

1

Subjectivity and Causality: A Corpus Study of Spoken Language

WILBERT SPOOREN, TED SANDERS, MIKE HUISKES AND LIESBETH DEGAND¹

1 Introduction

Connectives and the coherence relations they express are among the building blocks of discourse. It is an important issue to uncover the system behind the choice for a particular connective, because it can be seen as an act of categorization: By choosing one connective over another a speaker expresses that the relationship between two discourse segments is to be understood as an instance of one particular coherence relation and not another (Pander Maat and Sanders, 2001). In this paper we study one specific type of linguistic categorization: the way in which Dutch speakers categorize causally related events by expressing them with the connectives *omdat* ('because') and *want* ('for/because').

There is a subtle difference between sentences like (1) and (2), which both express backward causal coherence relations. .

- (1) Jan is niet thuis omdat hij weg moest.
 'Jan is not at home because he had to leave'

¹ Affiliations of the authors: W. Spooren: Vrije Universiteit Amsterdam, The Netherlands. T. Sanders and M. Huiskes: Utrecht institute of Linguistics OTS, Universiteit Utrecht, The Netherlands. L. Degand: Université catholique de Louvain, Belgium. The last author is a Research Associate at the Belgian Fund for Scientific Research (FNRS). This work was partially supported by grant F.R.F.C. n° 2.4535.02 Contact address: w.spooren@let.vu.nl.

- (2) Ik weet zeker dat Jan niet thuis is, want zijn auto is weg.
'I am sure that Jan is not at home, because his car is gone'

In a number of recent analyses (most recently, Degand and Pander Maat, 2003, Pander Maat and Degand, 2001 and Pit, 2003), the differences between (1) and (2) have been expressed in terms of subjectivity: The *want*-relations express a higher degree of subjectivity than the *omdat*-relations. Subjectivity is understood as the degree to which the conceptualizer –the person responsible for the causal relation– is present in the utterance. The notion of subjectivity has received much attention in cognitive linguistics (for instance, Langacker 1990 and Traugot 1995). In a recent discussion, Verhagen (in press) describes notions like subjectivization and perspectivization as follows:

What such notions [as perspective and subjectivity] have in common is that they relate to aspects of conceptualizations that cannot be sufficiently analyzed in terms of properties of the object of conceptualization, but necessarily involve a subject of conceptualization in one way or another (p. 1)

It can be said, then, that (2) is more subjective than (1). At the same time, there is a similarity between the epistemic relation (see Sweetser 1990 for this use of the term) in example (2) and the volitional type of relation in (1). Pander Maat and Sanders (2001) describe this similarity as “both crucially involve an animate subject, a person, whose intentionality is conceptualized as the ultimate source of the causal event, be it an act of reasoning or some “real-world” activity” (p. 251). From this point-of-view, (1) can be considered more subjective than a truly conceptualizer-independent example like (3), which describes a case of real-world physical causality.

- (3) De boom viel om omdat de bliksem insloeg.
'The tree fell because it was struck by lightning'

The notion of subjectivity that is used in this paper, differs therefore to some extent from that used by, e.g., Langacker. We will consider a causal relation subjective if it has its origin in a conceptualizer's mind, be that the speaker or an agent. In example (1) the conceptualizer is *Jan*, in example (2) the conceptualizer is the speaker. In this we will follow Pander Maat and Sanders (2001) and Pit (2003, in press). It is not our purpose in this paper to compare the merits of these different takes on perspective and subjectivity.

