

A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale

Liesbeth Degand (University of Louvain, Louvain-la-Neuve) &
Henk Pander Maat (Utrecht University)

1. Introduction

It is generally acknowledged that coherence relations and their linguistic markers can express meanings at different “levels” of the discourse. These differences in meaning and use have mainly been accounted for in terms of dichotomies: for instance, external/internal (Halliday & Hasan 1976; Martin 1992), semantic/pragmatic (van Dijk 1979; Moeschler 1989), subject matter/presentational (Mann & Thompson 1988). A very influential account is also Sweetser (1990), who has proposed to distinguish not two, but three domains of use for connectives: the content domain, relevant for (1), the epistemic domain, exemplified in (2), and the speech act domain, illustrated in (3).

- (1) John came back because he loves her.
- (2) John loves her, because he came back.
- (3) What are you doing tonight, because there’s a good movie on.

In previous work (Degand & Pander Maat 1999; Pander Maat & Degand 2001) we have proposed an alternative account of the distribution of causal connectives. Going beyond dichotomous and trichotomous classifications, we have proposed to represent (causal) coherence relations and connectives in a scalar way. This scalar representation reflects the fact that (causal) connectives are not strictly domain specific, but that they nevertheless impose constraints on the contexts in which they can occur, with some contexts being more “natural” than others¹ (see also Degand

¹ Actually, there are some connectives that seem to be rather domain-specific, like Dutch *daardoor* and *doordat* which are highly restricted to the non-volitional domain, but such a domain specificity seems to be the exception rather than the rule.

& Sanders 1999). In addition, a number of causal connectives seems to take an intermediate position between the traditional categories. According to us, this situation is an indication for the need of a scalar perspective on the spectrum reaching from non-volitional causality in the content domain to epistemic and speech act causality.

The scale we have developed is one of *speaker involvement*, on which the inherent expressive power of connectives can be represented. Our hypothesis is that the different causal relations can be ordered along a scale from minimal to maximal *speaker involvement*. Speaker involvement refers to *the degree to which the present speaker is implicitly involved in the construal of the causal relation*. More specifically, speaker involvement increases with the degree to which both the causal relation and the related segments carry assumptions and actions of the present speaker (see below). The different causal relations we distinguish are, in order of increasing Speaker Involvement: causal non-volitional and volitional content relations; causality-based and non-causality based epistemic relations, and causal speech-act relations. Pander Maat & Degand (2001) give a detailed account of this relational ordering on the scale. Here it will suffice to give an overview of the characteristics that come into play when determining the level of Speaker Involvement of a causal relation. Afterwards we will discuss the interaction between a number of Dutch and French causal connectives and the Speaker Involvement scale, and support our claims on the basis of a corpus analysis.

2. Speaker Involvement and causal relations

Four characteristics of coherence relations may enhance the prominence of speaker assumptions in the relation, and thus enhance the level of Speaker Involvement of the relation: The involvement of a protagonist, a lack of isomorphism between the relation and states of affairs in the real world, proximity of the relation to the present speaker and the time of speaking and the implicit realization of the protagonist.

First, causal coherence relations may be characterized in terms of the degree to which they necessarily imply the *subjective involvement* of a conscious participant, which may, but need not be the speaker. For instance: a volitional relation (see 4 below) involves such a participant (an actor), while a non-volitional relation (5) does not. Since conscious

participants may entertain assumptions and may be identified with by the speaker, volitional relations carry a higher degree of SI than non-volitional ones. Epistemic relations (6) are still higher on the scale because they involve beliefs of a concluding protagonist with whom the speaker must share a number of assumptions in order to understand the causal relation² (see examples 2 and 6). In fact the hearer must also share these assumptions to a certain extent (see also Pander Maat & Degand 2001). Finally, maximal SI is reached with speech-act relations in which the protagonist is identical to the Speaker by definition. In a speech-act relation the speech-act is motivated by reference to a situation constituting the reason for it (as in example 3 above).

- (4) He left, because he felt tired.
- (5) The temperature went up, because of the fact that the sun came up.
- (6) The children should be punished, because they have been mean.

Second, a given causal relation may be more or less *isomorphic with a real world causal relation*, that is, a causal relation in which the present speaker has no role. For instance, the volitional relation in (1) represents a real world causal relation. The same causal situation is present in (7) below, but this time it is no longer the primary causal relation expressed. In (7), a reason-consequence relation in the real world is transposed into the mental domain of making inferences, or in other words it is used to base an argumentation by the speaker upon. This is why the epistemic relation in (7) carries a higher degree of SI than the volitional relation in (1), repeated for convenience below.

- (7) John must have come back, since he loves her.
- (1) John came back, because he loves her.

In this kind of epistemic relation, real world causality is not represented, but it continues to impose constraints on the propositional content of the related segments. This is no longer the case for non-causal epistemic relations and speech act relations, which show maximal detachment from real world causality. Two examples of non-causal epistemic relations are

² In most cases, this concluding protagonist *is* the speaker himself.

presented in (8) and (2). In (8) the conclusion does not refer to a probable consequence of the situation referred in the premise, but to an evaluation by the speaker. Since this evaluation is not a real world situation, there can be no real world causality involved in its coming about. Hence, it can only be based on ‘inferential’ causality, operating in the domain of reasoning. In example (2), repeated below for convenience, this movement of detachment from real world causality is even more apparent: in this example, the real world causal relation is ‘turned around’ to yield a pattern of abductive inference. In the empirical study presented below, abductive epistemic relations will be discussed separately from other non-causal epistemic relations, since they will turn out to discriminate between connectives.

- (8) John just told me that he will accept our offer. This is an important message, since we assumed that he would not accept.
 (2) John loves her, because he came back.

Third, a given relation may be placed at *different distances from the present speaker*, and at different distances from the moment of speaking. The closer a given relation is to the present speaker, the more it constitutes a vehicle for the expression of speaker assumptions. After all, the speaker is more likely to accept the general assumptions underlying his own decisions than those underlying other persons’ decisions. Hence, the first-person relation in (10) has a higher degree of SI than the third-person relation in (9). And (11) is even higher in SI because it has present tense.

