

# Historical and comparative perspectives on subjectification: A corpus-based analysis of Dutch and French causal connectives\*

JACQUELINE EVERS-VERMEUL, LIESBETH DEGAND,  
BENJAMIN FAGARD, AND LIESBETH MORTIER

## *Abstract*

*In this article, we focus on the diachronic development of causal connectives and investigate whether subjectification occurs. We present the results of ongoing and previous corpus-based analyses of the diachronic development of Dutch want and omdat, and French car and parce que, all four causal connectives roughly meaning ‘because’. In addition, we try to show that “grammaticalization studies can gain from the systematic and principled use of large computerized corpora and the methods which have been developed within corpus linguistics” (Lindquist and Mair 2004: x). That’s why we have combined two historical and two comparative corpus methods to chart the diachronic development of these four causals. Our study reveals that subjectification is not an integral part of the diachronic development of these causals: subjectification does occur in the rise of these connectives, but in the later stages of their development only parce que undergoes subjectification. Our analyses show that the four methods all have their own merits and limitations, but they are most effective when combined.*

## **1. Introduction**

Research on grammaticalization has established itself as a major area in linguistic studies. A noteworthy development in the past decade is the growing interest of grammaticalization theorists in the use of corpora, and hence in the techniques developed in the field of corpus linguistics (cf. Rissanen et al. 1997; Lindquist and Mair 2004). Mair (2004) points out several commonalities between the two fields, and argues in favor of a closer collaboration between corpus linguists and grammaticalization specialists. In our article, we follow Mair’s recommendation and confront current ideas within grammaticalization theory with the results of various corpus studies. More specifically, we will show the advantages of taking a corpus-based approach to the study of causal

connectives, a subclass of discourse markers. “Discourse markers are ideal for observing variation and change: they originate in different grammatical categories, they often compete with many other forms, and they are sensitive to trends regarding language use” (cf. Vincent 2005: 191).

The diachronic development of discourse markers often involves a process of ‘(inter)subjectification’ (cf. Traugott and Dasher 2002; Athanasiadou et al. 2006), a shift from meanings pertaining to the characterization of the objective world first to meanings involving the expression of personal attitudes of the speaker (subjectification) and then to meanings linked to speaker-hearer interactions (i.e., intersubjectification). A famous example in the area of connectives concerns the diachronic development of English *while* (cf. Traugott 1995; González-Cruz 2007). A first instance of its subjectification can be found in the change of the adverbial phrase *pa hwile þe* (‘at the time that’) into the temporal connective *while*. Instead of profiling a specific time in the real world, the connective *while* profiles the ordering of events within the discourse structure, an ordering provided by the speaker. A second example of subjectification in the development of *while* can be found when temporal *while* turns into concessive *while*. This new use construes a relation between events that has no reference in the described situation, but only in the speaker’s belief about coherence.

Traugott (1995: 39) puts forward a specific claim regarding subjectification in the area of connectives. According to her, “historically almost all grammatical markers of clause combining have developed out of a more ‘objective’ function” (see also Dasher 1995). And indeed, many temporal, causal and conditional connectives have grown out of adverbial constructions (Genetti 1991). Examples in various languages include the subjectification of German *weil* ‘because’ (Keller 1995; Günthner 1996), Japanese *na* elements (Onodera 2000, 2004), English *because* and its Japanese counterpart *kara* (Higashizumi 2006), and Dutch *dus* ‘so’ (cf. Evers-Vermeul 2005, Evers-Vermeul and Stukker 2003), to mention just a few of many connectives from a variety of languages.

With the growing availability of digital diachronic corpora, the number of diachronic analyses of connectives in terms of subjectification has shown a tremendous increase, thus providing evidence in favor of Traugott’s (1995) claim. Note that this claim is not formulated as a strong or absolute hypothesis; Traugott does not predict that any change in the meaning or use of a connective must involve subjectification, but only that — in the majority of cases — subjectification will be involved at some stage (a weak or relative hypothesis). However, it would still be interesting to find out whether Traugott’s claim can be falsified or restricted in any way. More specifically, it seems worthwhile to test this weak hypothesis against various stages of the diachronic development of connectives. For example, it could be the case that subjectification typically occurs during the rise of connectives, a stage with both changes in meaning and changes at the grammatical level, such as internal bonding or reanalysis

(changes typically associated with grammaticalization). In fact, this is what Traugott (2010: 40–41) suggests herself, when she claims that subjectification “is more likely to occur in primary grammaticalization (the shift from lexical/constructional to grammatical) than in secondary grammaticalization (the development of already grammatical material into more grammatical material).” In order to test the weak subjectification hypothesis, we will focus on later stages of the diachronic development of connectives. More specifically, we will investigate whether shifts occur in the distribution over the meanings a connective can already express, but that are not necessarily accompanied by grammatical changes. Hence, our first research question is: Does subjectification occur in later stages of the diachronic development of causal connectives?

A second research question results from a few critical remarks that can be made on the methodology of various studies in this area. A first point concerns studies in which claims about subjectification or grammaticalization are based solely on synchronic data. For instance, Günthner and Mutz (2004) analyze the variation of *obwohl* ‘although’ and *wobei* ‘whereby’ in contemporary spoken German. They conclude on the basis of these synchronic data that *obwohl* and *wobei* “have developed discourse-pragmatic functions and have become, or are on their way to becoming discourse/pragmatic markers” (Günthner and Mutz 2004: 98). Vincent (2005) is another case in point (cf. also König and Van der Auwera 1988; Erman and Kotsinas 1993). Although she supports some of her claims about the diachrony of *par exemple* by reference to ancient dictionaries, she predominantly bases her claims on synchronic data.<sup>1</sup> In our view, subjectification studies as well as other studies with diachronic implications may take synchronic data as a starting point to build diachronic hypotheses, but can be validated only through its testing against diachronic data. For instance, synchronic variation between speech and writing (see Section 6 below) can serve as an indication that a specific type of language evolution is developing and thus lead to a given hypothesis; but then, this hypothesis should be tested against diachronic data in order to be validated.

A second methodological point is that studies which do meet this diachronic criterion and are based on authentic diachronic data have been predominantly qualitative in nature. Qualitative discourse studies typically take a small data set, a single text or a relatively small sample of texts, and examine it in depth. The majority of this type of research only provides anecdotic examples, and lacks quantitative underpinning (but see e.g., Prévost 1999, 2003, 2007, who systematically analyzes quantified data in the area of discourse markers). Although detailed qualitative analyses based on manual extraction are useful in themselves (cf. Traugott 1995 on *while*; Molencki 2007 on *since*), they can and should be fruitfully complemented by corpus-based methods. Stefanowitsch (2006: 12) formulates this urge for quantification in the area of research into metaphorical mappings: “many of the results are provisional, awaiting more

stringent quantification and statistical evaluation.” As Partington (2006: 268) puts it: “Complementing the qualitative with a more quantitative approach, as embodied in Corpus Linguistics, not only allows a greater distance to be preserved between observer and data but also enables a far greater amount of data to be contemplated. In addition, it can identify promising areas for qualitative forms of analysis to investigate.” Moreover, corpora enable researchers to meet the criterion of total ‘accountability’ (cf. Johansson 1985: 208; Labov 1994: 550), which demands that linguistic descriptions account for all the data in a body of texts, and not just for particular instances.

Given this characterization of previous studies on subjectification in the use of connectives, and given the new perspective corpus-based approaches seem to offer, the following two research questions can be formulated: 1) Does subjectification occur in later stages of the diachronic development of causal connectives? and 2) What does a corpus-based approach add to the study of this diachronic development? In the remainder of this article we will combine two historical and two comparative corpus methods to chart the diachronic development of four causal connectives. In Section 2 we give the rationale behind this method of using converging evidence. In Sections 3 to 6 we present the results of our four corpus studies. Not all four of our studies give rise to in-depth analyses here, our focus being on the type of results each of the four studies brings about, and on the methodological need to combine them. In Section 7, we will answer our research questions and put forward some points for discussion.

## 2. Method

We focus on the diachronic development of Dutch and French causal connectives and investigate whether subjectification occurs during these developments. To this end, we present the results of ongoing and previous analyses of the diachronic development of Dutch *want* and *omdat*, and French *car* and *parce que*, all four causal connectives roughly meaning ‘because’.

The initial premise of this article is that the techniques of corpus linguistics can assist the diachronic study of connectives. They can help reveal recurrent patterns of connective usage which reflect the systematic behavior and attitudes of the users. The choice of the corpus is fundamental; as Hoffmann (2004) points out, the

... reliability and meaningfulness of empirical data is heavily dependent on the assumption that language corpora constitute suitable mirrors of actual language use, either in its totality or at least in a wider functional domain. The choices made by the compilers of a corpus with respect to the selection and proportional representation of

Table 1. *Four corpora used in our subjectification research*

	Same texts	Different texts
Same period	Corpus 1 Present-day translation corpora	Corpus 4 Present-day spoken and written language
Different periods	Corpus 2 Bible translations from different periods	Corpus 3 Various written texts from different periods

different text domains consequently have direct influence on the relevance of the linguistic results. (Hoffmann 2004: 197)

This is why we will analyze the degree of subjectivity of causal connectives in four different corpora (see Table 1). Confronting the results of these different corpora applied to the same set of connectives should help us determine the contribution of each corpus when trying to trace subjectification processes.

Corpus 1 is a so-called parallel corpus, or more specifically, a translation corpus.<sup>2</sup> It is a collection of present-day original texts and their translations in another language. In our case, Corpus 1 is compiled of original Dutch texts and their French translations, and of original French texts and their translations into Dutch (see Section 3 for more details). Corpus 2 comprises translated texts from different periods; in our study we selected Dutch Bible translations from four periods.<sup>3</sup> Corpus 3 consists of a variety of Dutch and French original texts, both from present-day and from ancient written sources. Ideally, this is a comparable corpus, in the sense that it contains texts matched by such criteria as domain, genre, intended audience, etc. (cf. Johansson 1998: 5). This is the type of corpus that is most commonly used in diachronic research. Corpus 4 is compiled of present-day Dutch and French data; per language, it contains one sub-corpus of written data and another of spoken data, thus enabling a comparison of the two modalities. In Sections 3 to 6, we will discuss for each corpus a) its potential advantages and disadvantages, and b) the results it leads to in terms of the subjectification of causal connectives.

In order to measure the degree of subjectivity of each connective fragment in a reliable and quantifiable way, we took Sweetser's (1990) domains of use as an analytical instrument (cf. Pander Maat and Degand 2001; Pander Maat and Sanders 2001; Evers-Vermeul and Stukker 2003; Pit 2003). We therefore distinguish between (nonvolitional and volitional) *content*, *epistemic*, and *speech act* relations, as illustrated by the constructed examples in (1)–(4) below.