Dutch *want* seems to have a preference for subjective contexts, whereas *omdat* fits well in more objective, conceptualizer-independent contexts (note that replacing *omdat* with *want* in (3) changes the interpretation to an epistemic one). What we intend to do in this paper is to explore the relationship and the differences between the two Dutch causal connectives *omdat* and *want*. This is not the first enterprise of that nature, but interestingly enough, empirical studies of *want* and *omdat* to date are corpus studies of written texts. In this contribution we intend to extend the analysis to spoken language. We believe that this is important, for at least four reasons. The first is that extension to spoken language allows for a greater generalizability of the analysis. Second, spoken discourse can be seen as the canonical form of communication through language and consequently an adequate theory of language use should be able to deal with that type of language use (see Clark, 1996 for a similar argument). Third, it is interesting to find out whether the written versus spoken mode lead to a divergent distribution of these causal connectives (as has been found for French *car* and *parce que*, see Labbé, 2003 and Simon and Degand, submitted). Finally, the growing availability of large corpora of spoken language enables researchers to test hypotheses like this on a larger scale. In recent years a new corpus has come to the availability of the Dutch linguistic community, the Corpus of Spoken Dutch (Corpus Gesproken Nederlands, CGN). CGN is a 10 million words corpus of completely digitalized material, annotated in several ways (Oostdijk, 2000). We have used this corpus to analyze fragments with *want* and *omdat* for features of subjectivity. From the Corpus of Spoken Dutch we selected fragments with *want* and fragments with *omdat*. The fragments were analyzed for a number of subjectivity features, such as propositional attitude of the related segments, the nature of the main conceptualizer in the fragment, the syntactic completeness of the utterances, the type of coherence relation, etc. We also made some comparisons with written corpora. To that end we used two existing analyses, by Degand and Pander Maat (2003) and by Spooren, Bekker and Noordman (2001).

What did we expect to find in the analysis? If subjectivity is the right notion to analyze the difference between these connectives, it should go across the modalities of written and spoken language. Therefore our main hypothesis is that *want* expresses more subjective coherence relations than *omdat*. We also expected that the overall degree of subjectivity of spoken discourse is higher than that of written discourse, and that therefore the differences between *want* and *omdat* are smaller in spoken language than in written language.

2 Method

From the corpus, 149 fragments with *want* and 124 fragments with *omdat* were selected.² The corpus contains different text genres. We have only analyzed more or less spontaneous conversations and interviews. For the sake of interrater agreement each fragment was analyzed independently by two coders. In the case of disagreement, the difference was discussed. If the coders did not reach agreement, the fragment was disregarded.

We have looked at both the first and the second segment of the fragments. In this paper we will only report on the results for the first segment, as that is the segment in which we expected to find the clearest differences. Intuitively, in a maximally subjective relation of the type *X because Y* the *X* represents a segment that is particularly subjective; since it is the claim that is argued for.

We determined the amount of subjectivity in the corpus examples by analyzing the following properties, based on suggestions in the literature.

1. Propositional attitude of the first segment (S1)

Each segment was coded as expressing a fact, general knowledge, an intentional act, individual knowledge, a perception, an experience or a judgment, according to the following criteria.

The segment expresses a fact if a state of affairs is described that can be located at a particular moment in time and if it does not contain a conceptualizer who is responsible for the role of the segment in the causal event. Example: *The river has burst its banks.*

The segment expresses general knowledge if the information in the segment is based on general rules and/or is a generalization over times and individuals. Example: *Man is a social animal.*

The segment expresses an intentional act if there is a conscious protagonist who willfully carries out an act at a specific moment in time and place. Example: *I went to the pub.*

The segment expresses individual knowledge if there is a conceptualizer who carries out an act of understanding. This act can be located in time. Example: *Carl knew that it would be of no use anymore.*

The segment expresses an experience if there is a conceptualizer, who is non-agentive and whose experience can be located in time. Example: *Art became ill.*

² There is no principled reason for this difference, other than lack of time. Ultimately we aim at analyzing 200 fragments of each connective.