- (9) He felt tired. He left.
 (10) I felt tired. I left.
 (11) I feel tired. I’m going home.

In more general terms, SI increases when the distance between the present speaker and the protagonist decreases. In non-volitional relations SI is minimal because there is no protagonist to identify with. In volitional relations the situation is as sketched in examples (9-11). In epistemic relations the speaker is the unmarked protagonist and in speech act relations he is the only conceivable protagonist. This decreasing distance between speaker and causal protagonist leads to an increasing SI

from non-volitional to speech-act relations. The same happens with regard to the time distance to the moment of speaking: SI increases when the distance between speaking time and causal situation time decreases. Hence non-volitional and volitional relations are lowest on the SI scale because there are no temporal constraints on the expression of these types of causal relation with respect to the speaking time, i.e. the time of the causal event and the speaking time are independent. Speech act relations are highest on the SI scale because causal situation time (t_2) and speaking time (t_s) obligatory coincide, i.e. a speech-act cannot take place at another time than the speaking time. As for the epistemic relations, t_2 and t_s very often coincide but this is not obligatory the case, i.e. generally (speaker) conclusions are presented to be valid at the time of speaking but this is not necessarily so.

Fourth, the involvement of a conscious participant may vary in *explicitness*. We will illustrate this dimension with reference to the presence of the speaker in an epistemic relation.

- (12) He is Hungarian.
- (13) He is probably Hungarian.
- (14) I think he is Hungarian.

Taken in isolation, (12) expresses a factual statement that is produced outside the domain of the speaker's conceptualizing activity. That is, it presents itself as carrying a minimal degree of SI, because the speaker is only minimally involved in the production of the reported cognition. In contrast, (13) and (14) contain some explicit elements referring to the speaker's perspective. In example (14) the modal adverb *probably* invokes the speaker as the source for the probability judgement. This is even more explicit in (14) where the speaker's perspective *I think* is more or less "objectified". This objectification appears from the fact that it may be referred to by demonstrative pronouns in subsequent utterances. E.g. *that is not true* would refer to the *I think* clause, not to its complement. At this sentence level, (13) has a higher SI than (14), which is still higher than (12).

However, factual statements like (12) can also occur in a discourse context that heavily suggests an interpretation as a speaker conclusion.

For instance, this would be the case if (12) would be preceded by *he can only be Czech or Hungarian and I am sure he is not Czech*. In this case the speaker's inferential activity is part of the relational interpretation, but at the same time it is maximally implicit. Hence, this time it has a maximal degree of SI, higher than (13) and still higher than (14).

In sum, the involvement of the protagonist may be more or less implicitly realized in the discourse. As we proceed on the scale, explicit realization (in volitional relations) gradually changes into possibly implicit realization (in epistemic relations) and to obligatory implicit realization (in speech act relations).

Together the different dimensions mentioned determine the level of SI present in a causal relation³, that is the degree to which the relation is a vehicle for the expression of the speaker's assumptions and activities.

3. Speaker Involvement and causal connectives

We claim that the Speaker Involvement Scale is not only a way to classify coherence relations but that connectives – as prototypical markers of coherence relations – can be analyzed in terms of the scale too. One may expect that a valid classification of coherence relations, which are considered as cognitive entities, has some counterpart in language. After all, there are restrictions on the use of connectives. And there are many correspondences between connectives and the relational features they signal. Even though there is no one-to-one mapping between relations and connectives, the restrictions on the use of connectives imply an organization of the relations they can express; they do not just co-exist as a set of relations on one and the same level.

In previous work (Degand & Pander Maat 1999; Pander Maat & Degand 2001) we have demonstrated that the Speaker Involvement Scale can be used to analyze and classify forward causal connectives in Dutch and French. In particular, we have supported the following claims concerning the relationship between connectives and SI:

³ Of course, some of these dimensions may correlate. For instance, the dimension “identity of the protagonist” correlates with the dimension “realisation of the protagonist”, since third-person-protagonists are more often explicit than first-person-protagonists. However, correlation is not determination.

1. A connective *encodes a certain SI level*, which it contributes to the interpretation of its discourse environment. When this level is too low or too high to be combined with the level allowed for by the discourse environment, the use of the connective is inappropriate.
2. The set of relational environments of a connective can be represented as an area on the SI scale. The distribution of every connective occupies a *contiguous area* on the scale. That is, we do not expect to find a single connective that may express, e.g., volitional and non-causal epistemic relations but not causal epistemic relations.
3. The most frequent causal connectives in a given language should *differ significantly* from each other on the scale. That is, the prime reason for a language to have more than one causal connective is to be able to express several levels of SI.
4. These claims are *cross-linguistically valid* (starting with Dutch and French).

In the remainder of this paper, we will extend this account to three backward causal connectives in Dutch and French: *omdat/parce que* ('because'), *want/car* ('because/for'), *aangezien/puisque* ('since'). By 'backward causal connectives' we refer to connectives which are placed in the antecedent-segment of the relation. This does not always imply that this segment follows the consequent-segment, as has been the case in the examples given so far. In fact, some of the connectives are more or less evenly distributed over the two orders consequent-antecedent and antecedent-consequent (see Table 1 below).

