Subjectivity in Spanish Discourse: Explicit and Implicit Causal Relations in Different Contexts

Andrea Santana　　　　　　　　　　　A.C.SANTANACOVARRUBLAS@UU.NL

Utrecht Institute of Linguistic OTS, Utrecht University

Trans 10, 3512 JK, Utrecht, The Netherlands

Wilbert Spooren　　　　　　　　　　　W.SPOOREN@LET.RU.NL

Radboud Universiteit

Postbus/ PO Box 9103, NL-6500 HD Nijmegen

Dorien Nieuwenhuijsen　　　　　　　　D.NIEUWENHUIJSEN@UU.NL

Utrecht Institute of Linguistic OTS, Utrecht University

Trans 10, 3512 JK, Utrecht, The Netherlands

Ted Sanders　　　　　　　　　　　　　T.J.M.SANDERS@UU.NL

Utrecht Institute of Linguistic OTS, Utrecht University

Trans 10, 3512 JK, Utrecht, The Netherlands

Abstract

Corpus-based studies in various languages have demonstrated that some connectives are used preferentially to express subjective versus objective meanings, for example, *omdat* vs. *want* in Dutch. However, Spanish connectives have been understudied from this perspective. Moreover, most of the studies of subjectivity have focused on explicit relations and little is known about the subjectivity of implicit coherence relations. In addition, the role that context plays in the meaning and use of causal relations and their connectives is still under discussion. This study analyzes Spanish causal explicit and implicit relations in different contexts by carrying out manual text analyses, focused on subjectivity. 360 backward relations marked by three prototypical causal connectives and 120 backward implicit relations were extracted from academic and journalistic contexts. The analytical model is based on an integrative approach to subjectivity. Statistical analyses reveal no systematic profiles of Spanish connectives in terms of subjectivity. Furthermore, a significant three-way interaction between

subjectivity, context, and linguistic marking is observed. Based on a solid corpus analysis, this study reveals new insights into the expression of subjectivity in Spanish discourse relations.
Keywords: causality, subjectivity, Spanish connectives, implicit relations, manual analyses, context

1. Introduction

When interpreting discourse, experienced language users often infer causal relations between utterances, such as cause-consequence or claim-argument. Over the last two decades, the notions of causality and subjectivity have been identified as relevant cognitive principles that organize our knowledge of coherence relations (Sanders & Spooren, 2015). Causality is the implicational meaning (P→Q) that is inferred between connected discursive segments (Sanders et al., 1992; 1993). Subjectivity is a complex notion that has been defined from various theoretical approaches in linguistics (Lyons, 1977; Langacker, 1990; Traugott, 1995). Building on these insights, we consider subjectivity to be the degree to which the speaker is involved in the construal of the relation (Pander Maat & Sanders, 2000, 2001; Pander Maat & Degand, 2001; Degand & Pander Maat, 2003). By combining the principles of causality and subjectivity we distinguish between subjective and objective causal relations, depending on the presence of a Subject of Consciousness (SoC). This SoC is responsible for the causal relation that is constructed between two or more discourse segments. Thus, coherence relations defined in terms of domains by Sweetser (1990) can also be described in terms of subjectivity. For example, a speech-act relation like (1) is considered as a subjective causal relation since the SoC motivates her question by referring to the quality of the movie. It is not possible to interpret this relation without considering the speaker who performs the speech-act. The epistemic relation in (2) is also identified as a subjective causal relation because there is a SoC, again the speaker, who does the epistemic reasoning that John must love someone, based on the observation that he came back. Both relations differ from the content relation in (3), because in this case there is no SoC involved in the relation: the utterance described a causal link in the real world between John's love and his coming back[1]. That is why (3) is an example of an objective causal relation.

