Ted J. M. Sanders* and Wilbert P. M. Spooren

Causality and subjectivity in discourse: The meaning and use of causal connectives in spontaneous conversation, chat interactions and written text

Abstract: Many languages of the world have connectives to express causal relations at the discourse level. Often, language users systematically prefer one lexical item (*because*) over another (even highly similar) one (*since*) to express a causal relationship. Such choices provide a window on speakers' cognitive categorizations, and have been modeled in previous work in terms of subjectivity. However, a broader empirical basis and a more specific operationalization of subjectivity are urgently needed. This paper provides in these needs by developing an integrative empirical approach to the analysis of the Dutch connectives *omdat* 'because' and *want* 'since/for' in written text, conversation, and chat interactions. These can be considered a case in point for linguistic categorization since related European languages show similar distinctions. The construct of subjectivity is decomposed into characteristics like type of relation and s*ubject of consciousness* (who can be considered responsible for the causality?). The use of statistical methods specifically suitable for hypothesis testing in natural language corpora produces results that provide new insights into the division of labor between the two connectives, as well as into the notion of subjectivity.

Keywords: causality, connectives, subjectivity, Dutch, discourse

*Corresponding author: Ted J. M. Sanders:** Utrecht Institute of Linguistics, Universiteit Utrecht, Trans 10, NL-3512 JK Utrecht, The Netherlands. E-mail: t.j.m.sanders@uu.nl
Wilbert P. M. Spooren: Discourse studies, Radboud Universiteit Nijmegen, Erasmusplein 1, NL-6525 HT Nijmegen, The Netherlands. E-mail: w.spooren@let.ru.nl

1 Introduction

1.1 Discourse, causality, and connectives

People use language to communicate in various contexts and in various media, be it in spontaneous conversations, in writing and reading texts, or in chat interactions. Discourse is a crucial level in all types of human linguistic communication; it is impossible to communicate without understanding the coherence between utterances. One type of coherence relation language users often want to express is causality, for instance in the case of a *reason* or a *consequence-cause* relation (see (1)) between events in the world or by connecting a claim and an argument (see (2)). In English, both relations can be made explicit with the connective *because*.

(1) *The fields are wet because it has rained a lot this week.*

(2) *Surely all soccer games will be cancelled, because it has rained a lot this week.*

In this paper, we focus on this backward causality – that is, the order "S1, CONNECTIVE S2", where S stands for discourse segment, which is minimally a clause.

Many languages of the world have connectives to express causal relations at the discourse level (see Diessel and Hetterle 2011, who analyzed causal clauses in 60 languages from typologically different language families). Speakers of English, for example, can choose between *because* and *since* or *for*. We are interested in the system behind the meaning and use of such connectives. Also, we ask how different these choices in English are from the ones made by speakers of other languages, such as Dutch *omdat* versus *want*, German *weil* versus *denn*, and French *parce que* versus *car*. It seems as if language users often systematically prefer one lexical item over another (even highly similar) one to express a certain type of causal relationship. Systematic use of a particular lexical item to express a certain type of causal relationship implies that people distinguish between several types of causality. Hence, such choices could provide a window on speakers' cognitive categorizations of causality. The linguistic study of the meaning and use of causal connectives may reveal insights into human categorization of causality (see, among others, Sanders and Sweetser, 2009). In other domains, like the study of metaphor or that of causative verbs, similar studies of linguistic categories apparent in people's everyday language use have already produced many interesting insights into the working of the mind (see, for instance, Lakoff 1987; Verhagen and Kemmer 1997).

In her seminal work, Sweetser (1990) presents a domain approach in which she argues that a conjunction like English *because* is used in the content-domain when one event causes another in the real world (3). Epistemic use (4) concerns the speaker's reasoning and (5) illustrates the speech act use.

(3) *John came back because he loved her*. (i.e., the loving caused the return)

(4) *John loved her, because he came back*. (i.e., the observation that he came back is an argument for the claim that the loved her)

(5) *What are you doing tonight, because there's a good movie on*. (i.e., I invite you to come and I give a reason for performing this speech act).

In more recent years, we have seen related proposals in which distinctions like content, epistemic and speech act domains are described in terms of the subjectivity of *speaker involvement* (Pander Maat and Degand 2001; Pander Maat and Sanders 2000, 2001). In such an approach, content relations such as Cause-Consequence are objective (because the Speaker is not involved), whereas epistemic and speech act relations are subjective (because the Speaker is clearly involved) (see Section 1.2 for a further elaboration).

The distinction between, on the one hand, coherence between events in the world – named objective, semantic, propositional, internal, or content relations – and on the other hand coherence realized by the communicative acts or reasoning of the speaker – subjective, pragmatic, external relations – can be found in virtually all taxonomies and categorizations of coherence relations (Kehler 2002; Knott and Dale 1994; Mann and Thompson 1988; Martin 1992; Sanders et al. 1992; Sanders 1997). In addition, crosslinguistic studies suggest similar distinctions are useful to describe the organization of the lexicon of causal connectives in languages like Dutch, German, and French (Pit 2006; Evers-Vermeul et al. 2011). These languages show a more differentiated repertoire of connectives to express backward causality than English, where *because* can be used across the three domains (Ford 1993; Knott and Sanders 1998; Sweetser 1990). Dutch *want* can be used to express speech act and epistemic relations and is therefore considered more subjective than *omdat* (Degand 2001; Pit 2006; Verhagen 2005). Similarly, French *car* and *puisque* would be specifically used in the epistemic domain, whereas *parce que* sticks to content relations (*Groupe λ* 1975; Anscombre and Ducrot 1983; Degand and Pander Maat 2003; Zufferey 2010). Similarly, several German linguists have suggested that *denn* can only be used to express epistemic relations (Pasch 1983; Günthner 1993; Keller 1995, but see also Wegener 2000).

Since the mid 1990s we have witnessed a rise in corpus studies to investigate these and related ideas about the organization of the lexicon of connectives in

several languages, seeking to find the system behind the meaning and use of (causal) connectives, which has lead to an empirical test in actual language use of these challenging theories and hypotheses (see Sanders and Spooren 2009 and other contributions to Sanders and Sweetser 2009).

1.2 Subjectivity as categorization principle: inherent characteristic or context-dependent?

The corpus studies mentioned above have marked an important step forward in the field. However, there are fundamental challenges left, both of a theoretical and an empirical nature. The first issue concerns the notion of subjectivity, a term that is often used for different phenomena, which can be a source of confusion (Nuyts 2012). The second concerns the question whether the semantic profiles of causal connectives can actually be characterized in terms of subjectivity as a stable characteristic of their semantic profile, or whether this subjectivity is a context-dependent characteristic. A third issue concerns the empirical basis of the corpus studies, which is insufficient. We aim to address these three issues in this paper, in which we propose an integrative empirical approach to subjectivity in discourse. Below, we first address three and two, to end up with the first issue: the operationalization of subjectivity.

1.2.1 The empirical basis of current corpus studies

The empirical domain of the corpus studies that have been conducted until now is limited in many respects. First of all, there is a substantial amount of work on German (e.g., Günthner 1993; Keller 1995) and French (Anscombre and Ducrot 1983; Groupe λ 1975) causal connectives, some of which specifically investigates spontaneous conversation, but these studies have a limited empirical basis: they analyze only small amounts of cases, and statistical evaluation is often lacking. For Dutch, some recent studies providing such evaluations are available but these studies are dominated by analyses of written text (e.g., Pander Maat and Degand 2001; Pit 2006; Stukker and Sanders 2012); in fact, the only study on non-written media is a small pilot-study in which we compared spontaneous conversation and chat with written text (Spooren et al. 2010).

There is a certain urgency to add other than written data to the empirical foundation of theories on the categorization of connectives. Several studies of spontaneous conversations suggest a typical usage pattern of causal connectives in conversations. Günthner (1993) and Keller (1995) demonstrated that German

weil can express epistemic relations in spontaneous conversations, whereas in written language it seems to be reserved for the content domain. Recently, the work on German causal connectives has grown considerably, both in terms of quantity and in terms of empirical base; recent extensive and statistically evaluated corpus studies include Frohning (2007), Breindl and Walter (2009) and Volodina (2011a). In some of these studies insights from spoken language data have shown the profiles of causal connectives to be less specialized than was concluded on the basis of the analysis of only written texts (Breindl and Walter 2009). For French, Zufferey (2010, 2012) concludes that *puisque* has a strong preference for epistemic use in telephone conversations. Such results show that written language as the basis for analysis may lead to a distorted picture. This observation is the basis for addressing our second issue: how stable or context-dependent are characterizations of connectives in terms of subjectivity.

A principled point is that written language deviates from the prototypical communicative situation that spontaneous conversations provide in several respects (Clark 1996): Written language generally shows a large distance between authors and readers. In written language the communicators are relatively invisible. As a consequence, written language is detached from the deictic center of communication (Sanders et al. 2009). Authors can (re)consider and revise their lexical choices and formulations; the focus is usually on content and not on interpersonal issues. And whereas spoken language is fragmented, written language is integrated (Chafe 1994). These considerations lead to the conclusion that the use of causal connectives should be investigated systematically in different media. Such investigations are scarce (but see Zufferey 2010, 2012).

In this paper we present a systematic study of the use of *want* and *omdat* as a case in point of how European languages encode backward causal relations that differ in subjectivity. Our research question is: do *want* and *omdat* have a clear semantic profile that is constant across media, or do they have a vague profile and is their use mainly determined by the context in which they appear? The answer to this question is not obvious, given the limited empirical basis for Dutch. And it is not improbable that context plays a great role, given the frequency data presented in Table 1. This table shows strong differences in frequencies between *want* and *omdat* across the media in which they occur.

Why is that? Do *want* and *omdat* indeed have clear semantic profiles and do language users express different causal relations in different media? Or are the semantic profiles in fact not so clear, and do people use *want* and *omdat* in a less systematical way to simply express all kinds of causals? Is *omdat* relatively frequent in newspapers because it is a subordinating conjunction and thus supports the integrative nature of written language? Conversely, is *want* relatively frequent in chat not so much because speakers in chat express different types of causal

Table 1: Relative frequency of *omdat* and *want* (per million words).

Medium	Connective	
	Omdat	Want
Newspaper[a]	920	660
Spoken Discourse[b]	521	1640
Chat interaction[c]	445	1032

[a] based on pilot version of D-Coi; [b] based on Corpus of Spoken Dutch; [c] based on VU Chat corpus.

relations than in newspapers, but rather because the coordinating conjunction *want* fits in with the fragmented nature of chat language? In this paper we investigate whether there is a systematical relationship between the semantic profile of the connectives and the medium in which they occur.