The purposes of our study are twofold. First, we want to check whether the SI scale, which was developed on the basis of forward causal relations/connectives can also be used to describe backward connectives. If this were not the case, the generality of the scale would be doubtful. Second, there is a striking frequency difference between the backward translation "equivalents" in Dutch and French. Figure 1 displays the frequencies of the connectives under investigation in Dutch and French newspaper corpora.⁴

⁴ The corpora we used are electronic transcriptions of two daily newspapers. For Dutch, the 1994 issues of *NRC Handelsblad*, a Dutch national newspaper, provided by the INL (Leiden Institute of Dutch Lexicology) containing about

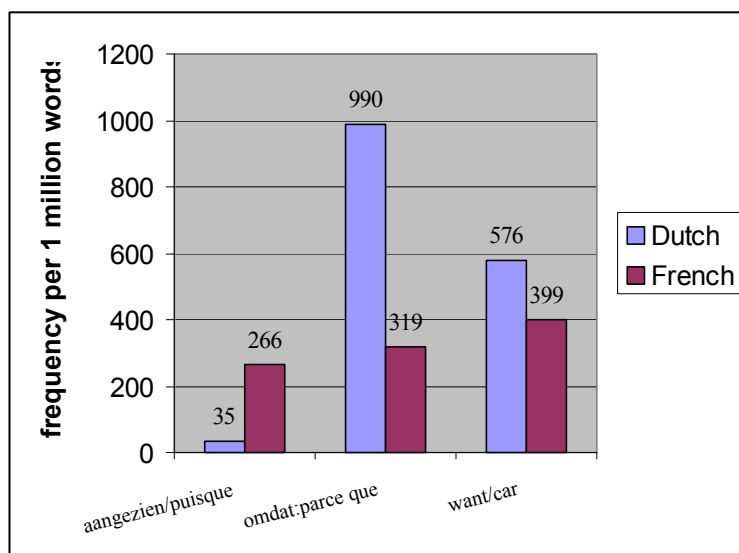


Figure 1: Frequency of Dutch and French causal connectives in newspaper corpora

The most striking frequency divergences occur with *aangezien* and *puisque*, since *puisque* is more than seven times more frequent than *aangezien*. *Omdat* and *parce que* also show highly diverging frequency patterns *omdat* being more than three times more frequent than *parce que* in comparable newspaper corpora. In addition, there is another intriguing difference between the connectives *puisque* and *aangezien*, namely a divergent syntactic pattern with respect to causal segment ordering. In Dutch, *aangezien* occurs nearly as often in antecedent-consequent (*aangezien Q, P*) ordering as in consequent-antecedent ordering (*P aangezien Q*). In French, however, *puisque* occurs nearly exclusively in consequent-antecedent ordering (*P puisque Q*). *Omdat* and *parce que* which also admit the two orderings do not show any divergences here (see Table 1).

27 million words. For French, the 1997 issues of *Le Soir*, a Belgian francophone newspaper distributed on CD-ROM containing about 26.8 million words.

	<i>aangezien</i>	<i>puisque</i>	<i>omdat</i>	<i>parce que</i>
P conn Q	27 (54%)	48 (96%)	45 (90%)	45 (90%)
conn Q, P	23 (46%)	2 (4%)	5 (10%)	5 (10%)
	50	50	50	50

Table 1: *Ordering of causal segments*

At the same time, in the traditional linguistic literature (e.g. Groupe λ -1 1975; Iordanskaja 1993; van Belle 1989), these connectives are described in similar terms. For instance, both *puisque* and *aangezien* are described as “utterance markers” rather than “operators”, and the antecedent is presented as already known to the hearer, while the speaker is not supposed to accept the truth of it automatically. This brings us to the question whether there are subtle differences between these backward connectives that could be accounted for in terms of the Speaker Involvement scale, but not on the basis of traditional descriptions.

3.1 A contrastive SI analysis of causal connectives

In our view, analyses of the SI potential inherent to connectives cannot do without systematic corpus analyses. Indeed, while attractive parallels and suggestive similarities between coherence relations and the linguistic devices that express them have been claimed, recent corpus studies do also reveal that existing categorization proposals *cannot* account for the data of connective distribution in a straightforward way (Degand 1998, 2001; Pander Maat 1998; Pander Maat & Sanders 1995). Hence, we carried out corpus analyses of the backward causal connectives in (written) Dutch and French: *omdat/parce que*, *want/car*, *aangezien/puisque*. For each of these connectives, we assembled 50 occurrences from a newspaper corpus.

First, we identified the coherence relation of each fragment by means of a paraphrase test.⁵ The results for Dutch and French are given

⁵ The non-volitional paraphrase was «this has/had the following cause»; the volitional paraphrase was «this action is/was the consequence of the following» and the epistemic paraphrase was «this conclusion follows from the following». For speech-act relations, two kinds of paraphrases were used: «(the use of) this word/phrase is motivated by the following» and «this speech-act is the consequence of/is motivated by the following». Finally, non-causal epistemic rela-

in Tables 2 and 3, respectively. In the two languages, the contiguity hypothesis (the second claim in section 3) is confirmed.

Relation	connective		
	<i>aangezien</i>	<i>want</i>	<i>omdat</i>
non-vol. causal	2 (4%)	1 (2%)	10 (20%)
vol. causal	8 (16%)	7 (14%)	16 (32 %)
causal epistemic	26 (52%)	20 (40%)	20 (40%)
non-causal epistemic	14 (28%)	17 (34%)	4 (8%)
abductive epistemic		1 (2%)	
speech act		4 (8%)	
	50	50	50

Table 2: *Relational interpretations co-occurring with backward causal connectives in a Dutch newspaper corpus*

For Dutch, *aangezien*- and *want*-fragments are clearly higher on the SI-scale in terms of relational interpretations than *omdat*-fragments. The two connectives do indeed differ significantly from *omdat* but not from one another (*aangezien-omdat*: $\chi^2 = 14.34$, $df = 3$ and $p < .01$; *want-omdat*: $\chi^2 = 23.94$, $df = 5$ and $p < .0001$; *aangezien-want*: $\chi^2 = 6.47$, $df = 5$ and $p > .05$; N.S.). This means that the third claim in section 3 is only partially supported in Dutch. However, since *aangezien* is fairly infrequent in Dutch, there is no functional motivation for *aangezien* to be completely distinctive from its backward alternatives. We will discuss this aspect below.

tions were distinguished from causally-based epistemic relations by checking whether substituting a ‘real-world’ causal paraphrase (non-volitional or volitional) results in a coherent sequence (though the meaning differs from the original fragment). If yes, the relation is causally-based, if not it is non-causal.