- (1) Nonvolitional content  
*The temperature rose quickly because the sun was shining.*
- (2) Volitional content  
*We went out in the garden because the sun was shining.*

- (3) Epistemic  
*The temperature is probably going to rise, because the sun is shining.*
- (4) Speech act  
*Let's have dinner in the garden, because the sun is shining.*

If subjectivity “implies some degree of integration of the perceiver in the description of an object or a process” (Cuenca 1997: 5), then it can be argued that the different domains can be used as a way to measure the degree of subjectivity of causal (and other) connective fragments. In a content relation, the speaker provides a description of facts that can be established objectively in reality. Content relations like the ones in (1) and (2) describe relations that can be objectively observed in the real world. Characteristic of nonvolitional relations is that the causal relation occurs without human intervention. Hence, this relation type is more objective than the volitional kind of content relations, in which human activity is involved. More subjective are epistemic relations (see (3)) in which the consequence is not a state of affairs in reality, but a mental state of the protagonist. The causal relation as a whole — which often involves argumentation — is not objectively observable and the speaker has to adopt the perspective of the protagonist in order to present the causal relation. Maximally subjective are speech-act relations like (4), since they do not concern a reality outside the speech event, but the structure of the ongoing discourse (cf. Pander Maat and Degand 2001: 216–228). The relation type of each connective fragment can be established using a paraphrase test (cf. Evers-Vermeul [2005] for a detailed discussion of this test, and Spooren and Degand [2010], for the operationalization of such categorizations).

### 3. Results Corpus 1: analysis of present-day translation corpora

Our aim in using Corpus 1 for diachronic research is twofold.<sup>4</sup> Our first purpose is not to track actual diachronic changes, but to gain insight into the precise meaning of the linguistic items under study. Translation corpora contain texts that are intended to express the same meanings and have identical or at least very similar discourse functions in the relevant languages. Successively using the source and target language as a starting point, we can establish paradigms of correspondences: the translations can be arranged as a paradigm where each target item corresponds to a different meaning of the source item. The use of translation corpora for this purpose is relatively new.<sup>5</sup> Traditionally, linguists have asked native informants to make judgments about meanings. Native speakers can distinguish different uses of the same polysemous item and say how they are related. However, at times too many uses may be distinguished, or too few (cf. Aijmer 2004: 58; Bybee et al. 1994: 44). Translations

are more reliable as sources of meanings and uses than native informants, because they are produced by trained translators without any theoretical concern in mind. Dyvik (1998, 2004) was one of the first to argue in favor of the use of translation corpora to establish the precise semantics of words (cf. also Doherty 1998). According to him, “the activity of translation is one of the very few cases where speakers evaluate meaning relations between expressions without doing so as part of some kind of metalinguistic, philosophical or theoretical reflection, but as a normal kind of linguistic activity. This inspires confidence in the intersubjectivity of such evaluations” (Dyvik 1998: 51). As such, the advantages of a translator-based approach to semantics and pragmatics are clear: “by taking the translator’s profile as a starting point, one is likely to acquire some information on the original propositional content of the message and on the potentially accompanying pragmatic implicatures” (Mortier 2007: 144). This type of analysis can thus also be used to place linguistic alternatives relative to each other on a subjectivity scale, specifying the semantic profile of closely related connectives.

Our second purpose in using translation corpora for diachronic research is that they reveal alternative markers expressing similar meanings in a specific genre. By performing back and forth translations, resulting in what we will call a *mirror analysis*, it becomes possible not only to track the most important synchronic translation equivalents, but also to reveal a field of competing markers for comparable meanings in one language (cf. Dyvik 2004; Lewis 2005; Mortier 2007; Mortier and Degand 2009). This is useful for subsequent diachronic analyses of these linguistic competitors. Although we acknowledge that diachronically there is no “need to see a new or alternative marker as contingent on the loss or dysfunction of another marker (cf. Bybee et al. 1994: 21)” (Aijmer 2004: 70), we do believe that changes in the system as a whole may have repercussions on the use of specific linguistic items. For example, it might be the case that certain causal connectives take over the function of their competitors or that the competitors take over one or more functions of the connective under investigation.

There are also some disadvantages associated with translation corpora. First of all, translations only provide synchronic insights and do not reveal diachronic changes. Second, although there is a growing body of translation corpora, especially for translation from and to the English language, the availability of translation corpora is still restricted. Also, as far as genre is concerned, the range of translated texts is restricted as compared with the range of original texts (cf. Johansson 1998: 4). This may have repercussions on the generalizability of the conclusions drawn from these data. A third disadvantage is that the data may be infected by *translationese* (cf. Gellerstam 1996), i.e., translation-based deviations from target language conventions. Translated texts may differ from original texts because of source language influence. Finally, it



Table 2. Frequencies of Dutch translations of *car* and *parce que*

Dutch translation	Translations of <i>car</i> (N = 45)	Translations of <i>parce que</i> (N = 53)
Want	25 (55.6%)	5 (9.4%)
Omdat	12 (26.7%)	41 (77.4%)
Doordat 'because of the fact that'	–	1 (1.9%)
Aangezien 'since'	3 (6.7%)	–
No translation / punctuation	4 (8.9%)	5 (9.4%)
Reformulation	1 (2.2%)	1 (1.9%)

is well-known that linguistic choices often depend on the individual translator's particular style and skill, and that there may be outright mistakes in translations. It is conceivable that "somewhere along the interpretation process, a mismatch occurs between the speaker's intentions and the hearer's interpretation" (Mortier 2007: 144). This, however, should not prevent us from using translations as linguistic evidence, because "it is exactly the translators' performance, not so much as good translators but as language users, which is of interest" and because "consistency in syntactic and lexical discrepancies between source and target texts is precisely what this kind of evidence hinges on" (Noël 2003: 779–780).

Our present-day translation corpus consists of a corpus of original Dutch texts and their translations into French, and of original French texts and their translations into Dutch. The size of Corpus 1 is approximately 550,000 words. Two main types of text are represented: fiction (literature) and nonfiction (newspaper texts). We selected all occurrences of the causal connectives *want*, *omdat*, *car* and *parce que* and their translations. Table 2 and Table 3 present the resulting lists of markers that were used as translations, as well as their respective frequencies.

As Table 2 shows, *want* and *omdat* are the most common translations of *car* and *parce que* (cf. examples (5) and (6)). Other linguistic alternatives (such as *doordat* 'because of the fact that' and *aangezien* 'since') are very infrequent, and are outnumbered by fragments in which the connective is replaced by punctuation (as in (7)) or is deleted altogether.

- (5) F: *Je lui ai dit que j'avais adopté sans peine la tactique du sourire car je suis convaincue qu'en effet cette histoire ne compte pas tant pour Maurice.*  
(Fiction-FR)
- D: *Ik heb haar verteld dat het me weinig moeite kostte om op de tactiek van de glimlach over te gaan, want ik ben ervan overtuigd dat deze affaire inderdaad niet zo veel voor Maurice betekent.*



Table 3. *Frequencies of French translations of want and omdat*

French translation	Translations of <i>want</i> (N = 127)	Translations of <i>omdat</i> (N = 153)
Car	76 (59.8%)	7 (4.6%)
Parce que	8 (6.3%)	79 (51.6%)
Comme ‘as, since’	1 (0.8%)	17 (11.1%)
Puisque ‘since’	3 (2.4%)	3 (2.0%)
No translation / punctuation	25 (19.7%)	15 (9.8%)
Gerund / (pour) infinitive	3 (2.4%)	12 (7.8%)
Reformulation	11 (8.7%)	20 (13.1%)

- E: ‘I have told her that it didn’t take me much effort to proceed to the tactics of the smile, car/want/because I am convinced that this affaire does not mean much for Maurice.’
- (6) F: *Je croyais aux couples, **parce que** je croyais au nôtre.*  
(Fiction-FR)  
D: *Ik geloofde in paren, **omdat** ik in onszelf als paar geloofde.*  
E: ‘I believed in couples, parce que/omdat/because I believed in us (as a couple).’
- (7) F: *Kerry n’a pas oublié l’échec subi par Bill Clinton, en 1995, **parce que** sa réforme faisait trop appel au budget et à l’intervention de l’Etat fédéral.*  
(Newspaper-FR)  
D: *Kerry is de mislukking van Bill Clinton in 1995 niet vergeten; zijn hervorming kostte een te grote hap uit het budget en deed een beroep op de inmenging van de federale staat.*  
E: ‘Kerry did not forget Bill Clinton’s failure in 1995, parce que;/ (because) his reform took too large a slice from the budget and appealed to interference of the federal state.’

Table 3 indicates that *want* is most often translated by *car* (see (8)), and *omdat* by *parce que* (see (9)). In addition, *comme* ‘as, since’ is a frequent equivalent of Dutch *omdat* (cf. (10)). The causal connectives are frequently omitted in the translations, or translated by nonconnective linguistic alternatives (e.g., syntactic alternatives such as the gerund, or the *pour*-infinitive, or lexical alternatives like *à cause du fait que* ‘because of the fact that’ or *en raison de* ‘for the reason that’).

- (8) D: *Gauw vrijlaten die fascist, **want** wij zijn geen fascisten, wij houden onze handen schoon.*  
(Fiction-NL)

- F: *Dépêchons-nous de le libérer, car nous ne sommes pas des fascistes, nous, nous gardons les mains propres.*
- E: 'Let us quickly release that fascist, want/car/because we are not fascists, we keep our hands clean.'
- (9) D: *Dat weet ik omdat je oom hier kort na de bevrijding is geweest.*  
(Fiction-NL)
- F: *Je le sais parce que ton oncle est venu ici juste après la libération.*
- E: 'I know that omdat/parce que/because your uncle has been here shortly after the liberation.'
- (10) D: *Omdat hij voorlopig toch ook in de weekends ergens heen moest, kocht hij nog een kleine boerderij in Gelderland (. . .)*  
(Fiction-NL)
- F: *Comme il lui fallait bien, pour le moment, aller quelque part en week-end, il acheta en Gueldre une petite ferme (. . .)*
- E: 'Omdat/Comme/Since he temporarily had to go places in the weekends, he bought a small farm in Ghelderland'

For all four connectives, we also investigated how the initial set of translations in the target language was translated back into the original language. For example, in the case of Dutch *want*, we made an inventory of all the French linguistic items that were used as its translations (e.g., *car*, *parce que*), and subsequently listed how these French items themselves were translated in the Dutch corpus. This back-and-forth translation revealed which Dutch counterparts could be regarded as the linguistic competitors of *want*.<sup>6</sup> From the results of this mirror analysis, we derived a translation network, within which the different markers are organized according to their respective importance within the field.<sup>7</sup> We used three criteria to determine this relative importance, see (11) (cf. also Mortier and Degand 2009).