1.2.2 The analysis of subjectivity in natural discourse: towards an integrative approach

The final fundamental issue that needs to be addressed is the operationalization of subjectivity. Broadly speaking, three fundamental approaches to subjectivity can be distinguished (Nuyts 2012): those by Lyons (1977), by Traugott (1995) and by Langacker (1990).[1] The three approaches highlight different aspects of the complex notion of subjectivity. Our aim is not to present a totally new approach but rather to combine crucial aspects of all three approaches, in order to operationalize subjectivity as a discourse phenomenon. In fact, we make use of the possibility that the different notions of subjectivity are to some extent "co-applicable" (Nuyts 2012: 22). Subsequently, we will test this integrative approach to subjectivity in an empirical way.

Reference to the speaker is generally considered a core component of linguistic subjectivity: "[Subjectivity is] the property (or set of properties) of being either a subject of consciousness (i.e., of cognition, feelings, and perception) or a subject of action (an agent). It denotes the property of being what Descartes called 'a thinking entity'" (Lyons 1995: 337). Traugott (1995: 31) defines subjectification as the process through which "meanings become increasingly based in the speak-

---

1 We are aware that these notions were intended to refer to different phenomena and show differences (Nuyts 2012). At the same time, we think that the basic insights represented in these approaches can be combined to develop a valid account of subjectivity in discourse connectives.

er's subjective belief state/attitude toward the proposition".[2] Consequently, an utterance is subjective if it requires reference to the speaker in its interpretation, and objective if it does not. In this paper we follow this characterization of subjectivity, as we will argue below. Additionally, we use Langacker's (1990) insight that subjectivity is defined by the way in which an entity is construed. It is construed with maximal subjectivity when it remains implicit and "off stage", and with maximal objectivity when it is put onstage as an explicit focus of attention. Crucial to this aspect of subjectivity is the way in which the speaker figures in the form of the utterance: explicit reference is objective, whereas an implicit reference is subjective. This implies that Langacker's categories of "subjective" and "objective" are both speaker-related, and therefore "subjective" in Traugott's terms (De Smet and Verstraete 2006: 369; see also Vis 2011; Nuyts 2012).

We consider both speaker-relatedness (Lyons; Traugott) and implicit presence of the speaker (Langacker) as important aspects of subjectivity. Furthermore, we draw attention to the distinction between speaker-subjectivity and character-subjectivity, as discussed for example by Sanders et al. (2012). Finally, we also include the nature of the causal relation itself as a characterization of subjectivity, since our topic of research is causal coherence relations in discourse.

In line with earlier work on causal connectives (Pander Maat and Sanders 2000), we define an utterance as subjective when its interpretation requires an active *Subject of Consciousness* (from now on SoC). A SoC crucially involves an animate subject, a person, whose intentionality is conceptualized as the ultimate source of reasoning, evaluating, describing, or acting "in the real-world". An utterance is subjective because there is some thinking entity in the discourse who evaluates. For instance, *He thought Paris was nice* is subjective because it involves an evaluation by a character in the discourse. Compare this with an utterance like *Paris is in France*, which is presented as a fact in the world that does not depend on the evaluation by a SoC. To be more precise, in the utterance *He thought Paris was nice* the validity of the proposition "Paris is nice" depends on the SoC *He*, whereas in the utterance *Paris is in France* the proposition "Paris is in France" can be verified directly in the non-linguistic reality.

Obviously, each utterance in a discourse comes from a speaker or author, and therefore each utterance is dependent on a SoC. However, in some utterances, the

---

2 This definition differs from the definition that Nuyts (2012) gives of subjectivity, a difference that is also noted by Nuyts, who considers an utterance like *They may have well left already* not subjective, but neutral, as it does not contain an explicit reference to the first person "assessor". For Traugott, such an utterance is indeed subjective.

SoC is manifest because the sequence cannot be interpreted without reference to a SoC. Such cases – typically feelings and evaluations of all kinds – are considered subjective; they simply cannot be interpreted without making reference to the SoC, her thoughts and feelings. In contrast, utterances that do not depend on such a manifest reference to the SoC are considered objective.

Authors/speakers can be SoC's, but characters can also function as such. The author/speaker is the first voice in the discourse who has constant access to her feelings and thoughts. She does not have access to the feelings and thoughts of a third person. As a result, *I think Paris nice* can be a direct report of an inner feeling, whereas *he thinks Paris nice* is a description of an evaluation. It follows that first person evaluations are more subjective than third person evaluations. The difference, then, between the speaker/writer versus a character as SoC is that the first type concerns a first voice, which is grounded in the Deictic Center of Communication (Sanders et al. 2009). This resembles Traugott's (1989, 1995) view on subjectivity as closeness to the communicative "here and now": the speaker here and now asserts that a particular state of affairs holds. By contrast, the character type concerns a third person in the discourse, which is more distant from the Deictic Center of Communication.

The examples mentioned so far, be it first person SoC's or third person SoC's, are descriptions of evaluations and consequently they are more or less objective. In terms of Sweetser this type of subjectivity may still be in the content domain. Yet, evaluations are often much more implicit. Especially when the speaker/author is first person SoC and the evaluation concerns the here and now, spontaneous evaluations typically are of the type *Paris is great*, i.e., a first person SoC. Such utterances express an evaluation and the SoC remains implicit. Indeed these are the most subjective type of utterances: those in which the speaker is SoC in first person, but remains off stage (Langacker 1990).

In sum, central to our integrative approach is that we consider utterances subjective when they cannot be interpreted without reference to a SoC; the SoC's thoughts, feelings, point-of-view are simply necessary for interpretation. This approach acknowledges that subjectivity ultimately is a cognitive notion. Although it can be signaled through linguistic expressions – such as explicit reference to the speaker / SOC, modal expressions, evaluative verbs, scalar predicates –, the Subjectivity of an utterance does not *depend on* the occurrence of such signals. It is typical for our take on Subjectivity to consider the author/speaker as the default SoC, who will often remain implicit and off stage, often producing subjective utterances.

In our analysis so far, we have presented subjectivity as a property of utterances. However, this is not enough, as subjectivity can also reside in the nature of relations between utterances, as the examples in (3–5), repeated below for con-

venience, have demonstrated: the link can be of a content, epistemic or speech act type. In fact, it is this kind of subjectivity that is the main focus of our interest (Sanders et al. 1992). Causal links in the content domain (example 3) are objective. Epistemic causality is inherently subjective because the speaker is actively reasoning towards a conclusion or concluding something on the basis of an observation (example 4) in the here-and-now. Cases of speech act causality (example 5) are also subjective: the speaker is performing a speech act and motivating that act on the basis of an observation.

(3) *John came back because he loved her.*

(4) *John loved her, because he came back.*

(5) *What are you doing tonight, because there's a good movie on.*

In our model of analysis we take the relational nature of subjectivity into account in two ways: we distinguish between different types of causal relation, and we specify the SoC that is responsible for the causal link. Example (3) is a content relation, with a third person SoC, *John*. Examples (4) and (5) are of an epistemic and speech act nature, respectively, with the speaker as the SoC. Note that this does not mean that there is a perfect correlation between relation type and SoC. Consider example (6), based on the epistemic relation in example (2).

(6) *That Saturday morning, Willem was sad. [S1 Now all soccer games would be cancelled], because [S2 it had rained a lot that week].*

Example (6) is an epistemic relation: the SoC concludes on the basis of an argument that the games will be cancelled. What is special about this example is that the SoC is not the speaker (as in examples (2) and (4)), but a third person (Willem). This, then, is a case of free indirect speech ("an unspeakable sentence," Banfield 1982; Sanders 2010) and shows that epistemic relations can occur in a third person SoC context. Authors like Banfield (1982) and Schlenker (2004) have argued that complex cases like Free Indirect Speech and the Historical Present constitute a challenge for the linguistic analysis of subjectivity. Our description of example (6) shows the potential of our approach for such cases.

1.2.3 An empirical approach

To summarize the discussion so far, we decompose the complex construct of subjectivity in terms of four characteristics of causal connections. In answer to

our research question about the semantic profile of Dutch causal connections, we investigate whether and to what extent these characteristics co-occur. To that end, we make use of statistical methods specifically suitable for hypothesis testing in natural language corpora. Such methods allow us to test whether these characteristics help to distinguish between two Dutch backward causal connectives, *omdat* and *want,* and to what extent the differences between these two connectives are determined by the context in which they occur. Thus we follow Gries (2012), who strongly advocates using corpus data in developing "a psycholinguistically informed, (cognitively-inspired) usage-based linguistics which should be located, firmly and deliberately, in the social/behavioral sciences".

The integrative empirical approach that we develop provides us with new insights into the notion of subjectivity. On top of that, we advance the research into linguistic categorization by a detailed and rigorous empirical study of a relatively large corpus of naturally occurring language from various media. The detailed study of Dutch *want* and *omdat* can be considered a case in point for linguistic categorization since related European languages show similar distinctions.

1.3 Backward causality in Dutch

Causality can be expressed using backward and forward causal connectives. In a forward causal construction the first segment introduces a cause or an argument, and the second segment expresses a consequence or a claim. In backward constructions, the first segment expresses a claim or a consequence, and the second segment expresses the argument or the cause. In backward constructions, the connective typically occurs at the beginning of the second segment. In Dutch the connective signaling a backward causal relation can be a coordinating conjunction (like *want*) or a subordinating conjunction (like *omdat, aangezien,* or *doordat*). The most frequently used causal connectives are *want* and *omdat*.[3] Usually,

---

3 This paper focuses on backward causality only. The reason is that forward causal connectives are treated in other work (cf. Stukker and Sanders 2012; Stukker, Sanders and Verhagen 2008). The reason that we confine ourselves to *want* and *omdat* is that these are by far the most frequent causal connectives in the Dutch language. For example, the connective *aangezien* occurs in our newspaper corpus only 59 times per million words, and most of these occurrences are forward causal connectives. The connective *doordat* is mostly used as a backward causal connective, but it occurs only 98 times per million words in our newspaper corpus.

the subordinator *omdat* is translated in English as *because*, whereas the coordinator *want* is often translated as *since* or *for*. Both translations of *want* seem too "formal" for *want*, which is very frequently used in relatively informal contexts, especially in spoken discourse, see table 1. The prototypical use of the Dutch backward causal connectives *want* and *omdat* can be illustrated by translating the English examples used so far.