The segment expresses a perception if there is a non-agentive conceptualizer, if what is conceptualized is an act of perceiving, if the perception can be located in time and if the segment contains an explicit verb of perception (*see, hear, smell*, etc). Example: *He saw that the car hit the tree.*

The segment expresses a judgment if it gives both a conceptualizer and that which is judged. The segment expresses a state and uses a so-called scalar predicate (a predicate that can be modified with degree expressions, such as *very much X; more than X*), which can be paraphrased with “I believe that...”. Example: *That is a pity.*

The propositional attitude of the segments can be ranked in terms of subjectivity: judgments are more subjective (more conceptualizer-dependent) than the other propositional attitudes:

Facts, general knowledge, intentional act, individual knowledge. / perception, experience < judgment

2. Relation type

The causal relation expressed in each fragment was analyzed in terms of domains (Sweetser, 1990): content (in which the speaker describes a causal relation in the world), epistemic (in which the speaker infers a conclusion on the basis of an argument) and speech act relations (in which the speaker motivates a speech act). Furthermore, within the content relations we distinguish between volitional and non-volitional relations: Does the relation involve an intentional act or not? Examples are:

- (4) Non-volitional content
The house burnt down because it was struck by lightning.
- (5) Volitional content
He went home because he was ill.
- (6) Epistemic
He must be on holiday because his car is gone.
- (7) Speech act
Can you hand me my coffee? I'm busy.

The causal relations can be ordered from least subjective to most subjective, as follows:

Non-volitional content . Volitional content < Epistemic, Speech act

3. Type of conceptualizer in the first segment (S1)

The conceptualizer is the person responsible for the causal relation that is constructed. In spoken dialogues there can be either no conceptualizer (as in example (4)), or the conceptualizer is a third person (examples (5) and (6)), a second person (example (8)) or a first person (example (7)).

(8) Second person conceptualizer

So you think that she will come back, because she loves him?

An interesting case is example (9):

(9) Hij is vast naar huis gegaan, want hij was ziek.

'He must have gone home, because he was ill'

Here we have a third person pronoun in combination with the speaker as conceptualizer. These cases have not been analyzed as third person conceptualizers, but as implicit conceptualizers (see the next subsection).

The different types of conceptualizers can be ordered in degree of subjectivity, as follows:

No Conceptualizer. Third person < Second person, First person

4. Linguistic realization of the conceptualizer (S1)

The final property we will report on is the linguistic realization of the conceptualizer. We have followed Langacker's (1990) suggestion that an explicit reference to the conceptualizer objectifies the conceptualizer. Consequently, implicit reference to the conceptualizer is more subjective than explicit reference:

Explicit reference to the conceptualizer < Implicit reference to the conceptualizer

In addition we have analyzed other properties of the corpus segments, that we expected to be relevant to determine their subjectivity. Among these other characteristics are Speaker continuity, position of the connective, syntactic completeness of the fragment, size of the segments and discourse genre. We will not discuss them here.

3 Results: comparing two connectives

The general hypothesis was that *want*-fragments show a higher degree of subjectivity than *omdat*-fragments. This hypothesis will be discussed for each of the four properties that we have looked at. Because of the low cell frequencies we have aggregated groups in a theoretically motivated way. We used chi2 to test (in)dependency relations between the various categories.

3.1 Propositional attitude

Our subhypothesis with respect to propositional attitude is that *want*-fragments occur relatively more often with judgments in S1 than *omdat*-fragments. This hypothesis is borne out ($\chi^2(1) = 4.78, p < .05$). Table 1 summarizes the data.

Table 1. *Number of judgments vs. other propositional attitudes as a function of connective (want, omdat) in the spoken corpus (for seven omdat-fragments and one want-fragment the data are missing).*

	judgments	other propositional attitudes
Omdat	49 (41.9 %)	68 (58.1 %)
Want	82 (55.4 %)	66 (44.6 %)

3.2 Type of coherence relation

With respect to coherence relation we expected to find epistemic and speech act relations to occur typically with *want* and non-volitional cause and volitional reason with *omdat*. In other words, we expected to find more content relations with *omdat* than with *want*. This hypothesis is borne out, as can be seen in Table 2 ($\chi^2(1) = 64.37, p < .01$).

Table 2. *Number of content relations vs. epistemic/speech act relations as a function of connective (want, omdat) in the spoken corpus.*³

³ In seven *omdat*-fragments and six *want*-fragments, the relation was a so-called textual relationship, in which the second segment gives a paraphrase or a metacomment. These have been discarded.

