 3. Results

79 These results confirm that the gerund clause is indeed an ambiguous cue, since only 60,8% of the candidates are *Elaborations*.

80 We chose to annotate the same number of candidates for each pattern despite the fact that each pattern differs considerably in size. Thus, we annotated half of the G + NV + NO cases but a smaller proportion of G cases. The result is a very wide confidence interval for the latter pattern. However, the performance difference between G and our two strategies (G, G + NV, G + NV + NO) is large enough to confirm that they both bring a significant improvement. With G + NV, 81% of the cases are *Elaborations*. The G + NV + NO strategy is very reliable, with 95% precision and a confidence interval below 3%. These results are highly promising.

81 The cases where our patterns failed were analyzed. In a few cases, the failure is caused by an irrelevant neighbourhood link. For example, in the context of example [22], the link between “marcher” (*to march*) et “incendier” (*to burn*) is irrelevant. In various cases, a different marker can be observed, which could be used to cancel the *Elaboration* inference. This is illustrated in example [23], where the strong lexical marker of *Contrast* relation “mais” (*but*) appears.

[22] (a) Ils marchent la campagne, (b) en incendiant toutes les habitations.

(a) *They marched the countryside, (b) burning down every dwelling they found.*

[23] (a) Le roi d'Espagne lui accorda une décoration qu'il accepta, (b) mais en refusant la pension qui y était attachée.

(a) *The king of Spain accorded him a decoration that he accepted, (b) while refusing the pension that was attached to it.*

82 These considerations suggest that our good results can still be improved upon,
 both by taking into account other types of markers (signaling other discourse
 relations) and by a finer grained filtering of the neighbours.

5. Conclusion and outlook

83 We have presented a practical experiment dedicated to the detection of *Elaboration*.
 While *Elaboration* is often considered as a relation without prototypical lexical
 discourse markers, our goal was to find signaling devices for its identification. We
 combined an ambiguous cue, the gerund clause, with lexical information provided
 by a distributional resource.

84 The results of our experiment are very encouraging. The fact that *Elaboration*
 relies on lexical device, suggested in the SDRT framework, has been validated on
 the basis of a corpus. With this contribution, we also fulfil one of the objectives of
 the ANNODIS project⁴, which aims at constructing an annotated corpus for the
 study of discourse organization in order to improve the description and formalization
 of discourse relations with real-world data.

85 The prevalence of lexical cohesion cues for discourse structuration is commonly
 accepted, but they are still neglected in NLP applications because of the difficulty
 associated with picking out lexical links in texts. This contribution confirms lexical
 neighbours as a relevant resource to be used to detect *Elaboration*. Within the scope
 of the VOILADIS project, we hope to generalize the use of lexical cohesion cues
 for all aspects of discourse analysis.

86 Only a few instances of *Elaboration* are detected by our practical experiment,
 given the size of the corpus. But as it validates our approach based on lexical
 cohesion detection, it suggests that other implementations of this approach could
 lead to a broader identification of *Elaboration*. For instance, improvements could
 be made by detecting *Elaborations* between sentences. We plan to combine lexical
 cohesion with other weak cues of *Elaboration* such as the adverbial expressions
 “d’abord” (*first*), “dans un premier temps” (*at first*). We also plan to investigate
 the role of distributional neighbours by taking into account a cohesion score
 depending on the number of neighbourhood links between two sentences and
 their syntactic positions.

References

- ADAM, C. forthcoming. *Voisinage lexical pour l’analyse du discours*. Unpublished PhD thesis.
 Université de Toulouse.
- ASHER, N. 1993. *Reference to Abstract Objects in Discourse*. Dordrecht – Boston: Kluwer
 Academic Publishers.