(1) D *De velden zijn nat omdat het veel geregend heeft deze week.*
         'The fields are wet OMDAT it has rained a lot this week'

(2) D *De voetbalwedstrijden worden vast afgelast, want het heeft deze week erg veel geregend.*
         'Surely all soccer games will be cancelled, WANT it has rained a lot this week'

(3) D *Jan kwam terug omdat hij van haar hield.*
         'Jan came back OMDAT he loved her'

(4) D *Jan hield van haar, want hij kwam terug.*
         'Jan loved her, WANT he came back'

(5) D *Wat doe jij vanavond, want er draait een goede film.*
         'What are you doing tonight, WANT there's a good movie on'

These examples illustrate how *want* 'for/since' is typically used to express epistemic and speech act relations, whereas *omdat* 'because' is typically used to express content relations. More specifically, *omdat* has a preference for *volitional content* relations, in which the intentions of a human actor propagate the actions. Several studies have shown that these characteristics are robust, and vary from strong preferences to clear restrictions on the relations they can express. Taken together, these observations show how the Dutch language "cuts up" backward causality (Degand 2001; Degand and Pander Maat 2003; Pit 2006).

The clearest case of this "cutting up" concerns a specific connective for non-volitional content relations: *doordat* ("as a consequence of the fact that"), see (7).

(7) *De temperatuur steeg, doordat de zon scheen.*
     'The temperature rose, DOORDAT the sun was shining'

There are clear restrictions on its use: it only expresses non-volitional content relations. In fact, Dutch *doordat* can never be used to express the relations (3)–(5).

Other divisions of labor are tendencies rather than clear-cut restrictions. For example, the relation in (3) can also be expressed by *want*, which gives the sequence a more epistemic flavor. And in an example like (8) *omdat* is used in an epistemic context (although this use requires a pause before *omdat*; Huiskes 2010; Persoon et al. 2010).

(8) *Het moet wel een slechtvalk zijn, omdat hij met een enorme snelheid omlaag dook.*
    'It must be a peregrine falcon, OMDAT it dove downwards with an enormous speed.'

Corpus studies also indicate that *want* regularly expresses volitional content relations, whereas *omdat* can express epistemic relations in a minority of cases (Bekker 2006; Degand 2001; Pit 2006). Hence, volitional content and epistemic relations are regularly lexicalized by the same connectives, an observation which can be taken as an argument against a strict domain-specific hypothesis, in which each connective correlates with a specific domain (content, epistemic, speech act). The study reported in this paper will shed new light on this discussion.

　　Before we move on to summarize the main research questions, it is important to elaborate somewhat on the syntactic differences between *want* (a coordinating conjunction) and *omdat* (a subordinating conjunction). It is known from the literature that there is a correlation between grammatical status and discourse function. For example, coordination constructions are syntactically independent and hence are better suited to express complete speech acts. Subordinating constructions are integrated, and therefore a less plausible site for expressing complete illocutions (van Dijk 1979; Verhagen 2005; Volodina 2011b). This raises the issue to what extent the grammatical difference between *want* and *omdat* is confounded with a difference in subjectivity. Admittedly, there is a strong correlation between syntactic integration and the tendency to express objective relations. Nevertheless, there are good reasons to assume that a difference in subjectivity between connectives cannot be reduced to a difference in their grammatical status. One reason is that despite the syntactic differences, it is not impossible for *want* to express objective relations (as will be shown in the results section), and similarly we find non-content subjective uses of *omdat* (cf. the use of *omdat* in spontaneous conversations to express subjective epistemic relations; Huiskes 2010; Persoon et al. 2010). A final reason is that in Dutch the subjectivity distinction is also relevant for categorizing connectives that do not differ grammatically, notably in the forward causal domain, where we have adverbial *dus* (*so*, *therefore*) preferring subjective epistemic and speech act rela-

tions, and the adverbial *daarom* preferring objective volitional content relations (Pander Maat and Sanders 2001; Pander Maat and Degand 2001; Stukker et al. 2008).

1.4 Research questions and hypotheses

If subjectivity is the right notion to analyze the difference between these connectives, it should go across the modalities of written, spoken and chat language. Therefore our main hypothesis is that *want* occurs in more subjective contexts than *omdat*, irrespective of the medium.

Generalization over media
Hypothesis 1: Across all media, *want* is used more often to express subjective relations (epistemic, speech act) than *omdat*.
Hypothesis 2: Across all media, *want* is used more often to support a judgment than *omdat*.
Hypothesis 3: Across all media, *want* is used more often with first and second person SoC's than *omdat*.
Hypothesis 4: Across all media, *want* is used more often with an implicit SoC than *omdat*.

In addition to these specific hypotheses, we formulate two explorative research questions. The first relates to differences between media. It is an open question whether the medium affects the degree of subjectivity, although one might argue that in relatively spontaneous media (spoken conversation, chat interaction) the Deictic Center of Communication is more salient than in relatively detached media (written text): a communicative situation of direct speaker-hearer interaction (spoken, chat) can be expected to be more subjectively grounded than written communication, because of the direct availability of Speaker and Addressee. The second question relates to the strength of the various subjectivity characteristics: are all characteristics of subjectivity (as formulated in the hypotheses) equally important in predicting the choice speakers or writers make between *want* and *omdat*?

In the following we further develop our integrative empirical approach to test these hypotheses and answer these research questions. By doing so, we tackle the two issues introduced in Section 1.2 – the need for broadening the empirical domain and the operationalization of subjectivity.

2 Method

2.1 Model of analysis

Below we present the specific discourse characteristics that were analyzed. Illustrative examples are taken from our corpora.

I. Propositional attitude of the first segment (S1)

As we argued above, evaluations are central for linguistic subjectivity. A prominent way in which evaluations manifest themselves in discourse is in the form of judgments. Consequently, each segment was analyzed as expressing either a judgment or another propositional attitude (fact, general knowledge, an intentional act, individual knowledge, a perception, an experience). Our analysis focuses on the first segment, because that is the site where relational subjectivity is most manifest in backward causals: for example, in a Claim-Argument relation the argument can be very factual and objective.

A segment expresses a judgment if it presents or implies a SoC – the person responsible for the causal relation; the Subject of Consciousness – and expresses what is judged. The segment expresses a state and uses a so-called scalar predicate (a predicate that can be modified with degree expressions, such as *very much X*; *more than X*), which is a judgment because it can be paraphrased with "I believe/feel that …". Fragment (9) gives an example.[4]

(9) Judgment in S1
    A *[S1 en ik vind 't niet meer leuk op die manier te*
          and I find it not more nice on that way to
        werken ook dat nog 'ns een keer.]
        work even that again
        'and I don't like it any longer to work that way also'
    B *nee ja.*
        'no yes'
    A *hè?*
        'right?'

---

4 In our presentation we use the following conventions: first and second segments in the relation are delimited by [S1] and [S2]; the interlocutors are indicated by capitals (A, B etc.); the English translations of the Dutch examples are rather idiomatic translations unless a more precise gloss is needed. In the translation the connective is indicated in capitals.

> A   *omdat*   *[S2* *er*   *geen*   *uh* *geen*   *wisselwerking*   *is.]*
>   because   there no   eh   no   interaction   is
>   'OMDAT there is eh no interaction'

Judgments were considered more subjective than other modalities.

II. Relation type

The causal relation expressed in each fragment was analyzed in terms of domains (Sweetser 1990): content (in which the speaker describes a causal relation in the world), epistemic (in which the speaker infers a conclusion on the basis of an argument) and speech act relations (in which the speaker motivates a speech act). Furthermore, within the content relations we distinguish between volitional and non-volitional relations (see Stukker et al. 2008): Does the relation involve an intentional act or not? Examples are:

(10) Non-volitional content
    [S1 De vogelstand gaat hard achteruit] omdat [S2 een hele voedselketen stelselmatig wordt vergiftigd].
    'The bird population decreases fast OMDAT a complete food chain is poisoned systematically.'

(11) Volitional content
    A *[S1 dan gingen ze Albert-Jan vragen of ie de achtste in de boot kon zijn] omdat [S2 Bas naar die inauguratie van die pastoor moest].*
       'then they went to ask Albert-Jan whether he could be the eighth man in the boat OMDAT Bas had to go that priest's inauguration.'

(12) Epistemic
    A *[S1 en en wa ik het nu heb dat is geen noodsituatie] want [S2 ik kan donders goed inschatten kwart over vier half vijf dat er dan geen student meer boven gaat kijken.]*
       'and and what I have it now that is not an emergency WANT I can estimate very well quarter past four half past four that no student will go and look upstairs then anymore.'

(13) Speech act
    A *[S1 en uh a als iemand mij belt ja dan ben ik er niet] want [S2 ik ben bezig met dit werk en dat moet vandaag af punt].*
       'and uh i if someone calls me yes then I am not in WANT I am busy working at this and it has to be finished today period.'

On the basis of our reasoning in Section 1.2, the causal relations can be ordered from least subjective to most subjective, as follows:

Non-volitional content < Volitional content < Epistemic / Speech act

III. Type of SoC in the first segment (S1)

The SoC is the person responsible for the causal relation that is constructed. There can be either no SoC (as in example (10)), or the SoC is a third person (example (11)), a second person (example (14)) or a first person (examples (12) and (13)).

(14) Second person SoC
    Speaker A *en dat is de enige manier via mij krijgen ze hun boeken.*
                'and that is the only way through me they get their books'
    Speaker B *ja.*
                'yes'
    Speaker A *ik ben de leverancier als het ware.*
                'I am the supplier so to speak'
    Speaker A *ja.*
                'yes'
    Speaker B *ja ja ja ja ja.*
                'yes yes yes yes yes'
    Speaker B *uh koopt u dan ook alleen maar gebonden uitgaven OMDAT dat mooier is in de boekenkast of ...*
                'eh does that mean that you only buy hardcover editions BECAUSE that is more beautiful on the bookshelves or ...'
    Speaker A *nee.*
                'no'
    Speaker A *neen niet altijd niet altijd.*
                'no not always not always'

These options can be ordered in degree of subjectivity, as follows:

No SoC < Third person < Second person, First person

In the analysis presented below we only use the distinction between third person SoCs and first/second person SoCs. Cases without a SoC were deemed irrelevant as these express facts. Cases of first and second person SoCs were collapsed. We did not distinguish between 1st and 2nd person SoCs, because both introduce

subjectivity in the *hic et nunc* of the speech/writing situation (speaker/writer subjectivity in case of 1st person SoCs and addressee subjectivity in case of 2nd person SoCs).