- (11) Criteria to determine the importance of a linguistic marker within the translation network
- a. The overall frequency of the marker
  - b. The number of relations
  - c. The strength of relations

Criterion (11) looks at the overall frequency of the marker: the primary markers have a high frequency in the corpus data (*omdat* occurs 4.3 times per 10,000 words, *want* 3.5, *car* 2.4, and *parce que* 2.8 per 10,000 words), when compared to alternatives such as *doordat* (0.02) or *puisque* (0.3). This also works the other way around: if markers occur extremely infrequently, they are probably less relevant for the semantic field, and hence are not placed at the core of the map.

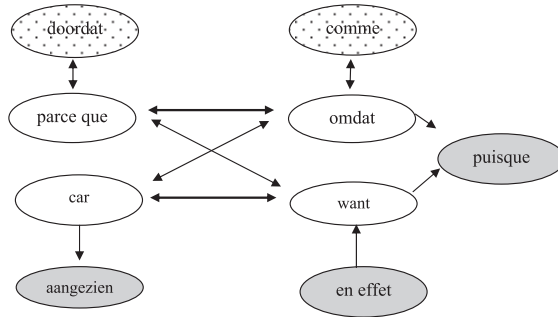


Figure 1. Translation network of *want*, *omdat*, *car* and *parce que*

Criterion (11) takes into account the number of relations: the more relations a marker entertains with other (primary) markers, the more it is at the core of the semantic field, especially when these relations are bidirectional (e.g., when *parce que* is translated by *omdat* and *omdat* by *parce que* in a significant amount of cases). Again, this argumentation can also be inverted: if L2 translations from L1 markers are entered into a back and forth analysis and do not provide L1 output with a meaning that is at least partially related to the original L1 markers, then they are less likely to be relevant for the semantic field. They probably belong to a different semantic field, they have a very general, non-specific meaning, or they are instances of lexical reformulations that are most likely the result of translator interference.

Criterion (11) investigates the strength of relations: a high frequency of translation pairs suggests a strong correlation between markers which thus have comparable semantic ‘weight’ in the given field. For example, *parce que* is translated by *omdat* in 41 out of the 54 cases (77.4%), whereas its translation by *doordat* does not exceed 1.9% (1 case). A hierarchy of equivalents is thus established, with three main categories: primary, secondary, and tertiary equivalents.

Our analysis resulted in the following translation network (see Figure 1). The white fields represent the primary equivalents, the core of the semantic field with the maximum of relations between the four constitutive connectives. The dotted fields represent the secondary equivalents associated with *parce que* and *omdat*; secondary, because they still have a two-way relationship with those markers. The grey fields represent the tertiary equivalents associated with *car* and *want*, on the one hand, and with *omdat* and *want*, on the other hand. They are tertiary because of their one-way relation with these causals. Bold-faced arrows indicate that equivalents are translated into each other over 50% of the time.

From Figure 1 it appears that *comme* and *en effet* are serious competitors for the two French causals. For Dutch, *doordat* and *aangezien* are the most

important linguistic alternatives to be looked at when starting diachronic analyses.

An additional analysis we could have performed on the translation data in Corpus 1 concerns an analysis in terms of subjectivity. By establishing the domain of each source fragment, we could find out whether connective use in specific domains results in different translations. Prior research on the “translation pair” *puisque* and *aangezien* (Degand 2004) has indeed shown that translators tend to respect the level of subjectivity expressed by the connective fragments. It would thus be conceivable that e.g., nonvolitional *doordat* would be translated into *parce que*, but not into *car*. Similarly, it is likely (given previous synchronic analyses of *want* and *omdat*, cf. Pit 2003, among others) that the French *car*-fragments that are translated into *omdat* are more objective than the fragments translated with *want*. Such a subjectivity analysis could result in a subjectivity scale of the various linguistic markers. However, given limitations of time and space, we have not performed such a subjectivity analysis on the present data.

#### 4. Results Corpus 3: analysis of various texts from different periods

We will now turn to Corpus 3, leaving Corpus 2 for discussion in the next section. Analyzing various texts from different periods is the most common way to investigate the historical development of linguistic phenomena. This allows researchers to perform both qualitative and quantitative analyses. An additional advantage is that an increasing amount of digital diachronic data is becoming available for research so that researchers can base their claims on actual diachronic data in context. Such information about contexts of use is lacking for diachronic data from dictionaries, which do not reveal frequency patterns either.

Disadvantages of Corpus 3 are that it restricts the researcher to the analysis of written data and that the results may suffer from a possible confounding with genre effects: historical “developments” may be the result of studying different text types in different periods. For example, corpora with ancient texts often contain charts, moralistic texts and rhyming literature, whereas corpora with modern texts are often compiled of journalistic texts and nonrhyming novels. Ideally, Corpus 3 should be a comparable corpus, in the sense that it should contain texts matched by such criteria as domain, genre, intended audience, etc. (cf. Johansson 1998: 5). It is not always possible to compile such a comparable corpus, however, because of the restricted availability of ancient texts.

We analyzed *want*, *omdat*, *car* and *parce que* in various texts from different periods. We included various spellings, which is an important point for Old French and Middle Dutch in particular. Table 4 shows more details on our se-

Table 4. *Corpus 3: data on the Dutch and French diachronic corpora containing various texts*

	Dutch	French
Periods	13th, 16th and 20th century	Old French (OF, 900–1300), Middle and Preclassical French (MF, 1351–1550), and Present-day French (PDF, 1990–2000)
Data	Equal distribution of rhyming, literary texts and nonrhyming, nonliterary texts, mostly from CD-ROMs ( <i>Middelnederlands</i> ‘Middle Dutch’ and <i>Klassieke literatuur</i> ‘Classical literature’) and the Internet (e.g., the project <i>Laurens Jansz. Coster</i> ) <sup>9</sup>	Exclusively literary texts, from the BFM Database (Base du Français Médiéval), the DMF Database (Dictionnaire du Moyen Français), and the Frantext database
Fragments	50 per connective per period	50 per connective per period
References	Evers-Vermeul (2005)	Degand and Fagard (2008), Degand and Fagard (submitted), Fagard (2008), Fagard and Degand (2008)

lection of periods, texts, and number of connective fragments. We selected 50 occurrences per period for each marker, and then registered formal and functional aspects of each occurrence. On the formal side, we looked at the categorical status of the connective, and at the positioning of the connective clause as a whole, distinguishing between pre- and postpositioning. On the functional side, we analyzed the relation type (causal, temporal, concessive or other) and the domain type (nonvolitional content, volitional content, epistemic, and speech act).<sup>8</sup>

Our French analyses reveal that both grammaticalization and subjectification are involved in the early stages of the evolutions of *car* and *parce que* (see for more details Degand and Fagard 2008; Fagard 2008; Fagard and Degand 2008). The rise of the connective *car* shows a series of features associated with grammaticalization: phonological reduction and internal bonding (from Latin *qua re* ‘for which/what reason’ to Middle French *quar/quer*, to PDF *car*). In addition, *car* changes from a complex subordinating conjunction to a simple coordinating conjunction, resulting in the loss of *car*’s ability to occur in preposed connective clauses. The original meaning of *res* ‘object, cause’ progressively fades to the point that the presence of the noun is completely hidden not only by phonetics (*re* > *r*) but also by semantics (an instance of semantic bleaching). This early stage in the development of *car* also involves subjectification. Where *qua re* functions at the relatively objective referential level, *car* functions at the more subjective textual level, indicating a causal coherence relation put forward by the speaker. The grammaticalization of *parce que* is shown by loss of variation, phonetic attrition and internal bonding (OF *par/porce que* ‘for this that’ > MF *parce* (. . .) *que* ‘because’ > PDF *parce que/paske*).

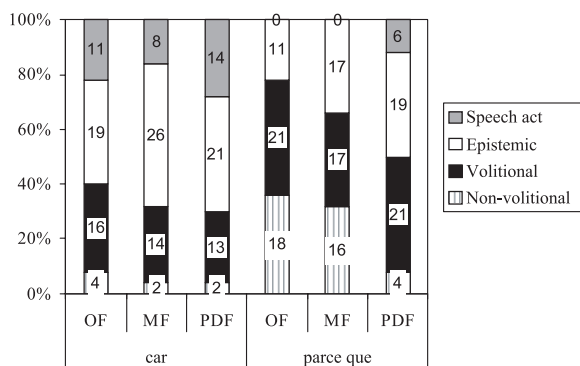


Figure 2. Distribution of *car* (left) and *parce que* over the domains of use in three periods<sup>10</sup>

Subjectification occurs during the process of semantic bleaching: *ce* ‘this’ is anaphoric in OF *por ce que*, but not in PDF *parce que*. Hence, the new connective *parce que* only functions at the more subjective textual or discourse level, and no longer at the referential level.

The corpus approach enabled us to quantify the later stages in the evolution of *car* and *parce que*, in which these words continued to be used as a connective (see Figure 2). Our subjectification study of the French part of Corpus 3 reveals that the use of both *car* and *parce que* remains relatively stable throughout the centuries (Degand and Fagard [2008], see also Degand and Fagard [submitted]).

For *car*, both speech act and epistemic uses are already present in Old French (cf. examples (12) and (13)). The use of *car* remains stable over time ( $\chi^2(6) = 4.1$ ;  $p = .66$ ). It is mainly used in volitional and epistemic contexts, it also occurs with speech-act uses (especially in speech-like contexts such as quoted speech). Its nonvolitional use is fairly rare. Thus, it can be described as a subjective connective as soon as it has become grammaticalized.

*Parce que*, which remains a subordinator throughout the ages, is predominantly used in the content domain (see (14)). The degree of subjectivity of *parce que* in written data seems to change over time ( $\chi^2(6) = 28.8$ ;  $p = .001$ ). In PDF, *parce que* witnesses a dramatic drop of its nonvolitional uses ( $z = -2.4$ )<sup>11</sup>, while it occurs for the first time in speech-act contexts ( $z = +2.8$ ). However, we have to be careful with the interpretation of these data. The overall frequency of speech-act uses of *parce que* being very low, no strong conclusions can be drawn at this stage. There seems to be a tendency for *parce que* to increase its intersubjective (speech-act) uses, but further analyses should help determine whether a genuine intersubjectification process is ongoing (cf. Degand and Fagard submitted). In addition, while it is true that the nonvolitional uses of *parce que* decrease over time, overall, its more objective contexts of

use (nonvolitional and volitional uses together) remain stable over time. In other words, it is not the case that *parce que* becomes less objective.