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

Coherence relations can, but need not, be marked explicitly through connectives. This paper focuses on backward relations, of the type consequence-cause (Q because P). (1), (2) and (3) are backward relations marked explicitly by *because*, but this connective could be omitted and the causal meaning between the discursive segments can be interpreted perfectly well. Connectives are defined as processing instructions on how one part of a text is related to another (Sanders & Spooren, 2007). Language users can choose between one or another connective to express causality, for instance, between *because*, *since* and *for* in English or between *omdat* and *want* in Dutch. The preference for one connective over another to express certain coherence relations is insightful, if it shows to be a systematic one. For instance, connectives may vary in the extent to which they can express subjectivity. In fact, several corpus-based studies have demonstrated that language users prefer to use some connectives to express subjective meanings, while others specialize in communicating objective meanings. This existing body of research has studied various languages. In Dutch, the connectives *dus* ('so'), *want* ('since/for') and *aangezien* ('since') have been identified as connectives used to express subjective meanings; *omdat* ('because'), *daardoor* ('as a result') and *doordat* ('because') have been recognized as connectives used preferentially to express objective meanings, while *daarom* ('that's why') occupies an

---

[1] Examples are taken from Sweetser (1990), who distinguishes between content, epistemic and speech act relations.

intermediate position (Pander Maat & Degand, 2001; Pander Maat & Sanders, 2000, 2001; Degand, 2001; Verhagen, 2005; Pit, 2006, 2007; Stukker & Sanders, 2009; Sanders & Spooren, 2015).

A similar pattern has been identified for German, where *denn* and *da* ('because') tend to express more subjective meanings, whereas *weil* (which also mean 'because') tends to express more objective meanings (Günthner, 1993; Keller, 1995; Wegener, 2000; Pit, 2007; Stukker & Sanders, 2012). The same situation has been observed in other, typologically less related languages. For example, in French, *car* ('for') and *puisque* ('since') have been recognized as connectives that are used preferentially to express subjective meanings as opposed to *parce que* ('because'), that has been identified as a connective that has a more general meaning and a preference to express objective relations (Degand & Pander Maat, 2003; Zufferey, 2012); in Mandarin-Chinese, *kějian* ('so/therefore') prefers to express subjective relations, whereas *yīn'er* ('as a result') tends to express objective meanings (Li, et al., 2013; Wei, et al., 2018).

These findings reveal that language users distinguish unconsciously between different types of causality. They also suggest that many languages of the world might have a comparable pattern in terms of subjectivity. That is, language users from various languages use different connectives to express more or less subjective causal relations. There are prominent exceptions, however, such as English *because*, which can be used to express speech act, epistemic as well as content relations, and the same holds for the connective *so* (Knott & Sanders, 1998; Sweetser, 1990). The crucial question is whether English is really exceptional, especially because we lack knowledge of many other languages. In a Romance language like Spanish, there is an extensive body of research on discourse markers and connectives (Briz, 1998, 2000; Martínez, 1997; Martín Zorraquino & Montolío, 1998; Montolío, 2001; Portolés, 2001; Pons, 1998, 2000; Vázquez Veiga, 2002; Santos Río, 2003; Garrido Rodríguez, 2004; Domínguez García, 2007; Fuentes, 2009; Martí Sánchez, 2008; Garcés Gómez, 2014), but in spite of this, the Spanish language has been understudied from the perspective of subjectivity (for noticeable exceptions, see Pit et al., 1996; Blackwell, 2016). This paper presents such a study.

In earlier studies of Spanish connectives, *porque* ('because'), *ya que* ('since') and *puesto que* ('given that') have been identified as prototypical connectives for expressing causality (Casado Velarde, 1998; Montolío, 2001; Domínguez García, 2007). *Porque* ('because') is the most common connective in Spanish (Vera, 1984; Domínguez García, 2007). Goethals (2002, 2010) indicates that *porque* ('because') differs from other connectives in Spanish because of its multifunctional properties: it is the only connective that can (i) fall in the scope of negation; (ii) be the focus of a question; (iii) be modified by any type of adverb; (iv) be included in a cleft sentence; and (v) introduce an answer to a WH-question. In sum, this causal connective seems to be like other subordinating causal conjunctions, like English *because*. Blackwell (2016) indicates that *porque* in oral narratives have either a semantic or a pragmatic[2] reading and, in some cases, they may have both. Given these features, it can be deduced that *porque* ('because') is indeed a general connective and, therefore, it can be expected that it is a connective used to express various types of causal relations, both objective and subjective.