	Content	epistemic/speech act
omdat	106 (90.6 %)	11 (9.4 %)
want	61 (42.7 %)	82 (57.3 %)

3.3 Type of conceptualizer

With respect to type of conceptualizer we expected to find more facts and third person conceptualizers with *omdat* and more first and second person conceptualizers with *want*. This hypothesis has to be rejected on the basis of the data ($\chi^2(1) = .52, p = .47$), as can be seen in Table 3.

Table 3. *Number of facts/3rd person conceptualizer vs. 2nd/1st person conceptualizer as a function of connective (want, omdat) in the spoken corpus.*

	fact/3rd person	2 nd /1st person
omdat	25 (21.4 %)	92 (78.6 %)
want	37 (25.2 %)	110 (74.8 %)

3.4 Coding of the conceptualizer

We expected the *omdat*-fragments to have more explicit codings of the conceptualizer than the *want*-fragments. In this analysis we excluded the fragments with a fact, because there is no conceptualizer in these cases. Table 4 gives a summary of the data, which show that the hypothesis is supported ($\chi^2(1) = 5.07, p < .05$).

Table 4. *Number of explicit vs. implicit conceptualizers as a function of connective (want, omdat) in the spoken corpus (NB. facts were excluded from the analysis).*

	explicit	implicit
omdat	80 (74.1 %)	28 (25.9 %)
want	84 (60.4 %)	55 (39.6 %)

3.5 Conclusion

There is substantial evidence that *want* is more subjective than *omdat* in our spoken corpus. Compared to *omdat*, *want* occurs more often with judg-

ments, with non-content relations and with implicit conceptualizers. The one exception is the identity of the conceptualizer, on which characteristic *want* and *omdat* do not show any difference. It has been noted (Pit, 2003) that *omdat* is often used to express explicit acts of speaking and thinking ('I assume this because...', 'I say this because ...'). This use of *omdat* might account for the high frequency of first person pronouns with *omdat*. If that is correct then the frequency of first person pronouns was not a good operationalization of high degree of subjectivity. We could even argue that these constructions show a low degree of subjectivity: The speaker describes his or her speaking or thinking and thereby creates a distance (see also Langacker, 1990).

4 Results: comparing spoken and written language

Our second hypothesis deals with the difference between spoken and written language. Our hypothesis is based on the assumption that the genres that we have analyzed differ with respect to the prominence of the speaker and addressee: In spoken discourse they are much more salient than in written discourse. Therefore, the speaker / writer is more clearly present in spoken discourse. This makes the overall degree of subjectivity of spoken discourse higher than that of written discourse. As for the comparison of *want* and *omdat*, this implies a smaller difference in *subjectivity* for the spoken corpus than for the written corpus. In other words, we expected a three-way interaction between type of connective, type of corpus and subjectivity: We expected that the differences between *want* and *omdat* would be smaller in spoken than in written corpora.

In order to be able to study higher order relations, we used loglinear analyses. Loglinear analyses are appropriate to study multi-way contingency tables like the ones presented below. The aim of loglinear analysis is to set up a model of the minimum amount of factors (main effects and interactions) that fits the data best. An introduction to loglinear analysis can be found in Rietveld and Van Hout (1993).

First of all, we want to compare the frequency of the connectives in the two types of corpora. As can be seen from table 5, *omdat* is more frequent in written language, *want* much more frequent in spoken language (since these are standardized frequencies, we have not calculated a χ^2 for this table).

Table 5. *Frequency per million words of want and omdat in written and spoken corpora.*

	written ^a	spoken ^b
omdat	938	521
want	686	1640

Notes: ^a. Based on data in Bestgen et al. (in press). ^b. Frequency in Corpus of Spoken Dutch in the selected genres.

We used loglinear analyses to analyze the data. We can only make a limited number of comparisons because our data about written language come from different authors, who have used different analytic categories. For our first three comparisons we have used a study by Degand & Pander Maat (2003).

4.1 Medium, connective and coherence relation

Our first comparison looks at the relationship with type of coherence relation. Table 6 summarizes the data, and table 7 summarizes the loglinear analysis.