- (12) Speech act connective use of *car* (12th century)  
*Di va! fet il, pour coi me celes? Car assez set on les noveles, que Edypus fist de son pere quant il l'ocist et prist sa mere.*  
 'Tell me! he exclaimed, why do you keep it a secret? **For** we know very well what Oedipus did to his father, killing him and taking his own mother.'  
 (*Le roman de Thèbes*, c. 1150, v. 865)
- (13) Epistemic connective use of *car* (12th century)  
*Par foi, ci le cuidai trover, sire, car ains de moi leva.*  
 'Truly, I thought I would find him here, sir, **because** he got up before me.'  
 (*Floire et Blancheflor*, 1150–1160, v. 2568)
- (14) Content connective use of *parce que* (13th century)  
*Li anfes ploroit de grant fin por ce que n'avoit que mengier.*  
 'The child cried of hunger, **because** he had nothing to eat.'  
 (*Roman de Renart*, early 13th c.)

Our analysis of the Dutch connectives in Corpus 3 reveals that the rise of *omdat* involves grammaticalization (see Evers-Vermeul 2005 for a more detailed analysis). *Omdat* shows internal bonding and reanalysis of the preposition *om* 'for' and the relativum *dat* 'that'; it became a fixed combination, the subordinator *omdat* 'because', in the 16th century. Its positioning properties remain stable over time (cf. (15)). The rise of the connective *omdat* also involves semantic bleaching and subjectification. *Om + dat* functions at the objective referential level, whereas the conjunction *omdat* functions at the more subjective textual or discourse level. In the 13th and 16th century, *want* could be used both as a subordinator and as a coordinator.<sup>12</sup> For example, in (16) *want* is used in a clause with a finite verb (*was*) in final position, a syntactic configuration typical of Dutch subordinating clauses. Also, *want* could appear in preposed connective clauses, a use that is not considered grammatical for Dutch coordinators. Over time, *want* changed into a pure coordinator: its occurrence in modern Dutch is restricted to postposed connective clauses with Verb Second, as in (17). Note, however, that this syntactic change in the use of *want* does not involve subjectification: the loss of the subordinating use is not accompanied by a loss of or decrease in content use (see Figure 3).

- (15) Subordinating content use of *omdat*  
*Zij (= Célestine) knikte hem met een fijn glimlachje toe, en gaf niet meer uitleg, omdat Sarah al teruggelopen was en naast haar stond.*



‘She (= Célestine) nodded to him with a subtle smile, and did not explain anymore, because Sarah had returned already and stood next to her.’

(*Het verbroken zegel*, 1991)

(16) Subordinating content use of *want*

*Doen dit ghedaen was, soo ghingen si ter maeltijt (wantet byder noenen was)*

‘When this was done, they went to have a meal (because it was almost noon)’

(*Historie van Malegijs*, 1556)

(17) Coordinating epistemic use of *want*

*Er moet hem een behoorlijk bedrag nagelaten zijn, want zijn ouders waren redelijk welgesteld.*

‘A substantial amount must have been left to him, because his parents were fairly well off.’

(*In de schaduw bloeien de rozen*, 1994)

During the later stages of the diachronic development, Corpus 3 shows a fairly stable profile for both *want* and *omdat*, not straightforwardly supporting the subjectification hypothesis of discourse markers. Figure 3 charts the distribution of *want* and *omdat* over time. It reveals that *want* is mainly used as a marker of epistemic causal relations, whereas *omdat* mainly occurs in content relations. Statistical analysis indicates that the domains profile of *omdat* is stable across ages ( $\chi^2(6) = 10.1$ ;  $p < .25$ ). The only change in its use is that, after the 16th century, *omdat* has lost its ability to mark result (‘so that’) causal relations. The connective *want* has hardly changed during the selected time span of 800 years; only in the 16th century was a significant increase in speech-

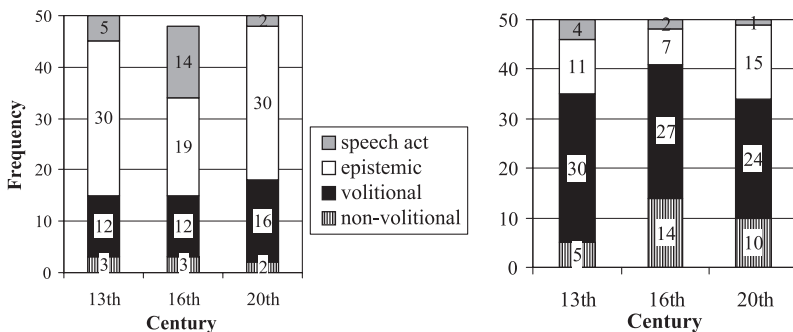


Figure 3. Distribution of *want* (left) and *omdat* over the domains of use in three periods<sup>13</sup>

act use found ( $\chi^2(1) = 12.2$ ;  $p < .001$ ). This increase seems to point to subjectification. However, this subjectification was not a lasting phenomenon, since the number of speech-act fragments decreased again in the 20th century.

Our analysis of the speech-act fragments from the 16th century revealed that seven of the fourteen *want*-fragments came from the same moralistic source — *Devoot ende profitelyck boecxken* ‘Devout and profitable book’ — in the sample of rhyme texts. In this text from 1539, advice and orders like (18) are presented for a “good life style”. This advice and these orders are frequently supported with arguments, which results in the high number of speech-act relations. The temporary increase in speech acts, then, should not be seen as a case of subjectification, but as a genre effect.

- (18) *God kent sijn schapen ende sij hooren na sijn stem Nyemant en machse trecken wt sijnder hant / Sijn woert aenhoert want God ghetuycht van hem*  
‘God knows his sheep and they listen to his voice No one can draw them from his hands / Listen to His (Jesus’) word because God testifies to him’

All in all, our study of Corpus 3 shows both grammaticalization and subjectification during the rise of three of the four connectives under investigation (*omdat*, *car*, *parce que*). The rise of these connectives involves a transition from use of linguistic items at the relatively objective referential level to use of these items (albeit in a condensed form) at the more subjective textual or discourse level. For *want*, our corpus did not contain any data on its rise as a connective. During the later stages of the development, once the items serve as connectives, subjectification is not a frequent phenomenon. However, inter-subjectification might be at stake for *parce que*, which could be driven by an increasing use in spoken language (cf. Section 6).

## 5. Results Corpus 2: analysis of same texts in different periods

Section 4 revealed that researchers who study a variety of texts in different periods run the risk of interpreting genre effects as a change of the linguistic phenomenon under investigation (cf. the temporary increase of *want* in speech-act relations in the 16th century, as shown by Figure 3). Only texts from certain genres tend to be preserved from any historical period, making it difficult to separate out the effects of diachrony from the effects of genre (cf. Herring et al. 2000, and the references cited there). Of course there are other factors that affect the homogeneity of the corpus. For example, if texts from different periods vary in register, dialect, and/or subject matter (cf. Biber et al. 1998: 248), this may also result in a diachronic difference being unjustly interpreted as a change

of the linguistic phenomenon under investigation. We will focus on genre effects here, because — as Condamines (2008: 115) puts it — “Within corpus linguistics, genre is one of the most crucial but also one of the most difficult problems to be tackled.”

There are at least two ways out of the problem of genre diversity. Gries (2006) proposes a rather sophisticated statistical solution, which enables researchers to calculate effect sizes of variability within and between corpora. For researchers who are less statistically equipped, a second solution may be more attractive: take genre into account during the analyses (cf. the suggestions in Condamines 2008). This is also useful when trying to identify regularities in the use of linguistic items which depend on the purpose of the text.

One way to take genre into account is to keep the genre constant throughout the ages and study a number of texts of the same genre or a limited set of genres in each period. Although this restricts the generalizability of the conclusions, it avoids the confounding problem by and large. A disadvantage of this method is that the characteristics of the genres themselves may change over time and that it introduces a new danger: confusing linguistic changes with genre changes (see, for example, Claridge and Wilson 2002 on the evolution of the genre “sermons”). A second way of taking genre into account is to use different translations of the same source text. Because the same source text forms the basis of the translation in each period, the effect of changing genre conventions is diminished. That’s why this approach enables the researcher to separate out the effects of diachrony from the effects of genre.

Our Corpus 2 contains the same text in different periods: four Dutch translations of the Bible (see Schoonenboom 2000 and Vogl 2007 for a similar approach). Because of the scarce availability of ancient French Bible translations, we could not perform a comparable study for the French language. Just like Corpus 3, this corpus may reveal actual semantic and/or structural diachronic changes. In addition, this corpus gives insight into the linguistic competitors of the connectives under investigation. Disadvantage of this method is that it concerns a very specific genre, namely religious texts, and that a translation effect may occur (cf. Gellerstam 1996). For example, Degand (2004) found that *aangezien* ‘because’ is more subjective in texts translated from French into Dutch than in original Dutch texts, because of transfer of the subjectivity profile of French *puisque*. Observations based on such a translation corpus, then, need to be checked against a control corpus consisting of comparable original texts in the same language.

The selection of Bible translations has its own merits, which are less likely to be found when using ancient translations of other source texts. First, the presence of chapter and verse numbers facilitates a comparison of the translations from different periods. Second, the Bible does not represent just one specific genre, but it contains a variety of text types, including stories, proverbs,

Table 5. *Corpus 2: data of the Dutch corpus of Bible translations*

Translation	Year of publication	Number of words in Genesis
Delftse Bijbel (DB, see Van der Sijs 2008a)	1477	35,659
Statenvertaling (SV, see Van der Sijs 2008b)	1637	37,754
Translation of the Dutch Bible Association (NBG)	1951	36,482
New Bible Translation (NBV)	2004/2007	35,319

Table 6. *Corpus 2: results on the diachronic corpus containing Dutch Bible translations (N = 248)*

Translation	Want (%)	Omdat (%)	Other or no marker (%)
DB 1477	45	23	32
SV 1637	52	10	38
NBG 1951	41	11	48
NBV 2004	24	15	61

songs, and argumentative texts. Third — although this certainly does not hold for all languages — for Dutch a relatively large amount of ancient Bible translations is available (thanks to the efforts of Nicoline van der Sijs, see Van der Sijs 2008a, 2008b). Fourth, the translations are often constructed by a team of translators, rather than by one individual. This diminishes the effects of personal preferences of individual authors. Finally, it is possible to compare the Dutch findings with translations in other languages.

The use of Bible translations has some disadvantages as well. First, there may be effects of the religious nature of the texts: the language use may be more formal, or lag behind compared to the language used in secular documents. Second, it is not possible to cover the complete history of a language using only Bible translations. Furthermore, it is hard to incorporate other factors causing diversity in the corpus. For example, it would be very difficult to take dialect variation into account. All these disadvantages affect the generalizability of the conclusions.

Details on the Dutch corpus of Bible translations can be found in Table 5. We focused on the book of Genesis, which consists of 50 chapters and 1533 verses.