IV. Linguistic realization of the SoC

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

Explicit reference to the SoC < Implicit reference to the SoC

2.2 Materials

For our analysis we used three corpora. For the written medium we made use of the pilot version of the D-COI corpus, a preparatory project which aimed at producing a blueprint and the tools needed for the construction of a 500-million-word reference corpus of contemporary written Dutch (D-COI 2006). The size of the corpus part that we used was 1,8 million words. We randomly selected 100 occurrences of *omdat* and 100 occurrences of *want*.[5] For the spoken medium we made use of the Corpus of Spoken Dutch (Corpus Gesproken Nederlands, CGN). CGN is a 10 million words corpus of completely digitalized material, annotated in several ways (Oostdijk 2000). From the spontaneous conversations and interviews in this corpus we randomly selected 100 fragments with *want* and 100 fragments with *omdat*. For the chat medium we have used the VU Chat corpus, a small corpus of chat conversations between secondary school children, collected at VU University Amsterdam. The size of this corpus is 217,000 words. From this corpus we selected all occurrences of *omdat* (39 cases) and *want* (90 cases). Because of the limited size of the corpus we had to add occurrences from other chat data: we selected all 27 occurrences of *omdat* in a pilot version of the CONDIV corpus (Grondelaers et al. 2000) and added ten randomly selected occurrences of *want*

---

5 Some corpus fragments contain more than one instance of the connective under analysis. In that case we have analyzed both instances. Consequently sometimes we have more than 100 instances per corpus per connective. Note that *omdat* can occur in sentence-initial position ("Omdat S1, S2") and in sentence-medial position ("S1, omdat S2"). As we are dealing with backward causals, we have only included the latter type of cases in our corpus.

from the same corpus. Only 12 *omdat*-instances from the CONDIV corpus could be used in the analyses reported below, as the IRC chat in the CONDIV corpus is extremely difficult to interpret and has many instances of *omdat* without an apparently appropriate context.

2.3 Procedure

In our analysis we followed the "complete double coding" strategy (Spooren and Degand 2010), in which the two authors coded the fragments independently and discussed discrepancies. We determined the subjectivity in the corpus examples by analyzing a number of properties of the discourse context (i.e., the segments surrounding the connectives) that provide information on the subjectivity of the relation. First we determined the size of the related segments. Then we analyzed the type of causal relation, the propositional attitude of the first segment, the SoC (if present), and the linguistic realization of the SoC.

2.4 Statistical analysis

As indicated earlier, the frequency of the two connectives differs per medium. The samples from which we collected the fragments also differed in size. For example, *omdat* is more frequent than *want* in the written corpus, and in the chat corpus there were not enough instances of *omdat* to create a sample of 100 occurrences (which was our initial target).

In order to compensate for the difference in size of the samples and the corpora from which these stem, we did not analyze the raw frequencies to test our hypotheses, but the logits of these frequencies. A logit is the natural logarithm of the frequency of a phenomenon, divided by the corpus size minus the frequency of the phenomenon (in formula: ln(frequency/(corpus_size – frequency))). We used the data in Table 1 to estimate the size of the corpus from which our samples were chosen.

To test the hypotheses that differences between *want* and *omdat* generalize over media, we carried out logit analyses, in which contributions from the variables to account for the variation in the data are evaluated in order to establish the best fitting model. Four separate analyses were carried out, one for each indicator of subjectivity (type of relation, propositional attitude in the first segment, type of SoC, linguistic realization of the SoC).

To answer the research question concerning the relative weight of each indicator of subjectivity, a so-called CART (Clustering and Regression Tree) analy-

sis was carried out (Baayen 2008: 148–154). With this analysis we tried to set up a model that predicts whether a fragment uses *omdat* or *want* on the basis of such factors as the type of relation, the type, and linguistic realization of the SoC, the propositional attitude in the first segment and the type of medium.

3 Results

For ease of reading we present the statistical details of the analyses in the appendices. In the main text we will present those parts of the analyses that directly test our four hypotheses. In footnotes we will present additional significant parts of the analysis.

3.1 Type of Relation

Our overall hypothesis is that, irrespective of medium, *want* occurs more often in subjective contexts than *omdat*. For Type of Relation this means that we expect to find more Epistemic/Speech Act relations with *want* than with *omdat*. The results are summarized in Table 2.

Examples (15)–(19) illustrate the findings as summarized in Table 2. Examples (15), (16), and (17) are prototypical examples of *omdat* expressing content (15) and *want* expressing an epistemic (16) and a speech act (17) relation. (18) and

Table 2: Type of relation in spoken, chat and written data, by connective (percentages are column percentages per medium).

		Omdat	Want
Spoken			
	Content	89 (89.9)	40 (40.4)
	Epist./Speech Act	10 (10.1)	59 (59.6)
Chat			
	Content	45 (88.2)	35 (35.0)
	Epist./Speech Act	6 (11.8)	65 (65.0)
Written			
	Content	95 (95.0)	39 (39.0)
	Epist./Speech Act	5 (5.0)	61 (61.0)

Note: Four fragments had relations with different possible readings and were coded as missing.

(19) are non-prototypical cases of *omdat* expressing an epistemic and *want* expressing a content relation.

Fragment 15 is from the spoken corpus, more specifically an interview with a school teacher who explains how he arrived at this school. S1 expresses a volitional action, which is explained in S2; the two segments are connected with *omdat*, expressing a content-volitional relation ("the reason was ...").[6]

(15) *Omdat* expressing a volitional content relation
   maar [S1 ik ben m wel hier meteen uh op school uh terecht gekomen na mijn examen van de PA].
   'but I did m manage uh to go to this school immediately uh after my final examination at the teacher training college'

   omdat [S2 mij dat gevraagd werd om hier les te komen geven en ik daar wel trek in had.]
   'OMDAT I was asked to teach here and I felt like doing it'

Fragment (16) was taken from a Dutch newspaper story about English football player Tony Adams, who is the SoC and Speaker in this fragment. In S1 Adams (Speaker=SoC) draws a conclusion about someone else's behavior (*he* – notably football player David Beckham) and explains this conclusion on the basis of knowledge of an ongoing state of affairs, signaled by *want*, expressing an epistemic relation.

(16) *Want* expressing an epistemic relation
   [S1 Ik weet niet meer wat hij zei maar hij moet het gewaardeerd hebben], want [S2 hij heeft er sindsdien vaak over gesproken]
   'I don't know what he said but he must have appreciated it WANT he spoke of it often since then'

Fragment (17) is part of a chat conversation between two middle school students, in which one asks a question and subsequently provides the reason for asking this question. This is a prototypical example of a speech act use of *want* in chat. The relation can be paraphrased as "I ask you what your address is and the reason for my asking (speech act) is that I do not have the address."

---

6 Note: "ggg" stands for guttural sounds, "xxx" means uninterpretable.

(17) *Want* expressing a speech act relation
   maarre tim … [S1 wat's jou egte adres]
   'But eh tim … what is your real address'

   want [S2 die heb ik niej]
   'WANT that I don't have'

Fragment (18) is a case of an epistemic relation, but expressed in an *omdat*-construction. It is from the written corpus, and an interviewee is quoted.

(18) *Omdat* expressing an epistemic relation
   De oefenmeester, zelf nog een groentje in het Europese topvoetbal, klampt zich maar vast aan de ervaring van vorige week op Old Trafford, toen zijn ploeg de offensieve intenties van Manchester United met verbluffend positiespel ontregelde.
   'The trainer, himself a newcomer in European top football, clings to his experience from last week at Old Trafford, when his team disorganized the offensive intentions of Manchester United using astonishing positional play.'

   "Het spel van Manchester United ligt ons wel.
   'Manchester United's type of play suits us nicely.'

   Bovendien [S1 zullen zij wederom op de aanval speculeren], omdat [S2 ze normaal gesproken moeten winnen".]
   'Moreover, they will again speculate on attacking, OMDAT normally speaking they have to win".'

Fragment (19) is a volitional content relation from a chat conversation between students. The speaker explains why Miranda's face was completely red, using a *want*-coordination.

(19) *Want* expressing a content relation
   A   *en Miranda's gezicht was helemaal rood. WANT ze had gemische peeling gehad ofzo.*
      'and Miranda's face was completely red. WANT she had had a chemical peeling or something.'

The logit analysis is summarized in Table A1 in the appendix (a short introduction to the interpretation of these tables is provided at the beginning of the appendix). The data are best described with a model containing main effects of Connective and Type of Relation and interactions of Connective * Medium and

Connective * Type of Relation (model 6 in Table A1). The fit of the resulting model is adequate ($\chi^2(4) = 3.43$, p = .49). Directly related to our hypothesis is the interaction between Connective and Type of Relation: In *omdat*-fragments there are relatively few Epistemic/Speech Act relations (21 out of 250 relations or 8.4%), in *want*-fragments the majority are Epistemic/Speech Act relations (185 out of 299 relations or 61.9%).[7]

3.2 Propositional Attitude

The Propositional Attitude hypothesis states that irrespective of medium, first segments of *want* fragments more often express an opinion, compared to *omdat* fragments. The data are summarized in Table 3.

Fragment (20), from the spoken corpus, illustrates a judgment in S1, which is the dominant propositional attitude for *want*-connections.

(20) Judgment in S1
   [S1 dat is gewoon krankzinnig].
   'that is simply insane'

Table 3: Type of propositional attitude in spoken, chat and written data, by connective (percentages are column percentages per medium).

		Omdat	Want
Spoken	Judgment	43 (43.4)	54 (54.0)
	Other propositional attitudes	56 (56.6)	46 (46.0)
Chat	Judgment	12 (23.5)	35 (35.0)
	Other propositional attitudes	39 (76.5)	65 (65.0)
Written	Judgment	42 (42.0)	73 (71.6)
	Other propositional attitudes	58 (58.0)	29 (28.4)

Note: One case is missing because it allowed for multiple readings.

---

7 The parameter estimates for model 6 (in Appendix, Table A2) allow for an interpretation of the other effects. The main effect of Connective shows that there are somewhat more *want*-fragments than *omdat*-fragments. The main effect of Type of Relation indicates that overall there are less instances of Epistemic and Speech Act relations compared to Content relations. The interaction between Connective and Medium reflects the fact that there are relatively few fragments with *want* in the Written medium (relative to the size of the corpora).

want [S2 als hij uhm mensen goed inschat moet ie ook weten dat ik m'n uiter-ste best doe om dat zo snel mogelijk voor elkaar te krijgen.]
'WANT if he uhm is such a good judge of character then he should also know that I am doing my very best to take care of that as soon as possible.'