Table 6. *Type of relation in written and spoken corpus, by connective.*

	written ^a		spoken	
	content	epist./sp. act	content	epist/sp. act
omdat	26	24	106	11
want	8	42	61	82

Note a: Data based on Degand & Pander Maat (2003).

Table 7. *Summary of loglinear analysis (Connective, Medium, Relation).*

Factor	DF	Partial X ²	Prob
Medium	1	73.660	.0000
Connect	1	1.879	.1704
Relation	1	4.911	.0267
Connect x Medium	1	13.851	.0002
Connect x Relation	1	84.143	.0000
Medium x Relation	1	39.992	.0000
Connect x Medium x Relation	1	1.853	.1735

The loglinear analysis shows that the data are best described with a model containing two-way interactions and two main effects. The main effect of Type of Medium is trivial in that it reflects the fact that we have more fragments from the spoken corpus than from the written corpus. The main effect of Type of Relation reflects the fact that there are more content relations than non-content relations. Of the two-way interactions, the relationship between Type of Connective and Type of Medium is trivial, as it is

determined by the way the corpora were collected. The relationship between Type of Connective and Type of Relation is that *omdat* occurs much more often with content relations (132 of 167 or 79.1 %), whereas *want* occurs more often with non-content relations (124 of 193 or 64.2 %). The relationship between Type of Medium and Type of Relation reflects the fact that in the written corpus the non-content relations are more frequent (66 out of 100 or 66.0 %), whereas in the spoken corpus the content relations are more frequent (167 of 260 or 64.2 %). Note that this latter finding is in contrast with our assumption about the overall higher degree of subjectivity of the spoken corpus. Note also that the expected three-way interaction was not found.

4.2 Medium, Connective and Identity of the Conceptualizer

The second property we compared was the identity of the conceptualizer. This analysis was restricted to volitional and epistemic relations by Degand & Pander Maat (2003). For the sake of comparability we have followed this restriction. Table 8 summarizes the data and table 9 summarizes the loglinear analysis.

Table 8. *Type of conceptualizer in the first segment in volitional and epistemic relations, by corpus and by connective (column percentages in parentheses).*

	Written		Spoken	
	omdat n=40	want n=45	omdat n=85	want n=106
First person	25 (62.5 %)	39 (86.7 %)	67 (78.8 %)	82 (77.4 %)
Third person	15 (37.5 %)	6 (13.3 %)	18 (21.2 %)	24 (22.6 %)

Note. In the spoken data, the fragments in which the identity of conceptualizer is the addressee (2nd person) have been disregarded.

Table 9. *Summary of loglinear analysis (Connective, Medium, Type of Conceptualizer).*

Factor	DF	Partial χ^2	Prob
Medium x Connect	1	.127	.7215
Medium x Concept	1	.216	.6421
Connect x Concept	1	1.623	.2027
Medium	1	41.775	.0000
Connect	1	2.453	.1173
Concept	1	86.101	.0000
Connect x Medium x Concept	1	5.214	.0224

The analysis shows that there are main effects of Type of Medium (which reflects the trivial difference that there are more spoken fragments than written fragments) and Type of Conceptualizer (there are more first person conceptualizers than third person conceptualizers in the corpora). None of the two way interactions is significant. The best fitting model is one which incorporates all three factors. Inspection of the frequencies in table 8 shows that in the spoken corpus there is no difference between *want* and *omdat* with respect to the identity of the conceptualizer (as we described in section 3.3, both *want* and *omdat* have predominantly first person conceptualizers), in written language there IS such a difference: the predominance of first person conceptualizers is stronger for *want* than for *omdat*. This is in accordance with the hypothesis.

4.3 Medium, Connective and Linguistic Expression of the Conceptualizer

The final comparison concerned the linguistic expression of the conceptualizer, which can be either explicit or implicit. Degand and Pander Maat have restricted their analysis to epistemic relations, a restriction which we have followed. In view of the limited number of data we did not carry out a log-linear analysis and restricted ourselves to a more descriptive report. Table 10 summarizes the data.