4. Financed by the French National Research Agency (ANR).

- ASHER, N. & LASCARIDES, A. 2003. *Logics of Conversation*. Cambridge: Cambridge University Press.
- BALDRIDGE, J. & LASCARIDES, A. 2005. Probabilistic Head-Driven Parsing for Discourse Structure. In *Proceedings of the 9th Conference on Computational Natural Language Learning (CoNLL)*. Michigan: University of Michigan: 96-103. Available online: http://dl.acm.org/ft_gateway.cfm?id=1706560&type=pdf&CFID=89822543&CFTOKEN=23024412.
- BEHRENS, B. & FABRICIUS-HANSEN, C. 2010. The Relation Accompanying Circumstance across Languages. Conflict between Linguistic Expression and Discourse Subordination? In D. SHU & K. TURNER (eds), *Contrasting Meaning in Languages of the East and West*. Berlin – Bern – New York: Peter Lang: 531-551.
- BOURIGAULT, D. 2002. UPERY: un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus. In *Actes de la 9^e conférence sur le Traitement automatique des langues naturelles (TALN 2002 – Nancy, 24-27 juin 2002)*. 75-84. Available online: http://www.atala.org/doc/actes_taln/AC_0053.pdf.
- BRAS, M. 2007. French Adverb *d'abord* and Discourse Structure. In M. AURNAGUE, K. KORTA & J.M. LARRAZABAL (eds), *Language, Representation and Reasoning. Memorial Volume to Isabel Gómez-Txurruka*. Bilbao: Universidad del País Vasco, Servicio Editorial: 77-102.
- BUSQUETS, J., VIEU, L. & ASHER, N. 2001. La SDRT: une approche de la cohérence du discours dans la tradition de la sémantique dynamique. *Verbum* 23 (1): 73-101.
- COHEN, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20 (1): 37-46.
- FELLBAUM, C. 1998. *WordNet: An Electronic Lexical Database*. Cambridge (Mass.): MIT Press.
- GREFENSTETTE, G. 1994. Corpus-Derived First, Second and Third-Order Word Affinities. In *Proceedings of Euralex*. Amsterdam: University of Amsterdam: 279-290. Available online: http://www.euralex.org/elx_proceedings/Euralex1994/31_Euralex_Gregory%20Grefenstette%20-%20Corpus-Derived%20First_Second%20and%20Third-Order%20Word%20Affinities.pdf.
- GROSZ, B.J. & SIDNER, C.L. 1986. Attention, Intentions and the Structure of Discourse. *Computational Linguistics* 12 (3): 175-204.
- HALMØY, J.-O. 1982. *Le gérondif. Éléments pour une description syntaxique et sémantique*. PhD thesis. University of Trondheim.
- HARRIS, Z. 1968. *Mathematical Structures of Language*. New York: Interscience Publishers.
- HOBBS, J.R. 1990. *Literature and Cognition*. Stanford: Center for the Study of Language and Information. Chapter V: The Coherence and Structure of Discourse: 83-114.
- KEARNS, K. 2003. Durative Achievements and Individual-Level Predicates on Events. *Linguistics and Philosophy* 26 (5): 595-635.
- KNOTT, A. 1996. *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. PhD thesis. University of Edinburgh, Department of Artificial Intelligence.
- KNOTT, A. et al. 2001. Beyond Elaboration: The Interaction of Relations and Focus in Coherent Text. In T. SANDERS, J. SCHILPEROORD & W. SPOREN (eds), *Text Representation: Linguistic and Psycholinguistic Aspects*. Amsterdam – Philadelphia: John Benjamins: 181-196.

- LASCARIDES, A. & ASHER, N. 1993. Temporal Interpretation, Discourse, Relations and Commonsense Entailment. *Linguistics and Philosophy* 6 (5): 437-493.
- LIN, Z., KAN, M.-Y. & NG, H.T. 2009. Recognizing Implicit Discourse Relations in the Penn Discourse Treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics: 343-351. Available online: http://dl.acm.org/ft_gateway.cfm?id=1699555&type=pdf&CFID=89822543&CFTOKEN=23024412.
- MANN, W.C. & THOMPSON, S.A. 1987. *Rhetorical Structure Theory: A Theory of Text Organization*. Technical report ISI/RS-87-190. Los Angeles: University of Southern California, Information Sciences Institute. 92p. Available online: http://www.sfu.ca/rst/pdfs/Mann_Thompson_1987.pdf.
- MARCU, D. 2000. The Rhetorical Parsing of Unrestricted Texts: A Surface-Based Approach. *Computational Linguistics* 26 (3): 395-448.
- MCKEOWN, K.R. 1985. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge: Cambridge University Press.
- MOORE, J. & WIEMER-HASTINGS, P. 2003. Discourse in Computational Linguistics and Artificial Intelligence. In A.C. GRAESSER, M.A. GERNSBACHER & S.R. GOLDMAN (eds), *Handbook of Discourse Processes*. Mahwah (N.J.): Lawrence Erlbaum: 439-486.
- MORTUREUX, M.-F. 1993. Paradigmes désignationnels. *Semen* 8: 123-141.
- PÉRY-WOODLEY, M.-P. et al. 2009. ANNODIS: une approche outillée de l'annotation de structures discursives. In *Actes de la 16^e conférence sur le Traitement automatique des langues naturelles (TALN 2009 – Senlis, 24-26 juin 2009)*. Available online: http://www-lipn.univ-paris13.fr/taln09/paper/paper_TALN_52.html.
- PÉRY-WOODLEY, M.-P. & SCOTT, D. 2006. Computational Approaches to Discourse and Document Processing. *TAL* 47 (2): 7-19.
- SÆBØ, K.-J. 2006. The Structure of Criterion Predicates. In J. DÖLLING, T. HEYDE-ZYBATOW & M. SCHÄFER (eds), *Event Structures in Linguistic Form and Interpretation*. Berlin – New York: Walter de Gruyter: 127-147.
- SCOTT, D. & DE SOUZA, C.S. 1990. Getting the Message across in RST-Based Text Generation. In R. DALE, C. MELLISH & M. ZOCK (eds), *Current Research in Natural Language Generation*. Boston: Academic Press Professional: 47-73.
- SUBBA, R. & DI EUGENIO, B. 2009. An Effective Discourse Parser that Uses Rich Linguistic Information. *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics: 566-574. Available online: http://dl.acm.org/ft_gateway.cfm?id=1620837&type=pdf&CFID=89822543&CFTOKEN=23024412.
- VERGEZ-COURET, M. 2010. *Étude en corpus des réalisations linguistiques de la relation d'Élaboration*. Unpublished PhD thesis. Université de Toulouse.
- WOLF, F. & GIBSON, E. 2006. *Coherence in Natural Language: Data Structures and Applications*. Cambridge (Mass.): MIT Press.