In (21), a *want*-construction without a judgment in S1, taken from a chat-conversation, a pupil explains why he cannot always watch his favorite TV-series. This is a non-volitional causal relation, which even could have been expressed by a *doordat*.

(21) *Want* expressing a non-volitional content relation
alleen [S1 kan t niet altijd kijken] want [S2 mn vader wil altijd journaal kijken]
'only cannot always watch it WANT my father always wants to watch the news.'

(22) shows the typical *omdat*-pattern: the propositional attitude in S1 is other than judgment and is presented in a construction expressing a volitional relation.

(22) *Omdat* expressing a non-judgment in S1
[S1 Drie vrouwen van middelbare leeftijd worden achterna gezeten] omdat [S2 ze het waagden te protesteren.]
'Three middle-aged women are chased OMDAT they dared to protest.'

Fragment (23), from a newspaper, shows a non-typical and infrequent occurrence of an *omdat*-construction with a clear judgment, expressing an epistemic relation.

(23) *Omdat* expressing a judgment in S1
[S1 Sint Maarten kan hier niet afgebeeld zijn] omdat [S2 het desbetref-fende portaal (...) aan de martelaren gewijd is en dat was Maarten niet.]
'It cannot be Saint Martin who is depicted here OMDAT the portal in ques-tion (...) is devoted to martyrs and Saint Martin wasn't a martyr.'

The logit-analysis is summarized in Table A3 (Appendix). The data are best de-scribed with a model containing a main effect of Connective and interactions of Connective * Medium, Connective * Propositional Attitude and Medium * Propositional Attitude (model 7 in Table A3). The fit of the resulting model is adequate ($\chi^2(2) = 4.34$, p = .11). Directly related to our hypothesis is the significant interaction between Connective and Propositional Attitude: In *omdat*-fragments

there are relatively few judgments (97 out of 250 relations or 38.8%), in *want*-fragments judgments are the majority (162 out of 302 relations or 53.6%).[8]

3.3 SoC-type

Our next analysis concerns the relationship between SoC, medium, and connective. In this analysis we compared 1st and 2nd person SoC on the one hand with 3rd person SoC on the other, see Table 4. As there are relatively few 2nd person SoCs in the medium, we grouped them together with 1st person SoCs. In the analysis we disregarded first segments without a SoC (facts) and fragments in which the SoC is a secondary speaker (a quoted character).

The logit analysis shows that the data in Table 4 are best described by a model containing all three variables (Table A5 in the Appendix shows that the fit of this model was perfect: $\chi^2(0) = 0.00$). Central to our research question is the two-way interaction between Connective and SoC, which means that the predominance of first/second person SoCs is much larger for *want*-fragments

Table 4: Type of SoC in spoken, chat and written data, by connective (percentages are column percentages per medium).

		Omdat	Want
Spoken	1st/2nd person	76 (82.6)	73 (76.8)
	3rd person	16 (17.4)	22 (23.2)
Chat	1st/2nd person	39 (81.2)	90 (91.8)
	3rd person	9 (18.8)	8 (8.2)
Written	1st/2nd person	19 (24.7)	57 (72.2)
	3rd person	58 (75.3)	22 (27.8)

Note: 64 cases are missing (either the first segment does not have a SoC because it is a fact, or the SoC is a quoted character).

---

8 The parameter coefficients for model 7 (table A4) suggest the following interpretation for the other effects. The main effect of Connective shows that there are relatively more *want*-fragments than *omdat*-fragments in this analysis. The interaction between Connective and Medium reflects the fact that there are relatively few fragments with *want* in the Written medium (relative to the size of the media). The interaction between Medium and Propositional Attitude reflects the fact that the predominance of other propositional attitudes over s is largest in the chat medium.

than for *omdat*-fragments (in conformity with our hypothesis), see examples (16), (17), and (18). However, this result is modified by a significant three-way interaction between Medium ＊ SoC ＊ Connective showing that this predominance of first/second person SoCs in *want* fragments is even stronger in the written medium, because there *omdat*-fragments generally occur with third person SoCs.[9]

Overall, *want* shows a consistent pattern over the media: it has predominantly 1st person SoCs. *Want*-cases in the chat medium have an even higher amount of 1st person SoCs than in the other media, see example (17), and (24) below.

(24) *[S1 geen praatjes he kleine man]*
     'no big mouth ay little man'

   want [S2 anders zet ik je in der prullenbakk]
   'WANT otherwise I will put you in the wastepaper basket'

Omdat has a clearly different behavior: in the spoken and chat medium, it resembles *want* with its abundance of 1st person SoCs as in example (9); in the written medium, however, this predominance has reversed, in that there are mainly 3rd person SoCs, as illustrated in example (25) below.

(25) *[S1 Maar de technocraten wilden per se aan de macht blijven], omdat [S2 ze hun economisch model in stand wilden houden].*
     'But the technocrats absolutely wanted to maintain power, OMDAT they wanted to hold on to their economic model.'

---

9 The main effect of Connective shows that there are relatively more *want*-fragments than *omdat*-fragments (note that this analysis is restricted to fragments in which SoCs (first/second person or third person) occur. The main effect of Medium reflects the fact that there are relative few fragments with a first or third person conceptualizer in the written medium, indicating that the written medium had relatively many factual relations like (10). The main effect of SoC shows that overall there were relatively less third person than first/second person SoCs. The interaction between Connective and SoC can be interpreted as follows: the predominance of *want* fragments in which conceptualizers occur is higher in the written medium than in the other two media.

The two-way interaction between Medium and SoC can be interpreted as follows: in the spoken medium and the chat medium there are relatively few third person SoCs compared to first/second person SoCs, whereas in the written medium there are more third person SoCs than first/second person SoCs. In other words, we see more first and second person SoCs in chat and spoken language.

Table 5: Linguistic marking of SoC in spoken, chat and written data, by connective (percentages are column percentages per medium).

		Omdat	Want
Spoken	explicit	67 (71.3)	57 (58.2)
	implicit	27 (28.7)	41 (41.8)
Chat	explicit	30 (62.5)	52 (53.1)
	implicit	18 (37.5)	46 (46.9)
Written	explicit	48 (53.9)	37 (36.6)
	implicit	41 (46.1)	64 (63.4)

Note. 25 cases are missing (the first segment does not have a SoC because it expresses a fact).

3.4 Linguistic marking of the SoC

Our next analysis concerns the relationship between connective, medium, and linguistic realization of the SoC. Remember that a SoC (if present) can be referred to explicitly in the first segment or that it can remain implicit. The latter is judged to be more subjective than the former. For that reason it is expected that implicit SoCs occur more often in the first segment of *want*-fragments than in that of *omdat*-fragments. The data are summarized in Table 5.

The logit analysis shows that the data are best described with a model containing all main effects and all two-way interactions. The fit of the resulting model is acceptable ($\chi^2(2) = 0.46$, $p = 0.80$). The parameter estimates for this model are presented in Table A8 (Appendix). Directly of interest for our hypothesis is the two-way interaction between Connective and Linguistic Marking: The predominance of explicit markings is less strong for *want* than for *omdat*. In other words, and as predicted, *want*-fragments have more implicit marking of the SoC than *omdat*-fragments (*omdat*: 86 out of 231 cases or 37.2%; *want*: 151 out of 297 or 50.8%).[10]

---

10 The main effect of Connective reflects the overall predominance of *want* fragments in this analysis of linguistic marking. The main effect of Medium reflects the fact that there are relatively few fragments with implicit or explicit conceptualizers in the chat medium. The main effect of Linguistic Marking shows that overall explicit marking is predominant. The interaction of Connective and Medium reflects the fact that the predominance of *want* fragments in this analysis does not hold for the written medium.

Example (20), repeated here for convenience, is a clear prototypical example of *want* with a (first person) implicit SoC, expressing a judgment in S1.

(20) First person implicit SoC in S1
[S1 dat is gewoon krankzinnig].
'that is simply insane'

want [S2 als hij uhm mensen goed inschat moet ie ook weten dat ik m'n uiterste best doen om dat zo snel mogelijk voor elkaar te krijgen.]
'WANT if he uhm is such a good judge of character then he should also know that I am doing my very best to take care of that as soon as possible'

Fragment (26) shows a *want*-case from the spoken corpus, with a first person explicit SoC.

(26) *Want* with explicit 1st person SoC
en dan wil je mensen zo snel mogelijk helpen.
'and then you want to help people as soon as possible'

en en [S1 wa ik het nu heb dat is geen noodsituatie] want [S2 ik kan donders goed inschatten kwart over vier half vijf dat er dan geen student meer boven gaat kijken.
'and and wha I am having it now that is not an emergency because I can estimate damned well quarter past four half five that no student is going to look upstairs anymore'

3.5 The relative importance of the indicators of subjectivity

In the previous subsections we have shown that all four indicators of subjectivity play a significant role in characterizing the difference between *want*- and *omdat*-fragments. An important question is whether each characteristic is equally important for this characterization. In order to answer that question we have made use of a so-called CART (classification and regression tree) analysis, described in Baayen (2008: 148–154). A CART-analysis predicts the classification of an object on the basis of a number of factors. In our case, it produces a tree (see Figure 1) that outlines a decision procedure for determining the realization of *want*. Each split in the tree is labelled with a decision rule. The leaf nodes of the tree specify a partition of the data into a series of non-overlapping subsets. The analysis stops when a new split does not have enough explanatory value or when there are too few observations left to make a new split. The length of the branches is an
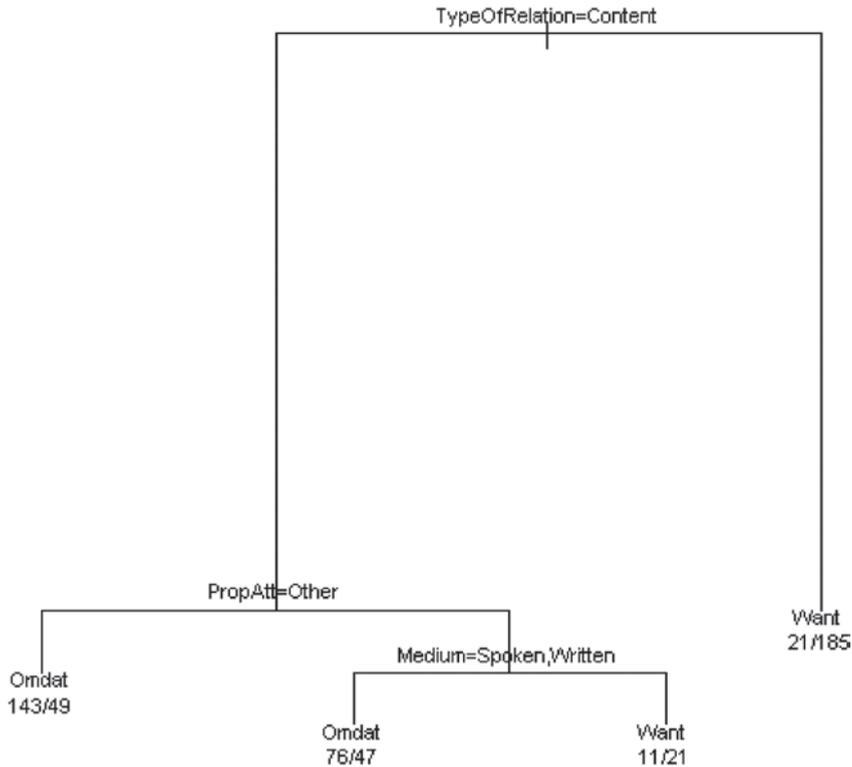
Fig. 1: A decision procedure for determining the realization of *want*, as produced by a CART (classification and regression tree) analysis (Baayen 2008: 148–154).

indication of the explanatory value of a split (the longer the branch, the higher the explanatory value).