There is a difference between the written and the spoken data: In the written corpus, the conceptualizers remain predominantly implicit, which is not the case in spoken language. However, there is no three-way interaction: in each of the two media, written and spoken, *omdat* and *want* behave similarly.

Table 10. *Linguistic realization of conceptualizer in the first segment of epistemic relations, by type of corpus and type of connective.*

	Written		Spoken	
	omdat n=24	want n=37	omdat n=7	want n=50
Explicit	5	3	3	25
Implicit	19	34	4	25

Note: In the spoken data, the fragments in which the first segment expresses a fact have been disregarded.

4.4 Conclusion

In sum, the question whether the difference between *want* and *omdat* is smaller in spoken language than in written language seems to require a negative answer. With respect to type of coherence relation and linguistic realization of the conceptualizer there was no indication of the predicted three-way interaction. With respect to identity of conceptualizer, there was such a relationship, but the effect was not very strong.

5 General conclusion

In conclusion, the first hypothesis, namely that *want* and *omdat* differ in terms of subjectivity in the spoken corpus, is strongly supported by the data. We find a number of differences in subjectivity between *want* and *omdat*. The only characteristic on which *want* and *omdat* do not differ is the identity of the conceptualizer.

With respect to the second hypothesis, namely that the difference between *want* and *omdat* is smaller in spoken language than in written language, the picture is less clear. There is some evidence supporting the hypothesis, namely the identity of the conceptualizer: In spoken language there is no difference between *want* and *omdat*, in written language there is such a difference.

A difference that we did find concerned the relative frequency of *want* and *omdat*: *Want* appears to be the default causal connective for spoken language, *omdat* is the default for written language. Furthermore, *want* can be used to express almost all types of relations.

Let us try to formulate an interpretation. In the spoken genres that we have analyzed, it looks as if choosing *want* is an attractive option for the speaker. The reason might be one of speaker economy (Knott and Sanders, 1998): Speakers tend to choose the connective that expresses a default ‘general’ causal relation because it costs less energy. This is especially the case in the context of spontaneous conversations, characterized by limiting time constraints and few planning and editing options. From the point of view of speaker economy, using *want* has several advantages. One is that *want* can be used to express a wide variety of coherence relations. Another is that, if the speaker is planning a straightforward main clause, there is no need to change the syntactic frame, since *want* is a coordinative conjunction. The connective *omdat* is more specific in this respect, as it is a subordinator, which needs verb final word order in Dutch.

On a side note, the question is whether this is a stable situation. We have encountered a significant number of cases of *omdat* with main clause

order. An example is (10), in which utterance 231 has main clause word order, despite the fact that is introduced by the subordinator *omdat*

- (10)
 230 N01134 oh maar ggg we doen er niet zo heel moeilijk over.
 ‘oh but ggg we are not very difficult about that’
 231 N01135 ja xxx OMDAT ze die heeft 't ook altijd wel moeilijk om
 moeite om snel tot bijvoorbeeld de clou te komen zullen
 we maar zeggen.
 ‘yes xxx BECAUSE she she always has some problems
 to for example let’s say get to the point’

Such main clause orders with *omdat* are absent in the writings of adult, proficient writers and it is our impression that its occurrence in spoken Dutch is a recent development. It may well reflect the same change for Dutch *omdat* that has been reported for German, where the subordinate connective *weil* seems to be taking the position of the coordinate connective *denn*, so it also gets main clause word order. This development was also first found in the analysis of spontaneous discourse (Günthner 1993, Keller 1995).

In sum, the difference between *want* and *omdat* can be described in terms of subjectivity, irrespective of the medium. What differences there are between written and spoken language do not relate so much to a different degree of subjectivity, but to a different amount of planning possibilities in spoken language.