Relation	connective		
	<i>puisque</i>	<i>car</i>	<i>parce que</i>
non-vol. causal		1 (2%)	11 (22%)
vol. causal	1 (2%)	10 (20%)	23 (46%)
causal epistemic	36 (72%)	27 (54%)	14 (28%)
non-causal epistemic	10 (20%)	9 (18%)	2 (4%)
abductive epistemic	1 (2%)		
speech act	2 (4%)	3 (6%)	
	50	50	50

Table 3: *Relational interpretations co-occurring with backward causal connectives in a French newspaper corpus*

For French, *puisque*-fragments are highest on the SI-scale in terms of relational interpretations, followed by *car*-fragments, which in turn are more subjective than *parce que*-fragments. All connectives differ significantly from one another. (*puisque-car*: $\chi^2 = 12.70$, $df = 6$ and $p < .05$; *puisque-parce que*: $\chi^2 = 49.18$, $df = 6$ and $p < .0001$; *car-parce que*: $\chi^2 = 25.03$, $df = 4$ and $p < .0001$). Hence our third claim is supported for French.⁶

So far, we contrasted the connectives in terms of the relations they occur in. These relational interpretations are the result of an interaction between the connective and the connected discourse segments. But what we are looking for here is the contribution of the connective to this interaction. We have characterized this contribution in terms of ‘a certain SI level that is added to the interpretation of its discourse environment’. Now, how do we determine the SI level encoded by a connective? A first way to proceed would be to postulate that the relational interpretations of the fragments containing a connective also constitute the area on the SI scale covered by the connective, i.e. its specific SI level. However, this line of reasoning only leads to the unsatisfactory conclusion that the SI areas of the different connectives show considerable overlap. For

⁶ Abductive epistemic relations are a specific category of non-causal epistemic relations. Hence, the absence of abductive uses for the connective *car* (Table 3) does not constitute a ‘gap’ in the scale. Furthermore, *car* can be used in such a context. E.g. *Il doit avoir faim, car il a déjà mangé trois sandwiches*. (‘He must be hungry, because he has already eaten three sandwiches’). So, the contiguity hypothesis (claim 2) is also supported for French.

instance, the distinction between causal volitional, causality-based epistemic and non-causality based epistemic relations is not straightforwardly lexicalized for the backward causal connectives, neither in Dutch nor in French.

However, the distribution shows ‘peaks’ at different points. A first clue to the different SI profiles of the connectives may be obtained by inspecting these differences in terms of statistical significance. For Dutch, the difference between *aangezien* and *omdat* was significant, which indicates that *aangezien* may encode a higher SI level than *omdat*. But this approach does not help us in accounting for the differences between *aangezien* and *want*.

We could, of course, make claims based on our own intuitions regarding the meanings of the connectives or based on intuitions expressed in the linguistic literature. However, we prefer a somewhat more cautious approach, which combines distributional data and semantic intuitions.

In order to determine the proper level of SI that a connective contributes to its environment, we propose to analyze the contexts they occur in according to the four SI aspects presented in section 2. That is, (i) the degree of subjective involvement of a conscious participant, (ii) the degree of iconicity of the causal relation, (iii) the distance to speaker and speaking time, (iv) the degree of explicitness of the participants involved in the causal relation. For each occurrence of a given connective, we analyze the related segments for a number of features related to these four aspects. The combination of these distributional data determines the level of SI encoded by the connective. Doing so, we can scale the connectives with respect to each other on the SI scale. From a contrastive point of view, it becomes also possible to compare the respective SI level of “equivalent” connectives in different languages. We will now first proceed with the analysis of the Dutch connectives, then with the French connectives, and finally compare the results in the two languages.

So far, we have established, on the basis of Table 1, that in Dutch the backward connectives *want* and *aangezien* are higher on the SI scale than *omdat*. There is indeed a significant tendency for *omdat* to express causal relations that are isomorphic with real-world causality, i.e. non-volitional, volitional, and causality-based epistemic relations, and which are at a certain distance from speaker and speaking time (non-volitional and volitional relations). These two characteristics place *omdat* lower on

the SI scale than *want* and *aangezien*. The latter connectives can indeed also express relations that are non-iconic with respect to real-world causality (non-causality based epistemic relations), with *want* occurring also in speech-act contexts, i.e. with minimal distance to speaker and speaking time.

Nevertheless, *want* and *aangezien* differ too. According to our intuitions *aangezien* cannot express speech-act relations (see example (16)). Moreover, its acceptability is also doubtful in abductive epistemic relations, especially when the premise precedes the conclusion (see 17))

- (16) a. ?Kom onmiddellijk naar binnen, *aangezien* het regent.
 ?Come inside immediately, [*aangezien*] it is raining.
 b. ?Wat doe je vanavond, *aangezien* er een leuke film draait.
 ?What are you doing tonight, [*aangezien*] there is a good movie on.
- (17) a. ?Hij rookt, *aangezien* hij sigaretten koopt.
 ?He smokes, [*aangezien*] he is buying cigarettes.
 b. **Aangezien* hij sigaretten koopt, rookt hij.
 **[Aangezien]* he is buying cigarettes, he smokes.

Hence, we hypothesize that *want* is higher on the SI scale than *aangezien*. This should be confirmed or disconfirmed on the basis of further distributional data. We will further investigate the difference in SI between *want* en *aangezien* by focussing on two SI dimensions that have been mentioned already in the introduction: the distance between the relation and the present speaker and the realization of the protagonist of the relation. To uncover these SI aspects, we proceed in two steps.

First we determine the identity of the causal protagonist in the first segment (S1). To this end, we first select the fragments which actually have a causal protagonist, thus leaving out the factual segments, and which show some variation with respect to the identity of this protagonist, thus excluding speech-act fragments for which the identity of the causal protagonist is always the speaker. This leaves us with volitional and epistemic relations for which the possible participants include the author, a group including the author and others, a cited speaker, a generic third person, a pronominal third person and a nominal third person. The principal distinction in this domain is the one between first-person participants and third-person participants (we don't find

second person protagonists in our newspaper corpora). The hypothesis is that a connective with a higher SI level more often occurs with first-person participants, since the actions and conclusions of speakers will generally be formulated with a higher SI level than the actions and conclusions of third persons. After all, the speaker is more likely to accept the general assumptions underlying his own decisions than those underlying other persons' decisions.