The tree should be read as follows. The first decision concerns the type of relation. If it is not a content relation then the connective is predicted to be *want*. This leads to a correct prediction for 185 out of 206 cases. If it is a content relation then the next decision concerns the propositional attitude of the first segment. If that attitude is not a judgment the connective is predicted to be *omdat* (143 correct predictions out of 182 cases). The final decision concerns the medium: if the medium is written or spoken (as opposed to chat), the connective is predicted to be *omdat* (76 correct predictions, 47 incorrect predictions), if it is chat, it is predicted to be *want* (21 correct predictions, 11 incorrect). Note that groupings are determined by the program and were not predetermined by the investigators.

Overall, the analysis makes 425 correct predictions out of 553 cases or 76.9% correct. Compared to a minimal model in which the most frequent connective (*want*) is the only predictor, this leads to an improvement of 22.3%.

The analysis clearly shows that Type of Relation is by far the most important predictor of the connective choice. Other factors only matter for fine-tuning the prediction of the connective for content relations. We also see that Linguistic Marking and Type of SoC do not contribute substantively to the quality of the prediction.

4 Discussion

In this paper we investigated the system behind the meaning and use of backward causal connectives in discourse. Our starting point was the distinction between content / semantic / objective relations versus epistemic-speech act / pragmatic / subjective relations, which is well-known in text linguistics (ever since van Dijk 1979), functional (Degand 2001; Martin 1992) and cognitive linguistic (Sweetser 1990), and from work on (causal) connectives, as well as from cognitive approaches to coherence relations (Sanders et al. 1992; Sanders 1997). In addition, several linguists have suggested that the choices made by speakers of languages like Dutch *omdat* versus *want,* German *weil* versus *denn*, and French *parce que* versus *car* and *puisque*, are not only systematical, but also exactly reflect this distinction (see Pit 2006).

Here, we have argued in favor of a principle of subjectivity to explain the systematic differences between connectives. We have set out on a corpus investigation of the meaning and use of the Dutch connective pair *want* and *omdat*. The corpus that we analyzed was larger and more varied in media than in previous work, and we analyzed it using state-of-the-art statistical methods. But that was not the only innovative aspect of our study. We adopted an integrative empirical approach in order to solve two major challenges in the field. First, we decomposed the complex construct of subjectivity in several characteristics, which were analyzed separately. Second, we argued that it is important to investigate whether insights from existing work on written corpora can be generalized to other media, especially because claims concerning *cognitive reality* are at stake. After all, our most natural and spontaneous way to communicate is not merely via discourse, but through *spoken* discourse. Now that spoken corpora have become available in many languages, it is possible to test hypotheses against spoken corpus data. Chat data also provide an interesting case, because they are spontaneous, like spoken language, but chat lacks the immediate feedback, the intonation, and prosody of face to face spontaneous conversations. A more specific reason to be

interested in more spontaneous, less-edited language data is directly related to the interpretation of causal connectives as acts of categorization: How "basic" is this act? Are the distinctions only realized by highly proficient language users in a production context with many editing opportunities? Or are the same differences realized in the totally different production context of spontaneous conversations, characterized by limiting time constraints and few planning and editing options?

We analyzed a corpus of *omdat-* and *want-*cases from written, spoken and chat discourse. We expected subjectivity to go across the modalities of written, spoken and chat language. Therefore our main hypothesis was that *want* occurs in more subjective contexts than *omdat*, irrespective of the medium. We formulated four specific hypotheses on the way in which the connectives *want* and *omdat* would show differences in terms of subjectivity. When we summarize the main results of our corpus research, we can say that all four hypotheses repeated below were confirmed. Across all media

– *want* is used more often to express subjective relations (epistemic, speech act) than *omdat* (hypothesis 1);
– *want* is used more often to support a judgment than *omdat* (hypothesis 2);
– *want* is used more often with first and second person SoCs than *omdat* (hypothesis 3);
– *want* is used more often with an implicit SoC than *omdat* (hypothesis 4).

We can conclude from this that the subjectivity hypothesis and the generalization over media hypothesis are supported by the data: *Want* and *omdat* show a clearly different pattern over all media we investigated.

In addition to the hypotheses, we formulated two explorative research questions. The first was whether the medium affects the degree of subjectivity. The results are not unequivocal. We did not find a significant interaction between Medium and Type of Relation. We did find a significant interaction between Medium and Propositional Attitude, but it is not easy to interpret that interaction: Contrary to expectation there are relatively few judgments in chat. A plausible explanation is that chat is subjective not because of its abundance of judgments but because it has relatively many speech act relations (spoken: 20 out of 198 or 10.1% speech act relations; written: 11 out of 200 or 5.0% speech act relations; chat: 42 out of 151 or 27.8% speech act relations): Even though speech act relations can be considered very subjective, the first segment in a speech act relation is not a judgment. There was also a significant interaction between Medium and SoC: the written fragments are heavily dominated by Third Person SoC's, suggesting that written texts have many objective reports of causal relations. Finally, there was also a significant interaction between Medium and Linguistic Realiza-

tion of the SoC: surprisingly the written corpus has relatively many implicit SoC's. However, those implicit SoC's predominantly are first person SoC's in the written corpus, which constitute the minority. In sum, although the various interactions with Medium can be interpreted, it is not the case that there is a straightforward relationship between Medium and Subjectivity. This obviously is an area for further research.

The second explorative question concerns the relative strength of the various subjectivity characteristics: are all operationalizations of subjectivity (as formulated in the hypotheses) equally important? The CART-analysis provides a clear answer to that question: Type of Relation (content versus epistemic/speech act) is by far the most important predictor of the connective. The second important factor is propositional attitude of the first segment: If that attitude is not a judgment the connective is almost certainly *omdat*. The final decision concerns the medium: if the medium is written or spoken, the connective is predicted to be *omdat*.

Hence, our corpus study clearly corroborates the hypotheses formulated for the two causal connectives we have studied. There are substantial differences in the meaning and use of *omdat* and *want*: *want* is subjective in that it typically signals an epistemic or speech act relation, whereas *omdat* typically signals a content relation. In addition, *want* often has a judgment in the first segment, a first person conceptualizer, which is more often implicitly realized than *omdat*. This pattern roughly replicates earlier results reported by Pander Maat and Degand (2001) and Pit (2006) on the distribution of these connectives in written language. These differences between *want* and *omdat* survive across media, as they are found in spontaneous conversations, chat communication as well as written text.

What is the theoretical interpretation of our main findings? First of all, it shows the relevance of the notion of subjectivity, which we have defined, operationalized and actually used in corpus analysis in such a way that it indeed explains the differences between the two connectives. However, there is a fundamental issue to address here. Even though distinctions like objective-subjective or content-epistemic/speech act seem relevant across languages, many studies have observed that the causal categories are not *always* reflected in connective use. Like our study, earlier corpus studies have shown how, in a minority of cases, causal connectives that seem to specialize in one type of relation, *can* in fact be used to express other causal categories. For instance, even though Dutch *want* specializes in expressing epistemic relations, it can be and – as our study and earlier corpus studies show – actually *is* used to express content relations (Pit 2006). Similar observations exist for French and German connectives (Stukker and Sanders 2012). Apparently we are not dealing with black-and-white distinctions, but rather with tendencies. A crucial question is what consequences

such empirical observations should have for a theory of causal connectives as categorization devices.

In our view, the conceptual basis of linguistic categories offers a natural explanation of the fact that causal connectives in actual language use do not always directly reflect conceptual categories of causality (Stukker et al. 2008, 2009). A crucial insight here is that causal categories show prototypicality structure. Classical categorization theory (Rosch 1973) argued that robins are better examples of the category of birds than ospreys and puffins are. Similarly, connective uses that seem counter-examples against our categorization hypothesis, should be regarded as less prototypical members of the same category to which the "normal" uses belong. In that respect, it is not a coincidence that we often used terms like prototypical and less prototypical when characterizing patterns and examples presented above. More specifically, we expect the non-prototypical uses to have a different status in the language user's mental representation of the connectives' meaning and use (Bybee 2007; Stukker et al. 2008, 2009).

Such a position requires a more detailed analysis in various respects, which goes beyond the scope of this paper. We briefly mention two. One is the crosslinguistic comparison of patterns in connective meaning and use. Results of a meta-analysis of existing corpus studies indeed suggest highly similar patterns indicating a prototypicality structure for French, German, and Dutch (Stukker and Sanders 2012). A second issue is to show in (qualitative) linguistic analyses how exactly *want* and *omdat* result in different conceptual representations (Sanders et al. 2012), and perhaps even more importantly, how non-prototypical examples still show resemblance to their prototype. For instance, when we observe that, in a minority of cases, *omdat* can express epistemic relations, it is important to explain that this use of *omdat* is not a coincidence, but that the *omdat*-context shows, for instance, more objective characteristics than a *want*-context does (Degand 2001; Sanders and Spooren 2013; Stukker and Sanders 2012). This could even be done using automated large-scale quantitative analyses like the one Bestgen et al. (2006) used to study the subjective nature of the context of *omdat* and *want* in newspaper language.