We analyzed all fragments that were marked with *want* or *omdat* in at least one of the translations (cf. Evers-Vermeul 2008). In total, 248 fragments were selected. For each fragment, we checked whether it was marked with *want*, *omdat*, or with some other or no marker. Table 6 shows the percentages of use of these markers in the selected fragments.

Statistical analysis reveals that the distribution over the connectives is not stable over time ( $\chi^2(6) = 28.4; p < .001$ ). The percentage of *want*-fragments is

relatively stable in the first three translations. Only in the translation of NBV 2004, there is a significant decrease in the number of *want*-fragments ( $z = -2.6$ ). This can be ascribed to a preference of the 2004 translators to leave the causal coherence relation implicit, which also results in a significant increase in other uses in that translation ( $z = 2.4$ ). For example, Genesis 3:19 contains *want* in the first three translations (see (19)a for the version of SV 1637), but it lacks a causal marker in the NBV 2004 (marked with  $\emptyset$  in (19)b). Translations from other languages do insert a causal connective (German *denn*, English *for* and French *car*), suggesting the source text contains a causal marker as well.

## (19) a. SV 1637

*In 't sweet uwes aenschijns sult ghy broot eten, tot dat ghy tot d'aerde wederkeert, dewijle ghy daer uyt genomen zijt: want ghy zijt stof, ende ghy sult tot stof wederkeeren.*

'In the sweat of thy face shalt thou eat bread, till thou return unto the ground; **for** out of it wast thou taken: for dust thou [art], and unto dust shalt thou return.'

(King James 1611)

## b. NBV 2004

*Zweten zul je voor je brood, totdat je terugkeert tot de aarde, waaruit je bent genomen:  $\emptyset$  stof ben je, tot stof keer je terug.*

'You will have to sweat for your bread, until you return to the ground, from which you were taken:  $\emptyset$  dust you are, and to dust you will return.'

The percentage of *omdat* drops after the first translation (in which it occurs more frequently than in the other three translations;  $z = 2.1$ ), but remains stable after that. This drop is probably due to the disappearance of the resultative use of *omdat*. 25 Fragments that would be marked with a *to*-infinitive in English, a *pour*-infinitive in French or (*auf*) *daß* or *zu*-infinitive in German, contain *omdat* in the translation of DB 1477, but *opdat* 'so that' or an *om*-infinitive in more recent translations (compare the a- and b-examples in (20) and (21)). In addition, the DB 1477 contains several *omdat*-clauses that are left out in the other translations.

## (20) Genesis 21:30

## a. DB 1477

*Du sulste die seuen oyen ontfaen. van mijnre hant. **om dat** si mi een oercontscap sullen sijn: dat ic desen put groef.*

## b. NBG 1951

*Voorzeker moet gij de zeven lammeren uit mijn hand aannemen, **opdat** het mij tot een getuigenis zij, dat ik deze put gegraven heb.*

'These seven ewe lambs shalt thou take of my hand, **that** it may be a witness unto me, that I have digged this well.'

- (21) Genesis 15:7
- a. DB 1477  
*Ik bin die here di v wleide van hur van chaldea: om dat ic di geuen soude dat lant ende dattu dat besitten souste*
  - b. NBG 1951  
*Ik ben de HERE, die u uit Ur der Chaldeeën heb geleid om u dit land in bezit te geven.*  
'I am the LORD that brought thee out of Ur of the Chaldees, to give thee this land to inherit it.'

What do these data tell about changes in the degree of subjectivity of *want* and *omdat*? Firstly, they confirm previous findings that *want* is the more subjective of the two: 28 fragments that contain *want* in three of the four translations, are fragments that contain the more subjective markers in other languages: German *denn*, English *for*, and French *car*. Fragments containing *omdat* are often more objective and equal fragments with German *darum daß*, English *because*, and French *parce que*.

Section 4 showed that the connective *want* lost the ability to be used as a subordinator. This finding is confirmed by this analysis (compare the three translations of Genesis 30:18 in (22)). However, this loss does not seem to affect the overall subjectivity profile of *want*, because the more objective subordinating use was far less frequent than the coordinating use of *want*.

- (22) Genesis 30:18
- a. DB 1477  
*God heuet mi desen loen gegeuen: want ic mijn ioncwijf minen man gaf*  
'God has given me my hire, because I gave my handmaid to my husband'
  - b. SV 1637  
*Godt heeft mijnen loon gegeven; na dat ick mijne dientmaecht mijnen man gegeuen hebbe*  
'God has given me my hire, after I gave my handmaid to my husband'
  - c. NBV 2004  
*God heeft mij beloond omdat ik mijn slavin aan mijn man heb gegeven*  
'God has rewarded me, because I gave my bondwoman to my husband'

In the most recent translations, *want* appears to loose ground to markers such as *namelijk* 'namely', *immers* 'indeed', *tenslotte* 'after all', and *toch* — which would occur as a question tag in English. These markers indicate that certain

information is already given, or accessible to the reader. For example, *want* in the SV 1637 translation of Genesis 4:25b in (23) shows up as *immers* ‘indeed’ in the NBG 1951, and as a relative clause in the NBV 2004. Because these fragments take the knowledge of the reader into account, they can be labeled intersubjective (in the sense of Traugott and Dasher 2002: 22). This suggests that *want* becomes more restricted in its more subjective use.

## (23) Genesis 4:25b

## a. SV 1637

*Godt heeft my een ander zaet geset voor Habel; want Kain hem dootgeslagen heeft.*

‘God has appointed me another seed instead of Abel, **for** Cain slew him.

## b. NBG 1951

*God heeft mij een andere zoon gegeven in plaats van Abel; hem immers heeft Kain gedood.*

‘God has given me another son instead of Abel; **after all** he was killed by Cain.’

## c. NBV 2004

*God heeft mij in de plaats van Abel, die door Kain is gedood, een ander kind gegeven.’*

‘God has given me another child to take the place of Abel, who was killed by Cain.’

Modern *omdat* replaces other causal markers such as *dewijl*, *overmits* (*dat*), *naardien* (*dat*), and *daarin dat*, which are not used in modern Dutch anymore. Because these archaic connectives are comparable in terms of subjectivity, this does not affect the subjectivity profile of *omdat*.

Much more can be said about the results of Corpus 2. The analysis so far, however, already shows the usefulness of analyzing ancient Bible translations for diachronic research. Because Bible translations are very convenient for tracing linguistic competitors, Corpus 2 appears to be especially suitable for an onomasiological (“function-to-form”) approach, which may supplement studies with a semasiological (“form-to-function”) approach. In addition, Corpus 2 can be used as a control corpus for corpora with different texts from different periods, in order to be able to distinguish real changes from possible genre effects.

## 6. Results Corpus 4: comparison of present-day spoken and written language

Section 4 showed the need for taking genre differences into account. This confirms the idea that the range of text categories (or *registers*) that samples are



selected from is one of the two major issues in corpus linguistics.<sup>14</sup> Biber (1988, 1993, and many other works) has argued repeatedly that register variation is inherent to natural language and that diversified corpora representing a broad range of register variation are required as the basis for general language studies, especially for the external validity of the corpus study (i.e., the extent to which it is possible to generalize from a sample to a larger target population).

Although we acknowledge the need for using multigenre corpora, we will focus here on the importance of distinguishing between the two primary modalities of language: speech and writing.<sup>15</sup> First of all, we think this distinction is the most basic one of all the distinctions that can be made in order to incorporate register variation. Second, this distinction is relatively easy to operationalize, for synchronic and even for diachronic data. For example, Biber (1988) lists five dimensions to define similarities and differences among spoken and written registers. In addition, Koch and Oesterreicher (1985) provide criteria differentiating between the more “speech-like” data (e.g., drama, conversations in literature) and real “written” data (cf. also Chafe and Danielewicz 1987; Jacobs and Jucker 1995: 8; Traugott and Dasher 2002: 47). A third reason to focus on differences between the two modalities is that the comparison of present-day spoken and written data may reveal diachronic change “in progress”. It is often claimed that changes first occur in spoken language and only gradually make their way into written texts (cf., among many others, Hansen and Rossari 2005: 181; Croft 2000: Chapter 4). Fourthly, if a researcher finds differences between oral and written language, that may invite the researcher to zoom into this finding and perform a subsequent diachronic analysis including a distinction between the two modalities (see e.g., Degand and Fagard 2011), or restricting the research to one of the two modalities. For example, Lindström and Wide (2005) study the diachronic development of Swedish discourse particles of the type *you know*, and restrict their study to historical texts that have at least some interactive properties, and therefore may reflect colloquial language use in which such markers can be expected.

Analyzing present-day data — whether speech-like or written — in order to gain insight into diachronic processes is not completely unproblematic. First, these data are not instances of actual diachronic data, and whether it is possible to recognize early grammaticalization from synchronic data is a point considered controversial in the literature. As Mair (2004: 131) points out, some researchers are extremely skeptical about the possibility of observing grammaticalization processes unfolding in the field (cf. Compes et al. 1993: 20), whereas others seem to take a middle road. For example, Lehmann (1991: 532) writes in his study of ongoing change in present-day German, “Given presently available methodological means, it is next to impossible to know which of the changes that speech habits currently exhibit are synchronic manifestations of ongoing language change, and which of them are but ephemeral fashions.”

Table 7. *Corpus 4: Data on the Dutch and French present-day corpora*

	Dutch	French
Written data	Journalistic: NRC Handelsblad 1994	Journalistic: Le Soir 1997
Spoken data	CGN: spontaneous conversations and interviews (see Oostdijk 2000)	Valibel: spontaneous conversations and interviews (see Francard et al. 2002)
Number of fragments	Written: 50 <i>want</i> , 50 <i>omdat</i> Spoken: 149 <i>want</i> , 124 <i>omdat</i>	Written: 50 <i>car</i> , 50 <i>parce que</i> Spoken: 50 <i>car</i> , 50 <i>parce que</i>
References	Degand and Pander Maat (2003); Spooren et al. (2010)	Degand and Pander Maat (2003); Simon and Degand (2007)

A second problem concerns the nature of the “spoken” data. For example, the *Corpus Gesproken Nederlands* (CGN) does not only contain spontaneous conversations and interviews, but also prepared speeches and stories that are read out loud. This means again that researchers studying spoken data need to be careful in their collection of data (cf. Section 4). A final problem concerning the use of spoken data is that they are most often analyzed from a transcribed version, which introduces problems of its own (cf. Halliday 2004: 15–21).

Our final corpus then, Corpus 4, is compiled of present-day French and Dutch data; for both languages, it contains one subcorpus of written data and another of spoken data, thus enabling a comparison of the two modalities. Table 7 introduces some relevant information about Corpus 4.