The second step is to determine how this responsible participant is realized in the consequence segment (S1), i.e. whether he is linguistically expressed or not, and if this is the case whether this protagonist is explicitly (*I think/ he thinks*) or implicitly (e.g. *probably*) present in the segment. In non-volitional relations there is no conscious causal participant, so the question does not arise. In volitional relations, this participant is nearly always explicitly mentioned (with the exception of some passive and impersonal constructions), so the question arises mainly in epistemic relations. For these relations, we first compare the number of factually presented consequence segments (no participant involved) with the number of subjective consequence segments. Within this latter category we distinguish the implicit from the explicit cases.

Of course, this latter distinction is not entirely independent of the distinction between first and third person protagonists, since generally first persons are implicit and third persons explicit. However, since first persons are occasionally explicit and third persons occasionally implicit, it makes sense to formulate a separate hypothesis regarding implicit and explicit realization of the participant responsible for the causal relation. Since we have defined SI as the degree of implicit involvement of the present speaker in the relation, our hypothesis must be that a more subjective connective more often occurs in first segments with implicit participants. When no participant is involved in the segment (factual conclusions in epistemic relations), the SI hypothesis predicts a higher occurrence of high-level SI connectives, since these connectives are capable of introducing subjectivity into the relational interpretation all by themselves.

The results regarding these three hypotheses on the difference between the Dutch forward connectives are presented in Figures 2 and 3, and in Table 4. Since we are most interested here in the contrast between the connectives *want* and *aangezien*, results for these connectives are given separately in addition to the global results.

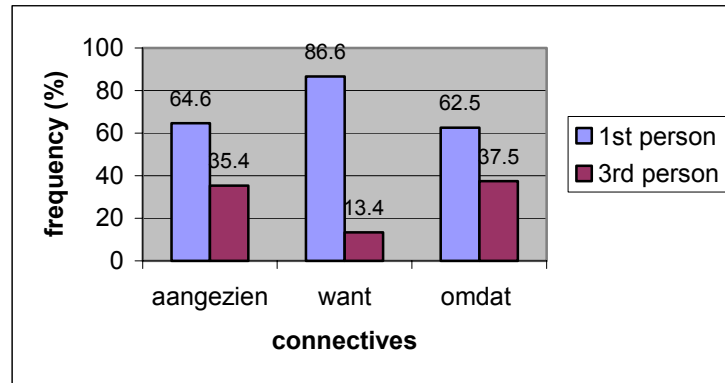


Figure 2: *Identity of the causal participant co-occurring with aangezien, want or omdat in volitional and epistemic relations*
 $\chi^2 = 7.785$, $df = 2$, $p < .05$ ($p = .020$)
want-aangezien: $\chi^2 = 6.085$, $df = 2$, $p < .05$ ($p = .014$)

Participant	connective		
	<i>aangezien</i>	<i>want</i>	<i>omdat</i>
Subjective consequence	40 (100%)	37 (97.4%)	24 (100%)
Factual consequence	0	1 (2.6%)	0
	40	38	24

Table 4: *Type of consequence segment co-occurring with aangezien, want or omdat in epistemic relations*

$\chi^2 = 1.701$, $df = 2$, $p > .05$ ($p = .427$) N.S.
want-aangezien: $\chi^2 = 1.066$, $df = 1$, $p > .05$ ($p = .302$) N.S.

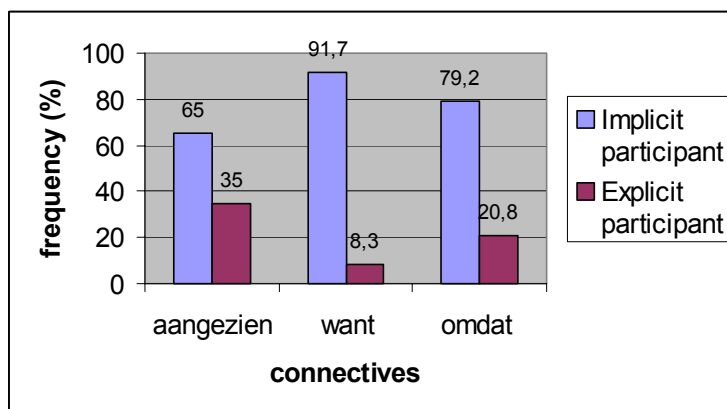


Figure 3: *Explicit or implicit presence of the causal participant co-occurring with aangezien, want, or omdat in epistemic relations (with participant)*

$$\chi^2 = 6.462, df = 2, p < 0.05 (p = .040)$$

$$want-aangezien: \chi^2 = 6.279, df = 1, p < .05 (p = .012)$$

The hypothesis that *want* encodes a higher SI level than *aangezien* is supported by two SI features: *want* more often has **first-person participants** (Figure 2) and these are more often left **implicit** in the first segment (Figure 3). If we consider these data together with those in Table 2, we believe we may conclude that the three Dutch causal connectives under investigation can be put on the SI scale in increasing order starting with *omdat*, followed by *aangezien* and *want* being highest. This is a striking result since the preliminary results for the French connectives placed *puisque*, supposedly the counterpart of *aangezien*, highest on the SI scale. So let us have a look at Figures 4 to 6 which display the SI features for the French connectives (again leaving out those cases which present no variation with respect to the causal participant involved, if at all).