In conclusion, we believe that causal categories are fundamental to human cognition and natural language at the discourse level. Causality and subjectivity are two cognitive principles that organize our knowledge of coherence relations. Notions like causality and subjectivity can help us explain the system and use of causal relations and their linguistic expressions in everyday language use, and following the methodological principle of converging evidence (cf. Gonzalez-Marquez et al. 2007, and contributions to this volume), we have shown elsewhere that they explain the acquisition of connectives and relations (Evers-Vermeul and Sanders 2009, 2011; Spooren and Sanders 2008; Sanders and Spooren 2009 and

the references cited there) as well as discourse processing and representation (Canestrelli et al. 2013). Furthermore, it seems worthwhile to feed theories of connectives and coherence relations with corpus studies of spontaneous language use in communicative situations that allow for direct interaction; for one thing, we have never before seen so many attested speech act relations as in our chat corpus. Systematic comparison of various communicative situations is imperative. Finally, we analyzed the causal connections in terms of detailed characteristics and subsequently investigated whether and to what extent they co-occur. We made use of statistical methods specifically suitable for hypothesis testing in natural language corpora. Such an enterprise provides new insights into the notion of subjectivity. We would like to see such methods used in studies that could reveal whether other, less-related languages, also encode such categories of causality, or other types of coherence relations. Recent results on Mandarin Chinese (Li et al. 2013), and results from studies looking into parallel corpora of translated texts (Cartoni et al. 2013) are promising.

References

Anscombre, Jean-Claude & Oswald Ducrot. 1983. *L'argumentation dans la langue*. Brussels: Pierre Mardaga.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Banfield, Ann. 1982. *Unspeakable sentences: Narration and representation in the language of fiction*. Boston: Routledge & Kegan Paul.

Bekker, Birgit. 2006. *De feiten verdraaid. Over tekstvolgorde, talige markering en sprekerbetrokkenheid* [The twisted facts: About text order, linguistic marking and speaker involvement]. Tilburg: Tilburg University dissertation.

Bestgen, Yves, Liesbeth Degand & Wilbert Spooren. 2006. Towards automatic determination of the semantics of connectives in large newspaper corpora. *Discourse Processes* 41(2). 175–194.

Breindl, Eva & Maik Walter. 2009. *Der Ausdruck von Kausalität im Deutschen: Eine korpusbasierte Studie zum Zusammenspiel von Konnektoren, Kontextmerkmalen und*

Diskursrelationen (Arbeiten und Materialien zur deutschen Sprache). Mannheim: Institut für Deutsche Sprache.

Bybee, Joan. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Canestrelli, Anneloes, Willem Mak & Ted Sanders. 2013. Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes* 28(9). 1394–1413.

Cartoni, Bruno, Sandrine Zufferey & Thomas Meyer. 2013. Annotating the meaning of discourse connectives by looking at their translation: The translation-spotting technique. *Dialogue & Discourse* 4(2). 65–86.

Chafe, Wallace. 1994. *Discourse, consciousness and time*. Chicago: Chicago University Press.

Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.

D-COI. 2006. The Dutch Language Corpus Initiative. http://lands.let.ru.nl/projects/d-coi/ (accessed 28 June 2010).

Degand, Liesbeth. 2001. *Form and function of causation: A theoretical and empirical investigation of causal constructions in Dutch*. Louvain: Peeters.

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.

De Smet, Hendrik & Jean-Christophe Verstraete. 2006. Coming to terms with subjectivity. *Cognitive Linguistics* 17. 365–392.

Diessel, Holger & Katja Hetterle. 2011. Causal clauses: A cross-linguistic investigation of their structure, meaning, and use. In Peter Siemund (ed.), *Linguistic universals and language variation*, 23–54. Berlin & New York: De Gruyter Mouton.

van Dijk, Teun. 1979. Pragmatic connectives. *Journal of Pragmatics* 3. 447–456.

Evers-Vermeul, Jacqueline, Liesbeth Degand, Benjamin Fagard & Liesbeth Mortier. 2011. Historical and comparative perspectives on subjectification: A corpus-based analysis of Dutch and French causal connectives. *Linguistics* 49(2). 445–478.

Evers-Vermeul, Jacqueline & Ted Sanders. 2009. The emergence of Dutch connectives; how cumulative cognitive complexity explains the order of acquisition. *Journal of Child Language* 36(4). 829–854.

Evers-Vermeul, Jacqueline & Ted Sanders. 2011. Discovering domains: On the acquisition of causal connectives. *Journal of Pragmatics* 43(6). 1645–1662.

Ford, Celia E. 1993. *Grammar in interaction: Adverbial clauses in American English conversations*. Cambridge: Cambridge University Press.

Frohning, Dagmar. 2007. *Kausalmarker zwischen Pragmatik und Kognition: Korpusbasierte Analysen zur Variation im Deutschen*. Tübingen: Niemeyer.

Gonzalez-Marquez, Maria, Irene Mittelberg, Seana Coulson & Michael J. Spivey. 2007. *Methods in cognitive linguistics*. Amsterdam & Philadelphia: John Benjamins.

Gries, Stefan T. 2012. Corpus linguistics, theoretical linguistics, and cognitive/psycholinguistics: towards more and more fruitful exchanges. In Joybrato Mukherjee & Magnus Huber (eds.), *Corpus linguistics and variation in English: Theory and description*, 41–63. Amsterdam: Rodopi.

Grondelaers, Stefan, Katrien Deygers, Hilde van Aken, Vicky van den Heede & Dirk Speelman. 2000. Het CONDIV-corpus geschreven Nederlands [The CONDIV corpus of written Dutch]. *Nederlandse Taalkunde* 5(4). 356–363.

Groupe λ. 1975. Car, parceque, puisque. *Revue Romane* 10. 248–280.

Günthner, Susanne. 1993. '... weil – man kann es ja wissentschaftlich untersuchen' – Diskurspragmatische Aspekte der Wortstellung in WEIL-Sätzen. *Linguistische Berichte* 143. 37–55.

Huiskes, Mike. 2010. *The role of the clause for turn-taking in Dutch conversations*. Utrecht: Utrecht University dissertation.

Kehler, Andrew. 2002. *Coherence, reference and the theory of grammar*. Chicago: University of Chicago Press.

Keller, Rudi. 1995. The epistemic *weil*. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectification: Linguistic perspectives*, 16–30. Cambridge: Cambridge University Press.

Knott, Alistair & Robert Dale. 1994. Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes* 18. 35–62.

Knott, Alistair & Ted Sanders. 1998. The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics* 30. 135–75.

Lakoff, George. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Langacker, Ronald. 1990. Subjectification. *Cognitive Linguistics* 1. 5–38.

Li, Fang, Jacqueline Evers-Vermeul & Ted Sanders. 2013. Subjectivity and result marking in Mandarin: A corpus-based investigation. *Chinese Language and Discourse* 4. 74–119.

Lyons, John R. 1977. *Semantics*. Cambridge: Cambridge University Press.

Lyons, John R. 1995. *Linguistic semantics: An introduction*. Cambridge: Cambridge University Press.

Mann, William & Sandra Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text* 8. 243–281.

Martin, James R. 1992. *English text: System and structure*. Amsterdam & Philadelphia: John Benjamins.

Nuyts, Jan. 2012. Notions of (inter)subjectivity. *English Text Construction* 5(1). 53–76.

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. 2000. Domains of use or subjectivity? The distribution of three Dutch causal connectives explained. In Elisabeth Couper-Kuhlen & Bernd Kortmann (eds.), *Cause, condition, concession and contrast: Cognitive and discourse perspectives*, 57–81. Berlin & New York: Mouton de Gruyter.

Pander Maat, Henk & Ted Sanders. 2001. Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics* 12(3). 247–273.

Pasch, Renate. 1983. Die Kausalkonjunktionen "da", "den", und "weil": drei Konjunktionen – drei lexikalische Klassen. *Deutsch als Fremdsprache* 20. 332–337.

Persoon, Ingrid, Ted Sanders, Hugo Quené & Arie Verhagen. 2010. Een coördinerende *omdat*-constructie in gesproken Nederlands? Tekstlinguïstische en prosodische aspecten [A coordinating because-construction in spoken Dutch? Text-linguistic and prosodic aspects]. *Nederlandse Taalkunde* 1. 259–282.

Pit, Mirna. 2006. Determining subjectivity in text: The case of backward causal connectives in Dutch. *Discourse Processes* 4. 151–174.

Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology* 4. 328–350.

Sanders, José. 2010. Intertwined voices: Journalists' modes of representing source information in journalistic subgenres. *English Text Construction* 3(2). 226–249.

Sanders, José, Ted Sanders & Eve Sweetser. 2012. Responsible subjects and discourse causality. How mental spaces and perspective help identifying subjectivity in causal connectives. *Journal of Pragmatics* 44(2). 191–213.

Sanders, Ted. 1997. Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes* 24. 119–147.

Sanders, Ted & Wilbert Spooren. 2009. Causal categories in discourse – Converging evidence from language use. In Ted Sanders & Eve Sweetser (eds.), *Causal categories in discourse and cognition*, 205–246. Berlin: Mouton de Gruyter.

Sanders, Ted & Wilbert Spooren. 2013. Exceptions to rules: a qualitative analysis of backward causal connectives in Dutch naturalistic discourse. *Text & Talk* 33(3). 399–420.

Sanders, Ted, Wilbert Spooren & Leo Noordman. 1992. Toward a taxonomy of coherence relations. *Discourse Processes* 15(1). 1–35.

Sanders, Ted & Eve Sweetser (eds.). 2009. *Causal categories in discourse and cognition*. Berlin: Mouton de Gruyter.

Sanders, Ted, José Sanders & Eve Sweetser. 2009. Causality, cognition and communication: A mental space analysis of subjectivity in causal connectives. In Ted Sanders & Eve Sweetser (eds.), *Causal categories in discourse and cognition,* 21–60. Berlin: Mouton de Gruyter.

Schlenker, Philippe. 2004. Context of thought and context of utterance: A note on free indirect discourse and the historical present. *Mind & Language* 19(3). 279–304.

Spooren, Wilbert & Liesbeth Degand. 2010. Coding coherence relations: reliability and validity. *Corpus Linguistics & Linguistic Theory* 6(2). 241–266.

Spooren, Wilbert & Ted Sanders. 2008. The acquisition of coherence relations: On cognitive complexity in discourse. *Journal of Pragmatics* 40(12). 2003–2026.