Analysis of the Dutch data in Corpus 4 reveals that *want* and *omdat* have different distributions in the two modalities (cf. Spooren et al. 2010 for a more detailed analysis): *want* is more frequent in spoken than in written Dutch (1640 vs. 686 instances per million words), whereas *omdat* is more frequent in written than in spoken Dutch (938 vs. 521 instances per million words). We believe that this higher frequency of *want* in the spoken data has to do with its multifunctional profile: *want* can be used to express almost all types of causal relations, thus working as a kind of default connective. This is less so for *omdat*, which in spoken language hardly occurs in highly subjective (epistemic and speech-act) contexts (see Table 8). Spooren et al. (2010) interpret this use of a default ‘general’ connective as an attractive option in terms of speaker economy: choosing the most general connective costs less energy. For Dutch, an additional explanation seems plausible as well: “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.” Table 8 shows the results of the domain analysis of the connective fragments.

This analysis confirms the findings of previous subjectivity analyses that *want* is more subjective than *omdat* ( $\chi^2(1) = 61.7$ ;  $p < .001$ ). *Omdat* occurs

Table 8. Domain type expressed by want and omdat in written and spoken Dutch<sup>16</sup>

	Written		Spoken	
	Content	Epist./speech act	Content	Epist./speech act
omdat	26	24	106	18
want	8	42	61	88

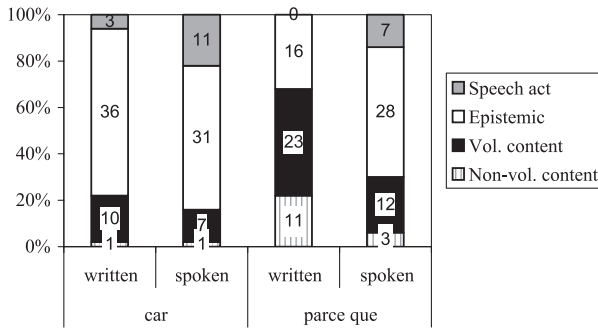


Figure 4. Domain type expressed by car and parce que in written and spoken French<sup>18</sup>

more often with content relations (132 of 174 or 75.9%), whereas *want* is more frequently found in noncontent relations (130 of 199 or 65.3%). Note that *want* occurs quite often in content relations<sup>17</sup> in spoken language, which confirms its status as “default” connective in speech. Nevertheless, both causals show a consistent semantic profile in writing and in speech; there is no three-way interaction between connective, medium and domain ( $\chi^2(1) = 1.9; p = .17$ ). Hence, the subjectivity profile of the connectives does not differ per medium. As such, no subjectification tendencies can be found for Dutch.

Analysis of the French data shows that *car* is more frequent in written than in spoken French (0.32% vs. 0.02%), whereas *parce que* is more frequent in spoken than in written French (3.70% vs. 0.40%). In written French, the two connectives occur in roughly the same frequency (0.32% vs. 0.40%), but in spoken French, *parce que* is 185 times more frequent than *car*. The frequencies of the two connectives in speech and writing are thus very different from those of their Dutch counterparts, a situation we believe to result from different causal paradigms in the two languages, related to different diachronic evolutions (cf. Section 4). The results of the domain analyses of *car* and *parce que* are given in Figure 4.

Previous subjectification analyses of these causals have shown that *car* has the same semantic profile in spoken and in written data ( $\chi^2(3) = 5.7; p = .14$ ), but that *parce que* shows divergent semantic profiles (see Degand and Pander

Maat 2003; Simon and Degand 2007). In written French, *parce que* is mainly used for objective functions; in spoken French, it can express both objective and subjective functions ( $\chi^2(3) = 18.3$ ;  $p < .001$ ). Example (24) shows an instance of the speech act use of *parce que* in the Valibel corpus.

- (24) *je crois que ça s'appelle en français mais excusez-moi parce que je vais peut-être [ . . . ] estropier le mot hein/un goupillon là*  
 'I think that in French this is called but excuse me because I might ruin the word / a "goupillon" [= sprinkler for holy water]'  
 (Valibel, chaBR1r)

It looks as if — in spoken French — *parce que* is taking over the role of *car*, including its more subjective functions. Hence, the difference between spoken and written *parce que* might reflect an ongoing change in French, one involving subjectification (see also Section 4). Although the future will have to prove or disprove this claim, we can conclude that this subjectification tendency has been strengthened by making the distinction between the two modalities (cf. Degand and Fagard submitted).

## 7. Conclusion

Our research started out with two questions: 1) Does subjectification occur in later stages of the diachronic development of causal connectives? and 2) What does a corpus-based approach add to the study of this diachronic development? In answer to the first question, we can state that — for the four connectives we studied — subjectification is not an integral part of the diachronic development of a causal connective. It appears, though, that subjectification does occur in the rise of these connectives. More specifically, we see this in the change a) from the preposition *om* 'for' plus the relativum *dat* 'that' to the subordinating connective *omdat*, b) from *par ce que* 'for this that' to the fixed phrase *parce que*, and c) from Latin *qua re* 'for which/what reason' to PDF *car*. These three changes are instances of grammaticalization, but they also show a shift from items being used at the relatively objective referential level to items being used at the more subjective textual or discourse level. In the later stages of their development, in which they remain to be used as a causal connective, only *parce que* seems to undergo subjectification. The other three causal connectives show hardly any subjectivity changes. On the contrary, *want* seems to lose some ground to intersubjectivity markers such as *immers* 'indeed', and *car* loses ground to *parce que* in the expression of speech-act causality in spoken PDF.

It appears that at least two general paths of development can be distinguished for causal connectives. The French case is an example of the first path, in which one causal marker (*parce que*) gradually takes over the function of an-

other marker (*car*). This development goes hand in hand with an increase in (inter)subjectivity: *parce que* can nowadays express speech act relations, which it could not express in earlier periods. The Dutch case is an example of the second path, in which no large changes occur in the system: both *want* and *omdat* remain stable markers of causality, each with their own degree of subjectivity.

Overall, the semantic profile of the connectives under investigation remains relatively stable. Sometimes connectives take over functions that were previously expressed by other markers (e.g., *omdat* nowadays expresses causal relations that were previously marked with *dewijl* or *overmits*), and sometimes a connective loses ground to another marker (e.g., the resultative use of *omdat* is taken over by *opdat* and *zodat* ‘so that’). This stability within the connective use need not come as a surprise: subjectification is not an obligatory characteristic of diachronic developments. An explanation of the stable subjectivity profile might be that three causal connectives did not show real changes in their domains of use in the sense that they came to be used in a domain in which they could not occur earlier. The connectives *want*, *omdat*, and *car* could be used in all three domains from the earliest century on. It is theoretically conceivable that subjectification of a connective occurs via a shift in the distribution over the meanings it can already express. However, this did not happen in the evolution of the connectives in our study. Hence, it may be the case that subjectification only occurs when items gain new meanings, and not when lexical items show a shift in the distribution over the meanings they can already express. This could indicate that, even though the two phenomena are not necessarily linked, a grammaticalized element undergoes further subjectification more easily as it grammaticalizes further (cf. Degand and Fagard submitted). Further research will have to reveal whether this restriction of the subjectification tendency applies to the development of other connectives and other linguistic markers as well, and hence should be seen as a fine-tuning of Traugott’s original subjectification claim.

Our second research question concerned the merits of a corpus-based approach: what does a corpus-based approach add to the study of this diachronic development? Our analyses have shown that each of the four methods introduced in Table 1 has its own merits and limitations. Although each of these methods is valuable in its own respect, we would like to stress here the fact that they are most effective when combined. An analysis of translation corpora provides the researcher with a semantic network, including synchronic linguistic competitors that may be relevant for the diachronic analysis as well. Hence, this method can be regarded as a good starting point for diachronic research or for a functional-semantic analysis. The analysis of data from different texts in different periods may reveal diachronic changes. This analysis should be at the heart of the study: it provides authentic data in their context of use, and allows

for both qualitative and quantitative research. The diachronic analysis of comparable texts (here, Bible fragments) may reveal whether these changes also occur when the genre is kept constant. This enables the researcher to separate out genre effects from real diachronic changes. This latter analysis may also reveal whether the proportion of use of linguistic competitors changes, and hence, whether changes in the connective system as a whole occur. Finally, the comparison of written and spoken data may reveal change in progress. The analysis of spoken and written materials is also useful in tracing stable differences between spoken and written language.

We hope to have shown that “grammaticalization studies can gain from the systematic and principled use of large computerized corpora and the methods which have been developed within corpus linguistics” (Lindquist and Mair 2004: x). We think that subjectification processes can be studied in much more detail, and that the results thus obtained will lead to a refinement of the theoretical model.

*Received 3 April 2009*  
*Revised version received*  
*19 March 2010*

*Utrecht University*  
*University of Louvain*  
*Lattice ENS & CNRS*

## Appendix. Primary sources

- BFM — *Base de Français Médiéval*. Lyon: UMR 5191 ICAR / ENS-LSH, 2005. <http://bfm.ens-lsh.fr> (accessed 21 January 2010).
- CD-Rom *Middelnederlands* ‘Middle Dutch’. 1998. Den Haag & Antwerpen: Sdu.
- CD-Rom *Klassieke literatuur: Nederlandse letterkunde van de Middeleeuwen tot en met de Tachtigers* [Classical literature: Dutch literature from the Middle Ages to the Eightiers]. 1999. Utrecht: Het Spectrum.
- DB 1477 — *Delfse Bijbel 1477*. Nicoline van der Sijs. 2008. <http://www.bijbelsdigitaal.nl> (accessed 21 January 2010).
- DMF — *Base du Dictionnaire de Moyen Français*, UMR 7118 ATILF / Nancy2, 2007. <http://atilf.atilf.fr/dmf> (accessed 21 January 2010).
- Frantext — *Base Textuelle Frantext*, UMR 7118 ATILF / Nancy2, 2008. <http://www.frantext.fr> (accessed 21 January 2010).
- NBG 1951 — *NBG-vertaling 1951*. Nederlands Bijbelgenootschap, 1951. <http://www.biblija.net> (accessed 21 January 2010).
- NBV 2004/2007 — *De Nieuwe Bijbelvertaling*. Nederlands Bijbelgenootschap, 2004/2007. <http://www.biblija.net> (accessed 21 January 2010).
- Oostendorp, M. van (ed.). *Project Laurens Jansz. Coster: Klassieke Nederlandstalige literatuur in elektronische edities* [Project Laurens Jansz. Coster: Classical Dutch literature in electronic editions]. <http://cf.hum.uva.nl/dsp/ljc/> (accessed 21 January 2010).

SV 1637 — *Statenvertaling 1637*. Nicoline van der Sijs. 2008. <http://www.bijbelsdigitaal.nl> (accessed 21 January 2010).