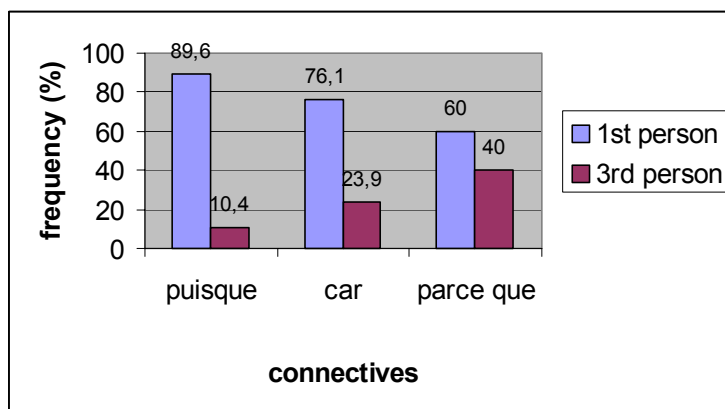


Figure 4: Identity of the causal participant co-occurring with *puisque*, *car*, or *parce que* in volitional and epistemic relations

$$\chi^2 = 14.542, df = 2, p < .001$$

puisque-car: $\chi^2 = 3.029, df = 1$, which is significant at $p = .041$ in a one-tailed test

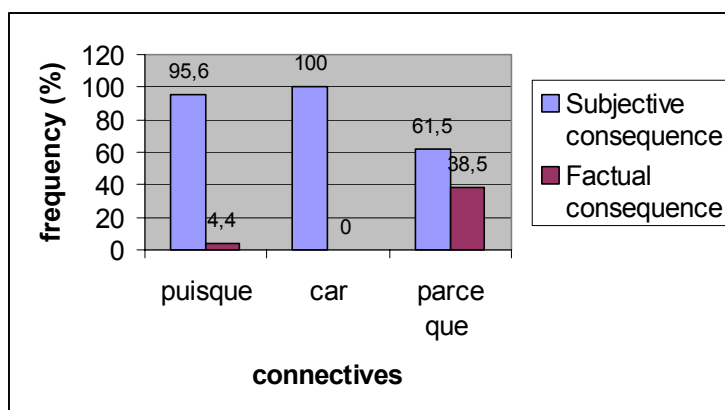


Figure 5: Type of conclusion co-occurring with *puisque*, *car*, or *parce que* in epistemic relations

$$\chi^2 = 2.258, df = 2, p > .05 (p = .323) \text{ N.S.}$$

puisque-car: $\chi^2 = 1.57, df = 1, p > .05 (p = .21) \text{ N.S.}$

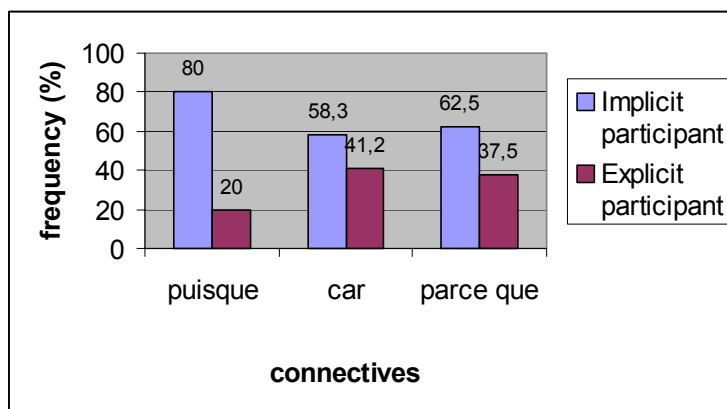


Figure 6: *Explicit or implicit presence of the causal participant co-occurring with puisque, car, or parce que in epistemic relations (if participant present)*

$$\chi^2 = 4.782, df = 2, p > .05 (p = .92) \text{ N.S}$$

$$\text{puisque-car: } \chi^2 = 4.5503, df = 1, p < .05 (p = .034)$$

Figures 4 and 6 confirm that in French *puisque* is higher on the SI scale than *car* and *parce que*, since *puisque* more often comes with implicit first person participants. Hence, in French the respective positions of the three connectives is different from that in Dutch.

In order to check whether these differences in distribution on the scale are the result of specific divergences in use of some of the connectives, or rather result from more general differences, we propose to make an inter-language comparison of the counterparts for each connective under investigation. The comparison will bear on the four SI aspects already accounted for. The results are given in Table 5 and Figures 7-8.

Relation	connective					
	<i>puisque</i>	<i>aangezien</i>	<i>car</i>	<i>want</i>	<i>parce que</i>	<i>omdat</i>
non-vol. causal		2	1	1	11	10
vol. causal	1	8	10	7	23	16
causal epis- temic	36	26	27	20	14	20
non-causal epistemic abductive	10	14	9	17	2	4
epistemic speech act	1			1		
	2		3	4		
	50	50	50	50	50	50

Table 5: *Inter-language comparison: relational interpretations*

puisque-aangezien : $\chi^2 = 12.72$, $df = 5$, $p < 0.05$

car-want : $\chi^2 = 5.18$, $df = 5$, $p = 0.39$ (N.S.)

parce que-omdat : $\chi^2 = 3.03$, $df = 3$, $p = 0.39$ (N.S.)

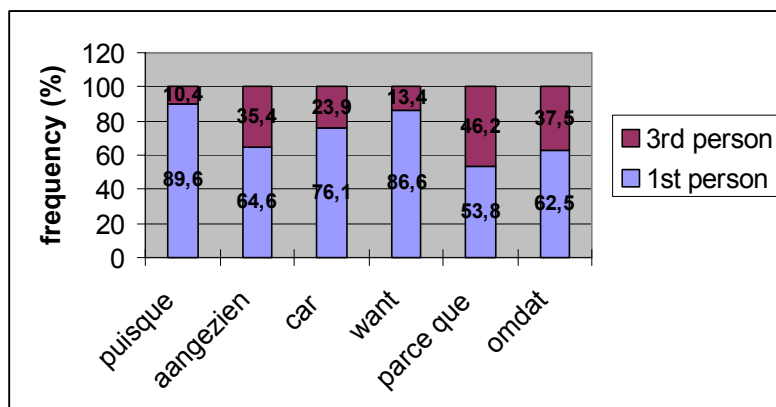


Figure 7: *Inter-language comparison: identity of the relational protagonist*

puisque-aangezien: $\chi^2 = 8.491$, $df = 1$, $p < 0.01$

car-want: $\chi^2 = 1.676$, $df = 1$, $p = 0.195$ (N.S.)

parce que-omdat: $\chi^2 = 0.608$, $df = 1$, $p = 0.436$ (N.S.)