Spooren, Wilbert, Ted Sanders, Mike Huiskes & Liesbeth Degand. 2010. Subjectivity and causality: A corpus study of spoken language. In John Newman & Sally Rice (eds), *Empirical and experimental methods in cognitive/functional research,* 241–255. Stanford CA.: CSLI.

Stukker, Ninke & Ted Sanders. 2012. Subjectivity and prototype structure in causal connectives: A cross-linguistic perspective. *Journal of Pragmatics* 44(2). 169–190.

Stukker, Ninke, Ted Sanders & Arie Verhagen. 2008. Causality in verbs and in discourse connectives. Converging evidence of cross-level parallels in Dutch linguistic categorization. *Journal of Pragmatics* 40. 1296–1322.

Stukker, Ninke, Ted Sanders & Arie Verhagen. 2009. Categories of subjectivity in Dutch causal connectives: a usage-based analysis. In Ted Sanders & Eve Sweetser (eds.), *Causal categories in discourse and cognition*, 119–171. Berlin & New York: Mouton de Gruyter.

Sweetser, Eve. 1990. *From etymology to pragmatics.* Cambridge: Cambridge University Press.

Traugott, Elizabeth. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 57. 33–65.

Traugott, Elizabeth. 1995. Subjectification in grammaticalization. In Dieter Stein & Susan Wright (eds.), *Subjectivity and subjectivisation: Linguistic perspectives*, 31–54. Cambridge: Cambridge University Press.

Verhagen, Arie. 2005. *Constructions of intersubjectivity: Discourse, syntax and cognition*. Oxford: Oxford University Press.

Verhagen, Arie & Suzanne Kemmer. 1997. Interaction and causation: Causative constructions in modern standard Dutch. *Journal of Pragmatics* 27. 61–82.

Vis, Kirsten. 2011. *Subjectivity in news discourse: A corpus-linguistic analysis of informalization*. Amsterdam: VU University Amsterdam dissertation.

Volodina, Anna. 2011a. *Konditionalität und Kausalität im Diskurs: Eine korpuslinguistische Studie zum Einfluss von Syntax und Prosodie auf die Interpretation komplexer Äußerungen*. Tübingen: Narr.

Volodina, Anna. 2011b. Sweetsers Drei-Ebenen-Theorie: Theoretische Überlegungen vor dem Hintergrund einer korpuslinguistischen Studie über konditionale und kausale Relationen. In Gisella Ferraresi (ed.), *Konnektoren im Deutschen und im Sprachvergleich: Beschreibung und grammatische Analyse*, 127–155. Tübingen: Narr.

Wegener, Heide. 2000. *Da, denn* und *weil* – der Kampf der Konjunktionen: Zur Grammatikalisierung im kausalen Bereich. In Rolf Thieroff, Matthias Tamrat, Nanna Fuhrhop & Oliver Teuber (eds.), *Deutsche Grammatik in Theorie und Praxis*, 69–81. Tübingen: Niemeyer.

Zufferey, Sandrine, 2010. *Lexical pragmatics and theory of mind: The acquisition of connectives*. Amsterdam & Philadelphia: John Benjamins.

Zufferey, Sandrine, 2012. "Car, parce que, puisque" revisited: Three empirical studies on French causal connectives. *Journal of Pragmatics* 44(2). 138–153.

Appendix

Introduction: How to read the tables

Each logistic regression analysis has resulted in two tables: a summary of the analyses and an estimate of the parameters in the resulting model. In the summary, a progressively more complex model is tested for its fit to the data. For example, the first line in Table A1 states that a model with only the constant as a predictor does not fit the data well, because its χ^2 is highly significant (p < .001). The second line states that adding the factor "Connective" improves the model significantly ($\chi^2(1) = 30.31$, p < .001), but the resulting model is still not a very good fit ($\chi^2(10) = 286.32$, p < .001). The table shows in line 7 that adding the interaction Medium*Type of relation does not improve the model compared to the model specified in line 6. That is why the model specified in line 6 is the resulting model.

Table A2 gives the parameters for this model. Positive estimates indicate that the odds that the fragment is of the described type go up, compared to the reference category, whereas negative estimates indicate that the odds go down. For example, the fifth line states that when the relation type is Epistemic/SpeechAct, the estimate is –2.39 and that this estimate significantly differs from zero (p < .01). This means that the number of epistemic/speech act relations is significantly lower than that of the reference category, content relations.

Analysis 1: Type of relation

Table A1: Summary of logit analysis

Component	χ² model	df	p model	χ² factor	df	p factor
1. constant	316.63	11	<.001			
2. + Connective	286.32	10	<.001	30.31	1	<.001
3. + Medium	280.88	8	<.001	5.44	2	0.07
4. + Type of Relation	246.31	7	<.001	34.57	1	<.001
5. + Connective*Medium	188.31	5	<.001	58.00	2	<.001
6. + Connective*Type of Relation	3.43	4	0.49	184.88	1	<.001
7. + Medium*Type of Relation	1.97	2	0.37	1.46	2	0.48
8. + Connective*Medium*Type of Relation	0.00	0	1.00	1.97	2	0.37

Table A2: Parameter estimates for model 6 (Table A1).

Coefficients	Estimate	SE	z value	p
(Intercept)	−7.65	0.10	−74.26	<.01
Connect:Want	0.27	0.16	1.68	n.s.
Medium:Chat	−0.16	0.17	−0.92	n.s.
Medium:Written	0.57	0.14	4.01	<.01
RelType:Epist/SA	−2.39	0.23	−10.48	<.01
Connect:Want;Medium:Chat	−0.31	0.22	−1.37	n.s.
Connect:Want;Medium:Written	−1.48	0.20	−7.38	<.01
Connect:Want;RelTtype:Epist/SA	2.87	0.26	11.17	<.01

Analysis 2: Propositional attitude

Table A3: Summary of loglit analysis

Component	χ² model	df	p model	χ² factor	df	p factor
1. constant	143.01	11	<.001			
2. + Connective	112.76	10	<.001	30.25	1	<.001
3. + Medium	107.12	8	<.001	5.64	2	0.06
4. + Prop.Att.	105.02	7	<.001	2.10	1	0.15
5. + Connective*Medium	46.65	5	<.001	58.37	2	<.001
6. + Connective*Prop.Att.	34.49	4	<.001	12.16	1	<.001
7. + Medium*Prop.Att.	4.34	2	0.11	30.15	2	<.001
8. + Connective*Medium*Prop.Att.	0.00	0	1.00	4.34	2	0.11

Table A4: Parameter estimates for model 7 (Table A3).

Coefficients	Estimate	SE	z value	p
(Intercept)	−8.50	0.14	−58.65	<.01
Connect:Want	1.54	0.17	8.95	<.01
Medium:Chat	−0.78	0.24	−3.22	<.01
Medium:Written	0.76	0.18	4.15	<.01
PropAtt:NoJudg	0.44	0.17	2.56	<.05
Connect:Want;Medium:Chat	−0.15	0.23	−0.64	n.s.
Connect:Want;Medium:Written	−1.54	0.20	−7.56	<.01
Connect:Want;PropAtt:NoJudg	−0.77	0.18	−4.22	<.01
Medium:Chat;PropAtt:NoJudg	0.89	0.23	3.83	<.01
Medium:Written;PropAtt:NoJudg	−0.34	0.20	−1.66	n.s.

Analysis 3: Type of SoC

Table A5: Summary of logit analysis

Component	χ² model	df	p model	χ² factor	df	p factor
1. constant	297.21	11	<.001			
2. + Connective	263.47	10	<.001	33.74	1	<.001
3. + Medium	257.27	8	<.001	6.20	2	<.05
4. + SoC	155.58	7	<.001	101.69	1	<.001
5. + Connective*Medium	107.20	5	<.001	48.38	2	<.001
6. + Connective*SoC	85.09	4	<.001	22.10	1	<.001
7. + Medium*SoC	22.99	2	<.001	62.11	2	<.001
8. + Connective*Medium*SoC	0.00	0	1.00	22.99	2	<.001

Table A6: Parameter estimates for model 8 (Table A5).

Coefficients	Estimate	SE	z value	p
(Intercept)	−7.75	0.11	−67.55	<.01
Connect:Want	1.08	0.16	6.56	<.01
Medium:Chat	−0.17	0.20	−0.89	n.s.
Medium:Written	−0.64	0.26	−2.49	<.05
SoC:3rd	−1.56	0.28	−5.67	<.01
Connect:Want;Medium:Chat	−0.11	0.25	−0.44	n.s.
Connect:Want;Medium:Written	−0.33	0.31	−1.07	n.s.
Connect:Want;SoC:3rd	0.36	0.37	0.98	n.s.
Medium:Chat;SoC:3rd	0.09	0.46	0.20	n.s.
Medium:Written;SoC:3rd	2.68	0.38	7.01	<.01
Connect:Want;Medium:Chat;SoC:3rd	−1.31	0.64	−2.06	<.05
Connect:Want;Medium:Written;SoC:3d	−2.43	0.52	−4.69	<.01

Analysis 4: Linguistic marking of the SoC

Table A7: Summary of logit analysis

Component	χ² model	df	p model	χ² factor	df	p factor
1. constant	122.38	11	<.001			
2. + Connective	92.70	10	<.001	29.68	1	<.001
3. + Medium	86.27	8	<.001	6.43	2	<.05
4. + Ling.Mark.	80.73	7	<.001	5.54	1	<.05
5. + Connective*Medium	25.93	5	<.001	54.80	2	<.001
6. + Connective*Ling.Mark.	16.14	4	<.01	9.79	1	<.01
7. + Medium*Ling.Mark.	0.46	2	n.s.	15.68	2	<.001
8. + Connective*Medium*Ling.Mark.	0.00	0	1.00	0.46	2	n.s.

Table A8: Parameter estimates for model 7 (Table A7).

Coefficients	Estimate	SE	z value	p
(Intercept)	−7.90	0.12	−68.20	<.01
Connect:Want	0.94	0.16	5.94	<.01
Medium:Chat	−0.24	0.19	−1.26	n.s.
Medium:Written	0.26	0.17	1.52	n.s.
Mark:Implic	−0.91	0.18	−4.96	<.01
Connect:Want;Medium:Chat	−0.34	0.23	−1.49	n.s.
Connect:Want;Medium:Written	−1.59	0.21	−7.58	<.01
Connect:Want;Mark:Implic	0.58	0.18	3.15	<.01
Medium:Chat;Mark:Implic	0.27	0.23	1.17	n.s
Medium:Written;Mark:Implic	0.82	0.21	3.85	<.01