## Notes

- \* This research was conducted with the support of a IUAP-grant P6/44 Grammaticalization and (Inter)Subjectification financed by the Belgian Federal Government. The second author is a research associate of the Belgian Science Foundation FRS-FNRS. Correspondence address: Jacqueline Evers-Vermeul, Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, NL-3512 JK Utrecht, The Netherlands. E-mail: J.Evers@uu.nl.
1. This is very striking, since the sociolinguistic interviews she analyzes stem from three different years: 1971, 1984, 1995, and therefore would have allowed for a short-term diachronic analysis according to the real time method. Unfortunately, however, Vincent (2005) has chosen to group all the data together.
  2. See Noël (2003: 278) for some remarks on this terminology.
  3. Unfortunately, since there is no electronic database with Old and Middle French translations of the Bible, we could not do a comparable study on French connectives.
  4. Cf. Aijmer and Altenberg (1996: 12) for additional advantages of using translation corpora.
  5. See Noël (2003) for an overview.
  6. This was not possible for *pour*-infinitives and gerunds.
  7. Our concept of *translation network* bears similarities with the more common concept of *semantic map*, used among others by Haspelmath (2003) and Zwarts (2006), insofar as it congregates forms which all associate with a particular meaning in one or more languages. However, our map is a simplified version of such a semantic map, since no abstraction of conceptual features common to the set of forms was carried out. For that reason, we prefer the concept of *translation network* to that of *semantic map*.
  8. It should be noted that the Dutch and the French studies differ slightly in their operationalization of the epistemic relation: mental states such as “being sad” were classified as epistemic in the French studies, but as nonvolitional content in the Dutch studies. In the Dutch studies, the epistemic category only includes argumentative relations. For the purpose of this article, it was not necessary to correct this discrepancy; we did not compare the Dutch and French data in one encompassing statistical analysis, but only studied the diachronic development of the two languages separately. However, it would be interesting to reanalyze the data in order to find out whether this yields different results.
  9. See Chapter 4 in Evers-Vermeul (2005) for more details on the exact compilation of the Dutch corpus.
  10. For *car*, instances can be found in the earliest texts in the Old French period. *Parce que* does not occur before 1200. For the purpose of this article we performed a random selection of 50 occurrences per connective (in its causal connective usage) per period.
  11. The z-score gives the probability that a particular score will occur. A z-score of  $+/-1.96$  is significant at the 0.05 level,  $+/-2.58$  at the 0.01 level, and  $+/-3.29$  at the 0.001 level.
  12. Note that the rise of the connective *want* could not be attested in our corpus, because it already functioned only as a connective. Given the lack of written Dutch data before the 13th century, we could not trace the rise of the causal use ourselves. However, according to etymological dictionaries the connective *want* originated from the adverb *wan* ‘when’. This would imply a process of grammaticalization (a change from adverb to conjunction), but not necessarily of subjectification (both the temporal use and the causal use function at the discourse level, and are relatively subjective).



13. In the 16th century, two *want*-fragments were discarded from the analysis, because they marked a temporal instead of a causal relation.
14. The second major issue is the size of the corpus.
15. Of course we are aware of the fact that there are a number of “in between” realizations of language, such as “written-to-be-spoken”, chat, SMS, etc., but these finer distinctions do not matter to us here.
16. These data are taken from Table 6 in Spooren et al. (2010). Because the written data contained hardly any speech act relations, the speech acts were grouped together with the epistemic relations. 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. In Spooren et al. these were discarded from the analysis, but here we decided to include them in the data in order to ensure comparability with the written data from Degand and Pander Maat (2003) in which these metatextual uses were categorized as speech-act cases.
17. Of these content relations with *want* are volitional causal relations.
18. These data are taken from Figure 3 and Figure 4 in Simon and Degand (2007).

## References

- Aijmer, Karin. 2004. The semantic path from modality to aspect: *Be able to* in a cross-linguistic perspective. In Hans Lindquist & Christian Mair (eds.), *Corpus approaches to grammaticalization in English*, 57–78. Amsterdam & Philadelphia: John Benjamins.
- Aijmer, Karin & Bengt Altenberg. 1996. Introduction. In Karin Aijmer, Bengt Altenberg & Mats Johansson (eds.), *Languages in contrast: Papers from a symposium on text-based cross-linguistic studies*, 11–16. Lund: Lund University Press.
- Athanasiadou, Angeliki, Costas Canakis & Bert Cornillie (eds.). 2006. *Subjectification: Various paths to subjectivity*. Berlin & New York: Mouton de Gruyter.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas. 1993. Using register-diversified corpora for general language studies. *Computational Linguistics* 19(2). 219–241.
- Biber, Douglas, Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago & London: The University of Chicago Press.
- Chafe, Wallace & Jane Danielewicz. 1987. Chapter 3: Properties of spoken and written language. In Rosalind Horowitz & S. Jay Samuels (eds.), *Comprehending oral and written language*, 83–113. San Diego: Academic Press.
- Claridge, Claudia & Andrew Wilson. 2002. Style evolution in the English sermon. In Teresa Fanego, Belén Mendez-Naya & Elena Seoane (eds.), *Sounds, words, texts, and change: Selected papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000. Volume 2*, 25–44. Amsterdam & Philadelphia: John Benjamins.
- Compes, Isabel, Silvia Kutscher & Carmen Rudolf. 1993. *Pfade der Grammatikalisierung: Ein systematischer Überblick* (Arbeitspapier 17). Cologne: Sprachwissenschaftliches Institut der Universität.
- Condamines, Anne. 2008. Taking *genre* into account when analysing conceptual relation patterns. *Corpora* 3(2). 115–140.
- Croft, William. 2000. *Explaining language change: An evolutionary approach*. Essex: Pearson Education.

- Cuenca, Maria Josep. 1997. Form-use mappings for tag questions. In Wolf-Andreas Liebert, Gisela Redeker & Linda R. Waugh (eds.), *Discourse and Perspective in Cognitive Linguistics*, 3–19. Amsterdam & Philadelphia: John Benjamins.
- Dasher, Richard. 1995. *Grammaticalization in the system of Japanese predicate honorifics*. Stanford: Stanford University dissertation.
- Degand, Liesbeth. 2004. Contrastive analyses, translation and Speaker Involvement: the case of *puisque* and *aangezien*. In Michel Achard & Suzanne Kemmer (eds.), *Language, Culture and Mind*, 251–270. Stanford: CSLI Publications.
- Degand, Liesbeth & Benjamin Fagard. 2008. Intersubjectification des connecteurs: Le cas de *car* et *parce que*. *Linguística: Revista de Estudos Linguísticos da Universidade do Porto* 3(1). 119–136.
- Degand, Liesbeth & Benjamin Fagard (2011). *Alors* between discourse and grammar: The role of syntactic position. *Functions of Language* 18(1). 29–56.
- Degand, Liesbeth & Benjamin Fagard (submitted). Subjectify or die! On the evolution of the causal pair: *car* — *parce que*. *Journal of Pragmatics*.
- Degand, Liesbeth & Henk Pander Maat. 2003. A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In Arie Verhagen & Jeroen van de Weijer (eds.), *Usage-based approaches to Dutch*, 175–199. Utrecht: LOT.
- Doherty, Monika. 1998. Clauses or phrases — a principled account of *when*-clauses in translations between English and German. In Stig Johansson & Signe Oksefjell (eds.), *Corpora and cross-linguistic research: Theory, method, and case studies*, 235–254. Amsterdam & Atlanta: Rodopi.
- Dyvik, Helge. 1998. A translational basis for semantics. In Stig Johansson & Signe Oksefjell (eds.), *Corpora and cross-linguistic research: Theory, method, and case studies*, 51–86. Amsterdam & Atlanta: Rodopi.
- Dyvik, Helge. 2004. Translations as semantic mirrors: From parallel corpus to WordNet. *Language and computers* 1. 311–326.
- Erman, Britt & Ulla-Britt Kotsinas. 1993. Pragmaticalization: The case of *ba'* and *you know*. *Studier i Modern Språkvetenskap* N. S. 10. 76–93.
- Evers-Vermeul, Jacqueline. 2005. *The development of Dutch connectives. Change and acquisition as windows on form-function relations*. Utrecht: Utrecht University dissertation.
- Evers-Vermeul, Jacqueline & Ninke Stukker. 2003. Subjectificatie in de ontwikkeling van causale connectieven? De diachronie van *daarom*, *dus*, *want* en *omdat* [Subjectification in the development of causal connectives? The diachrony of Dutch *daarom* ‘that’s why’, *dus* ‘so’, *want* ‘for/because’ and *omdat* ‘because’]. *Gramma/TTT* 9(2/3). 111–139.
- Evers-Vermeul, Jacqueline. 2008. “Want ghy zijt stof . . .”; Over het belang van oude bijbelvertalingen voor diachroon onderzoek [Dutch ‘For dust thou art’; On the importance of ancient Bible translations for diachronic research]. Paper presented at Cogling 3, Leiden University, 19–20 December.
- Fagard, Benjamin. 2008. *Parce que, perché, porque* dans les langues romanes médiévales: l’utilité des études sur corpus. *Corpus* 7. 83–113. <http://corpus.revues.org/> (accessed 21 January 2010).
- Fagard, Benjamin & Liesbeth Degand. 2008. La fortune des mots: Grandeur et décadence de *car*. *Actes du Congrès Mondial de Linguistique Française* (CD-ROM). <http://www.linguistiquefrancaise.org/index.php?option=article&access=standard&Itemid=129&url=/articles/cmlf/pdf/2008/01/cmlf08213.pdf> (accessed 21 January 2010).
- Francard, Michel, Geneviève Geron & Régine Wilmet. 2002. La banque de données VALIBEL: des ressources textuelles orales pour l’étude du français en Wallonie et à Bruxelles. In Claus D. Pusch & Wolfgang Raible (eds.), *Romanistische Korpuslinguistik — Korpora und gesprochene Sprache | Romance Corpus Linguistics — Corpora and Spoken Language* (ScriptOralia 126), 71–80. Tübingen: Gunter Narr.