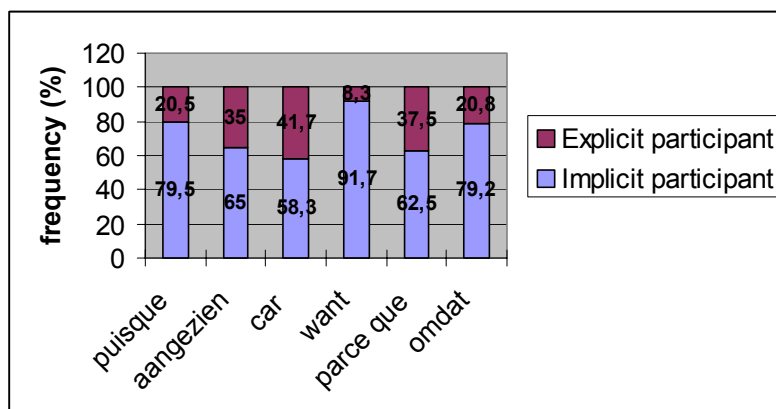


Figure 8: *Inter-language comparison: realisation of protagonist*
puisque-aangezien: $\chi^2 = 2.414$, $df = 1$, $p = 0.12$ (N.S.)
car-want: $\chi^2 = 9.023$, $df = 1$, $p < 0.01$ ($p = 0.003$)
parce que-omdat: $\chi^2 = 1.338$, $df = 1$, $p = 0.247$ (N.S.)

Table 5 and Figures 7-8 show that French *puisque* partially differs from Dutch *aangezien*. There are indeed significant differences between the two connectives in terms of relational interpretations (Table 5) and in terms of the identity of the causal participant (Figure 7). This leads us to the conclusion that *puisque* takes a higher position on the SI scale than *aangezien*. Furthermore, French *car* appears to be very close to Dutch *want*, the latter connective having, however, a higher SI since it occurs more often with implicit participants (Figure 8). Our data do not show any significant differences between the two low SI connectives *parce que* and *omdat*.

3.2. Speaker Involvement and connectives in use

In the previous section we have ordered a number of backward causal connectives on the speaker involvement scale. Now, what does it mean for a causal connective to be higher or lower on the Speaker Involvement scale? According to us, every connective brings some of its semantic content to the causal relation it is building. In addition to making explicit a causal relation between two segments, a connective also encodes a certain SI level which it contributes to the interpretation of its discourse environment. It is our claim that the proper SI level of the connective, i.e. its position on the SI scale, and the SI level of the given stretch of

discourse, i.e. the underlying causal relation, should concord. In other words, when the SI level of the connective is too high or too low to be combined with the level allowed for by the discourse environment, the use of the connective is inappropriate.

In the next three examples the differences in SI manifest themselves in the degree to which the speaker is presenting his stealing as justified. With *want* the speaker conveys the assumption that stealing from your boss is a perfectly natural thing to do for somebody in need of extra money for his family. With *omdat*, the speaker does not justify his behavior, but merely explains it: he presents a reason without tempting to generalize the validity of the relation between reason and action. The assumptions conveyed by *aangezien* are more like those conveyed by *want* than those conveyed by *omdat*. (Though *aangezien* may be somewhat weaker than *want*. Intuitions get rather subtle at this point).

- (18) a. Ik stal van mijn baas, *omdat* ik extra geld nodig had om mijn grote gezin te onderhouden.
 ‘I stole from my boss, [*omdat*] I needed extra money to support my large family.’
- b. Ik stal van mijn baas, *aangezien* ik extra geld nodig had om mijn grote gezin te onderhouden.
 ‘I stole from my boss, [*aangezien*] I needed extra money to support my large family.’
- c. Ik stal van mijn baas, *want* ik had extra geld nodig om mijn grote gezin te onderhouden.
 ‘I stole from my boss, [*want*] I needed extra money to support my large family.’

Another way of demonstrating differences in SI profiles is substituting the connectives for each other in contexts with a low degree of speaker involvement. For instance, when the speaker clearly distances himself from another protagonist by criticizing his actions, a low-SI connective should be more appropriate than a high-SI connective. Compare the next two examples, taken from the Dutch *omdat* corpus, in which *aangezien* appears inappropriate.

- (19) Afwezig waren vertegenwoordigers van de Nederlandse vakbeweging. “Dat is raar, maar ze willen niet *omdat*/**aangezien* het Europese werk voor hen een ver-van-mijn-bed-show is en ze ons een beetje zien als concurrenten”, aldus X.
 ‘The representatives of the Dutch trade unions were not there. “That is strange, but they don’t want to come [*omdat*/**aangezien*] they consider the European work as less important than their domestic affairs and they consider us as competitors”, says X.’
- (20) Er wordt met name in de sociale wetenschappen maar al te vaak verkrampt geschreven, *omdat*/**aangezien* men zich van tevoren tegen kritiek tracht in te dekken.
 ‘Especially in the social sciences we often encounter over-cautious writing, [*omdat*/**aangezien*] one tries to protect oneself from criticism in advance.’

Finally, a word should be said about the cross-linguistic comparison of the backward causal connectives. The data in section 3.1 clearly show that *puisque* and *aangezien* partially diverge from each other in that they do not naturally occur in the same types of discourse segments. In fact, our data with respect to the relational interpretations and the identity of the primary participant in the causal relation show that *puisque* has a higher SI than *aangezien*. As a result, *aangezien* is not used in the same way as *puisque* is. This might explain their diverging frequencies. In effect, the role of *puisque* seems to be largely taken over by *want* in Dutch, whose “official counterpart” *car* seems to be a little lower on the SI scale, *car* co-occurring less with implicit participants than *want*. At first view, *parce que* and *omdat* do occupy similar positions on the SI scale as appears from their occurrence in similar discourse environments.