Our interpretation is the breeding ground of a number of questions, which await future research. The most urgent one is, we think, the following: If the difference between spoken and written language is a difference in planning rather than a difference in subjectivity, do we find differences in the use of *omdat* and *want* in spoken genres that vary in amount of planning and editing? What we intend to do in the future is look at genres like sports reports (which require extremely low level of planning) versus prepared lecture (which require extremely high level of planning; for interesting first results in this direction see Troost, Jansen and Sanders, under review).

References

- Bestgen, Y., Degand, L., & Spooren, W. in press. Towards automatic determination of the semantics of connectives in large newspaper corpora. *Discourse Processes*.

- Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Degand, L. & H. Pander Maat 2003. A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. *Usage Based Approaches to Dutch*, eds. A. Verhagen & J. van de Weijer, 175-199. Utrecht: LOT.
- Günthner, S. 1993. '...weil - man kann es ja wissenschaftlich untersuchen' - Diskurspragmatische Aspekte der Wortstellung in WEIL-Sätzen. ['...because - it can be analysed scientifically' - Discourse pragmatic aspects of word order in weil-sentences], *Linguistische Berichte*, 143, 37-59.
- Keller, R. 1995. The epistemic 'Weil'. *Subjectivity and Subjectivisation: Linguistic Perspectives*, eds. D. Stein & D. Wright, 16-30. Cambridge: Cambridge University Press.
- Knott, A. & T. Sanders 1998. The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics* 30: 135-175.
- Labbé, D. 2003. Coordination et Subordination en Français Oral. Presentation at IVE Journées de l'ERLA, *Coordination/Subordination en Français de Spécialité*, Brest, 14-15 november 2003 (Also available as <http://www.upmf-grenoble.fr/cerat/Recherche/PagesPerso/LabbeBrest03.pdf>).
- Langacker, R. 1990. Subjectification. *Cognitive Linguistics* 1: 5-38.
- Oostdijk, N. 2000. The Spoken Dutch Corpus Project. *The ELRA Newsletter* 5 (2): 4-8.
- Pander Maat, H. & L. Degand 2001. Scaling causal relations and connectives in terms of Speaker Involvement. *Cognitive Linguistics*, 12 (3): 211-245.
- Pander Maat, H. & T. Sanders 2001. Subjectivity in causal connectives; An empirical study of language in use. *Cognitive Linguistics*, 12 (3): 247-273.
- Pit, M. 2003. *How to Express Yourself with a Causal Connective. Subjectivity and Causal Connectives in Dutch, German and French*. Amsterdam: Rodopi. Doctoral Dissertation, Utrecht University.
- Pit, M. in press. Determining subjectivity in text. The case of backward causal connectives in Dutch. *Discourse Processes*.
- Rietveld, T. & R. van Hout 1993. *Statistical Techniques for the Study of Language and Language Behaviour*. Berlin etc.: Mouton de Gruyter.
- Simon, A-C. & L. Degand submitted. Connecteurs de causalité, implication du locuteur et profils prosodiques. Le cas de *car* et *parce que*. [Causal connectives, speaker involvement and prosodic contours. The case of *car* and *parce que*].

- Spooren, W., B. Bekker & L. Noordman 2001. Reversed order and subjectivity in different text types. *Multidisciplinary Approaches to Discourse* eds. L. Degand, Y. Bestgen, W. Spooren, & L. Van Waes, 61-72. Amsterdam: Stichting Neerlandistiek VU.
- Sweetser, E. V. (1990). *From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.
- Traugott, E.C. 1995. Subjectification in grammaticalization. In: *Subjectivity and Subjectivisation* eds. D. Stein and S. Wright, 37-54.: Cambridge University Press.
- Troost, C, F. Jansen & T. Sanders under review. Bepaalt cognitieve economie de plaats van temporele en causale bijzinnen? [Does Cognitive Economy Determine the Location of Temporal and Causal Subordinate Clauses?].
- Verhagen, A. in press. Construal and perspectivisation. *Handbook of Cognitive Linguistics* eds D. Geeraerts and H. Cuyckens: Oxford: Oxford University Press. Available as PDF-file from <http://www.arieverhagen.nl/Documenten/Publications.htm>