- Gellerstam, Martin. 1996. Translations as a source for cross-linguistic studies. In Karin Aijmer, Bengt Altenberg & Mats Johansson (eds.), *Languages in contrast: Papers from a symposium on text-based cross-linguistic studies*, 53–62. Lund: Lund University Press.
- Genetti, Carol. 1991. From postposition to subordinator in Newari. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. II, 227–255. Amsterdam & Philadelphia: Benjamins.
- González-Cruz, Ana I. 2007. On the subjectification of adverbial clause connectives: Semantic and pragmatic considerations in the development of *while*-clauses. In Ursula Lenker & Anneli Meurman-Solin (eds.), *Connectives in the history of English*, 145–166. Amsterdam & Philadelphia: John Benjamins.
- Gries, Stefan Th. 2006. Exploring variability within and between corpora: some methodological considerations. *Corpora* 1(2). 109–151.
- Günthner, Susanne. 1996. From subordination to coordination? Verb-second in German clausal and concessive constructions. *Pragmatics* 6(3). 323–356.
- Günthner, Susanne & Katrin Mutz. 2004. Grammaticalization vs. pragmaticalization? The development of pragmatic markers in German and Italian. In Walter Bisang, Nikolaus P. Himmelmann & Björn Wiemer (eds.), *What makes grammaticalization? A look from its fringes and its components*, 77–107. Berlin & New York: Mouton de Gruyter.
- Halliday, Michael A. K. 2004. The spoken language corpus: A foundation for grammatical theory. In Karin Aijmer & Bengt Altenberg (eds.), *Advances in corpus linguistics*, 11–38. Amsterdam & New York: Rodopi.
- Hansen, Maj-Britt Mosegaard & Corinne Rossari. 2005. The evolution of pragmatic markers: Introduction. *Journal of Historical Pragmatics* 6(2). 177–187.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language* 2, 211–242. Mahwah, NJ: Lawrence Erlbaum.
- Herring, Susan C., Pieter Th. van Reenen & Lene Schøsler. 2000. On textual parameters and older languages. In Susan C. Herring, Pieter Th. van Reenen & Lene Schøsler (eds.), *Textual parameters and older languages*, 1–31. Amsterdam & Philadelphia: John Benjamins.
- Higashizumi, Yuko. 2006. *From a subordinate clause to an independent clause: A history of English because-clause and Japanese kara-clause*. Tokyo: Hituzi Shobo.
- Hoffmann, Sebastian. 2004. Are low-frequency complex prepositions grammaticalized? On the limits of corpus data — and the importance of intuition. In Hans Lindquist & Christian Mair (eds.), *Corpus approaches to grammaticalization in English*, 171–210. Amsterdam & Philadelphia: John Benjamins.
- Jacobs, Andreas & Andreas H. Jucker. 1995. The historical perspective in pragmatics. In Andreas H. Jucker (ed.), *Historical pragmatics. Pragmatic developments in the history of English*, 3–33. Amsterdam & Philadelphia: John Benjamins.
- Johansson, Stig. 1985. Grammatical tagging and total accountability. In Sven Bäckman & Göran Kjellmer (eds.), *Papers on language and literature presented to Alvar Ellegård and Erik Frykman* (Gothenburg Studies in English 60), 208–220. Gothenburg: Acta Universitatis Gothoburgensis.
- Johansson, Stig. 1998. On the role of corpora in cross-linguistic research. In Stig Johansson & Signe Oksefjell (eds.), *Corpora and cross-linguistic research: Theory, method, and case studies*, 3–24. Amsterdam & Atlanta: Rodopi.
- Keller, Rudi. 1995. The epistemic *weil*. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, 16–30. Cambridge: Cambridge University Press.
- Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe — Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36(85). 15–43.

- König, Ekkehard & Johan van der Auwera. 1988. Clause integration in German and Dutch conditionals, concessive conditionals, and concessives. In John Haiman & Sandra A. Thompson (eds.), *Clause combining in grammar and discourse*, 101–133. Amsterdam & Philadelphia: John Benjamins.
- Labov, William. 1994. *Principles of linguistic change* (I: Internal Factors). Oxford: Blackwell.
- Lehmann, Christian. 1991. Grammaticalization and related changes in contemporary German. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to grammaticalization, vol. II*, 493–535. Amsterdam & Philadelphia: John Benjamins.
- Lewis, Diane. 2005. Corpus comparables et analyse contrastive: L'apport d'un corpus français/anglais de discours politiques à l'analyse des connecteurs adversatifs. In Geoffrey Williams (ed.), *La linguistique de corpus*, 179–190. Rennes: Presses Universitaires de Rennes.
- Lindquist, Hans & Christian Mair (eds.). 2004. *Corpus approaches to grammaticalization in English*. Amsterdam & Philadelphia: John Benjamins.
- Lindström, Jan & Camilla Wide. 2005. Tracing the origins of a set of discourse particles: Swedish particles of the type *you know*. *Journal of Historical Pragmatics* 6(2). 211–236.
- Mair, Christian. 2004. Corpus linguistics and grammaticalisation theory: Statistics, frequencies, and beyond. In Hans Lindquist & Christian Mair (eds.), *Corpus approaches to grammaticalization in English*, 121–150. Amsterdam & Philadelphia: John Benjamins.
- Molencki, R. 2007. The evolution of *since* in medieval English. In Ursula Lenker & Anneli Meurman-Solin (eds.), *Connectives in the history of English*, 97–113. Amsterdam & Philadelphia: John Benjamins.
- Mortier, Liesbeth. 2007. *Perspectives on grammaticalization and speakers' involvement: The case of progressive and continuative periphrases in French and Dutch*. Leuven: Catholic University of Leuven dissertation.
- Mortier, Liesbeth & Liesbeth Degand. 2009. Adversative discourse markers in contrast; the need for a combined corpus-approach. *International Journal of Corpus Linguistics* 14(3). 338–366.
- Noël, Dirk. 2003. Translations as evidence for semantics: an illustration. *Linguistics* 41(4). 757–785.
- Onodera, Noriko. 2000. Development of *demo* type connectives and *na* elements: Two extremes of Japanese discourse markers. *Journal of Historical Pragmatics* 1(1). 27–55.
- Onodera, Noriko O. 2004. *Japanese discourse markers: Synchronic and diachronic discourse analysis*. Amsterdam & Philadelphia: John Benjamins.
- Oostdijk, Nelleke. 2000. The spoken Dutch corpus project. *The ELRA Newsletter* 5(2). 4–8.
- Pander Maat, Henk & Liesbeth Degand. 2001. Scaling causal relations and connectives in terms of speaker involvement. *Cognitive Linguistics* 12(3). 211–245.
- Pander Maat, Henk & Ted Sanders. 2001. Subjectivity in causal connectives: An empirical study in language use. *Cognitive Linguistics* 12(3). 247–273.
- Pit, Mirna. 2003. *How to express yourself with a causal connective. Subjectivity and causal connectives in Dutch, German and French*. Amsterdam & New York: Rodopi.
- Partington, Alan. 2006. Metaphors, motifs and similes across discourse types: Corpus-assisted discourse studies (CADS) at work. In Anatol Stefanowitsch & Stefan Th. Gries (eds.), *Corpus-based approaches to metaphor and metonymy*, 267–304. Berlin & New York: Mouton de Gruyter.
- Prévost, Sophie. 1999. *Aussi* en position initiale: évolution sémantico-syntaxique du 12e au 16e siècle. *Verbum* 21(3). 351–380.
- Prévost, Sophie. 2003. *Quant à*: analyse pragmatique de l'évolution diachronique (14e–16e siècles). In Bernard Combettes, Anne Theissen & Catherine Schmedecker (eds.), *Actes du colloque 'Ordre et distinction dans la langue et le discours'*, Metz 99, 443–459. Paris: Honoré Champion.
- Prévost, Sophie. 2007. *À propos de, à ce propos, à propos*: Évolution du 14e au 16e siècle. *Langue Française* 156. 108–126.

- Rissanen, Matti, Merja Kytö & Kirsi Heikkonen (eds.). 1997. *Grammaticalization at work: Studies of long-term developments in English*. Berlin & New York: Mouton de Gruyter.
- Schoonenboom, Judith. 2000. *Analyse, norm en gebruik als factoren van taalverandering: Een studie naar veranderingen in het Nederlands onzijdig relativum* [Analysis, norm and use as factors in language change: A study into changes in the Dutch neuter relative]. Amsterdam: University of Amsterdam dissertation.
- Simon, Anne-Marie C. & Liesbeth Degand. 2007. Connecteurs de causalité, implication du locuteur et profils prosodiques. Le cas de *car* et de *parce que*. *Journal of French Language Studies* 17. 323–341.
- Spooren, Wilbert & Liesbeth Degand (2010). Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory* 6(2). 241–266.
- Spooren, Wilbert, Ted Sanders, Mike Huiskes & Liesbeth Degand (in press). Subjectivity and causality: A corpus study of spoken language. In Sally Rice & John Newman (eds.), *Empirical and experimental methods in cognitive/functional research*, 241–255. Stanford: CSLI Publications.
- Stefanowitsch, Anatol. 2006. Corpus-based approaches to metaphor and metonymy. In Anatol Stefanowitsch & Stefan Th. Gries (eds.), *Corpus-based approaches to metaphor and metonymy*, 1–16. Berlin & New York: Mouton de Gruyter.
- Sweetser, Eve E. 1990. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.
- Traugott, Elizabeth C. 1995. Subjectification in grammaticalisation. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, 31–54. Cambridge: Cambridge University Press.
- Traugott, Elizabeth C. (2010). (Inter)subjectivity and (inter)subjectification: a reassessment. In Kristin Davidse, Lieve Vandelanotte, Hubert Cuyckens (eds.), *Subjectification, intersubjectification and grammaticalization*, 29–71. Berlin & New York: de Gruyter Mouton.
- Traugott, Elizabeth C. & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Van der Sijs, N. 2008a. *Verantwoording van de digitale uitgave van de Delfse bijbel uit 1477* [Account of the digital edition of the 1477 Delft Bible]. [http://www.bijbelsdigitaal.nl/info/DB1477\\_Verantwoording.pdf](http://www.bijbelsdigitaal.nl/info/DB1477_Verantwoording.pdf) (accessed 21 January 2010).
- Van der Sijs, N. 2008b. *Verantwoording van de digitale uitgave van de Statenvertaling 1637* [Account of the digital edition of the 1637 Statenvertaling]. [http://www.bijbelsdigitaal.nl/info/SV1637\\_Verantwoording.pdf](http://www.bijbelsdigitaal.nl/info/SV1637_Verantwoording.pdf) (accessed 21 January 2010).
- Vincent, Diane. 2005. The journey of non-standard discourse markers in Quebec French: networks based on exemplification. *Journal of Historical Pragmatics* 6(2). 188–210.
- Vogl, Ulrike. 2007. Het belang van conditionaliteit voor de ontwikkeling van temporeel naar causaal voegwoord. De geschiedenis van *dewijl*, *terwijl*, *weil* en *while* [The importance of conditionality in the development from temporal to causal complementizer. The history of *dewijl*, *terwijl*, *weil* and *while*]. *Nederlandse Taalkunde* 12(1). 2–24.
- Zwarts, Joost. 2006. *Om* en *rond*: een semantische vergelijking [Dutch *om* ‘about’ and *rond* ‘about’: a semantic comparison]. *Nederlandse Taalkunde* 11(2). 101–123.