A last question remains about the functional use of the connective *aangezien*. It is indeed fairly infrequent with respect to the other backward causal connectives, although not specifically bound to a restricted text genre (Degand 2000), and very close in meaning to *want*, although with a lower SI. A possible explanation could be syntactic in nature. Contrary to *want*, *aangezien* is a subordinating conjunction that can be used in preposed constructions like (21). Actually, these types of constructions are highly frequent in Dutch since nearly 50% of the

aangezien-segments occur in this ordering, as was shown in Table 1 above.

- (21) Aangezien we dit jaar niet op vakantie kunnen, zullen we maar veel eendagstripjes maken.
 '[Aangezien] we can't go on holiday this year, we will probably make a lot of one-day trips.'

A possible explanation for the high frequency of *aangezien* in preposed position could be that it is used instead of *want* when this latter connective is excluded for syntactic reasons, or at least reasons of information flow. Recent research (Degand 2000) has shown that the use of preposed vs. intermediate positioning in causal sequences is strongly linked to the preceding information. In particular, the first causal segment tends to pick up information contained by the previous theme or rheme. Since the connective *want* does not allow such syntactic manipulations and the connective *omdat* is too low on the SI scale; Dutch seems to have favored the use of preposed *aangezien* in such cases.

In French this phenomenon does not take place. The three connectives *parce que*, *car* and *puisque* differ significantly from one another. At the same time, substitution hardly ever leads to unacceptable sentences, but rather to semantic differences which we have explained in terms of SI.

4. Conclusion

In this paper, we have investigated whether an alternative account to the categorization of connectives, the Speaker Involvement scale, could be extended to backward causal connectives in Dutch and in French. To this end, we performed a corpus analysis of a number of backward causal connectives in terms of SI. In concrete terms this means first investigating the discursive context in which the connectives naturally occur and re-analyzing them in terms of a scale of increasing speaker involvement in the construction of the causal relation: from non-volitional to speech act relations via, in increasing SI-order volitional, causal epistemic and non-causal epistemic relations. The SI profiles thus developed were further supported by observations regarding the semantic effects of substituting a connective by one with a different SI level. According to

us, such a semantic analysis leads to fine-grained distinctions uncovering subtle meaning differences within a language but also cross-linguistically. In particular, it appeared that two supposedly translation equivalents *puisque* and *aangezien* do in fact not appear in the same contexts of use. According to our analyses, *puisque* encodes a higher level of SI than *aangezien* which should explain these divergences in use, as well as the high frequency difference of the two connectives in the two languages. Further research on bilingual and translated corpora is needed to show whether it is indeed the case that *puisque* is most often **not** translated by its “official” equivalent *aangezien*. The question remains open on which Dutch connectives fulfil the role of *puisque* in naturally occurring texts. According to our first observations, *want* should be frequent in these contexts but there might be other constructions playing a similar role.

An important outcome of this research is that connectives cannot be considered in isolation from one another. We believe to have shown that causal connectives, both in Dutch and in French, should be considered on a continuum from lowest to highest speaker involvement. The variation in the use of these connectives is due to the relative independence between connective meanings and categories of relational interpretations. This yields expressive possibilities for speakers who want to introduce assumptions in, or remove them from the interpretation of a certain relation. Such findings are of course important both for translation and for foreign language teaching. Finally, since the results we obtained are similar to those for forward causal connectives (Pander Maat & Degand 2001), we can conclude that the SI scale has proven to be valid for backward as well as forward causal connectives in Dutch and in French.

Acknowledgements

Liesbeth Degand is a research fellow at the Belgian Fund for Scientific Research (F.N.R.S.). At the time of this research, she was also supported by an I.U.A.P project from the Belgian Federal Government (I.U.A.P 4/19).

References

- Belle, W. van (1989). Want, omdat, en aangezien: een argumentatieve analyse. *Leuvense Bijdragen* 78, 435-556.
- Degand, L. (1998). Het ideationele gebruik van *want* en *omdat*: een geval van vrije variatie? *Nederlandse Taalkunde* 3, 309-326.
- Degand, L. (2000). Contextual constraints on causal sequencing in informational texts. *Functions of Language* 7, 173-201.
- Degand, L. (2001). *Form and Function of Causation. A theoretical and empirical investigation of causal constructions in Dutch*. [Studies op het gebied van de Nederlandse Taalkunde]. Peeters, Leuven.
- Degand, L. & H. Pander Maat (1999). Scaling causal relations in terms of Speaker Involvement. *Levels of Representation in Discourse, Working Notes of the International Workshop on Text Representation*, Edinburgh University, 45-54.
- Degand, L. & T. Sanders (1999). Causal connectives in language use. Theoretical and methodological aspects of the classification of coherence relations and connectives. *Levels of Representation in Discourse, Working Notes of the International Workshop on Text Representation*, Edinburgh University, 3-12.
- Dijk, T.A. van (1979). Pragmatic connectives, *Journal of Pragmatics* 3, 447-456.
- Groupe λ -I (1975). Car, parce que, puisque. *Revue Romane* 10, 248-280.
- Halliday, M.A.K & R. Hasan (1976). *Cohesion in English*. Longman, London.
- Iordanskaja, L. (1993). Pour une description lexicographique des conjonctions du français contemporain. *Le Français moderne* 2, 159-190.
- Mann, W.C. & S.A. Thompson (1986). Relational propositions in discourse. *Discourse Processes* 9, 57-90.
- Martin, J.R. (1992). *English Texts: system and structure*. John Benjamins, Amsterdam.
- Moeschler, J. (1989). Pragmatic connectives, argumentative coherence and relevance. *Argumentation* 3, 321-339.
- Pander Maat, H. (1998). Classifying negative coherence relations on the basis of linguistic evidence. *Journal of Pragmatics* 30, 177-204.
- Pander Maat, H. & L. Degand (2001). Scaling causal relations and connectives in terms of Speaker Involvement. *Cognitive Linguistics* 12, 211-245.
- Pander Maat, H. & T. Sanders (1995). Nederlandse causale connectieven en het onderscheid tussen inhoudelijke en epistemische coherentie-relaties. *Leuvense Bijdragen* 84, 349-374.
- Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge University Press, Cambridge.