The Spanish connective *ya que* ('since') has been characterized as a specific connective that expresses specialized uses. Some studies have claimed that it is used to justify or legitimize the speaker's acts, to express a collective voice, or to introduce information that is accepted by most of the people; it has also been identified as a connective that is associated with the presence of third singular or plural person (Borzi & Detges, 2011). Other studies have argued that it is a connective that refers to evident information in the context to justify other statements (Santos Río, 2003). Yet other studies have indicated that among its rhetorical functions *ya que* ('since') is specialized in expressing irony (Goethals, 2002, 2010). These specific features would suggest that

---

[2] Blackwell (2016) refers to semantic and pragmatic relations considering the distinction presented in Sanders (1997). These correspond to objective and subjective relations, respectively. See also Note 4.

ya que ('since') is a connective used preferentially for expressing subjective relations. Indeed, some studies have suggested this idea by claiming that *ya que* ('since') introduces speech acts of justification in which a conceptualizer expresses a subjective point of view, but without implying that this point of view is also the speaker's (Goethals, 2002, 2010). Pit et al. (1996) identified that *ya que* ('since') only occurred in epistemic relations and that the segments associated to this connective presented more evaluations than other connectives.

The connective *puesto que* ('given that') has also been identified as a connective that has specific meanings in that it introduces the justification of categorical judgments and deductions (Santos Río, 2003). Some authors have considered it as a connective that is associated with a proposal or suggestion (Galán Rodríguez, 1999). Others have stated that it is a connective that signals subjectivity, co-occurring generally with a speaker in an evaluative role (Pit et al., 1996). Therefore, these features lead us to infer that it is also used preferentially to express subjective relations.

In a recent corpus-based study (Santana et al., 2017), we explored the degree of subjectivity of Spanish connectives by carrying out automatic analyses. The results showed all three connectives occur with a relatively high frequency. However, the subjectivity profile observed in these connectives did not coincide with the claims in the literature discussed so far. We found that *porque* ('because') tends to occur in a subjective environment, i.e., in a context containing relatively many subjective words, which would suggest that it is used for expressing subjective relations. By contrast, *ya que* ('since') did not occur very often in a subjective environment. So, it is not necessarily a connective that expresses subjective relations. For *puesto que* ('given that') the results were mixed: whereas the segment preceding the connective (S1) was associated with many subjective words, the other segment (S2) contained a relatively low percentage of subjective words, which could indicate that this connective is used to express both subjective and objective relations. As these results (Santana et al., 2017) differ from the literature on Spanish connectives, this raises the issue of the semantic-pragmatic profile of the causal connectives. Are the differences of use between the three connectives so subtle that they do not turn up in an automatic analysis of subjectivity? For this reason, in the current study, we used another method to explore the profile of Spanish connectives: we carried out a manual analysis of corpus data.

Hence, the first research question this paper aims to answer is: *To what extent do Spanish connectives show a systematic variation in terms of subjectivity in a manual corpus analysis?* Our hypothesis is that *porque* ('because') is a general connective used for expressing both subjective and objective meanings; we also expect that *ya que* ('since') and *puesto que* ('given that') are specific connectives used predominantly to express subjective meanings. In our study, we annotated coherence relations using an analytical model of subjectivity that involves different subjectivity features analyzed in previous studies (Degand & Pander Maat, 2003; Sanders & Spooren, 2009, 2015; Spooren et al., 2010) (see Section 2 for more details).

So far, we have focused on coherence relations made explicit by connectives. However, in natural discourse, coherence relations often appear without a connective (Taboada, 2006, 2009; Taboada & Das, 2013); these are so-called implicit relations. Still, hardly any studies have concerned themselves with implicit relations (see Spooren, 1997; Taboada & Das, 2013; Das & Taboada, 2017 for noticeable exceptions), extensive research has concentrated on the study of explicit coherence relations. The main reason is that implicit relations are difficult to recognize (Lin et al., 2009), precisely because they do not show clear textual cues to indicate the relation. Therefore, little is known about the subjectivity of implicit relations. The past two decades have seen the availability of discourse annotation schemes like the Penn Discourse Treebank (PDTB, Prasad et al., 2008), the Rhetorical Structure Theory (RST) Treebank (Carlson, Marcu and Okurowski, 2003) and the Segmented Discourse Representation Theory (SDRT, Asher & Lascarides, 2003), which have allowed us to analyze explicit relations and their connectives, but also implicit relations. Most available resources contain English data, but gradually similar

resources have been created for other languages. The RST Spanish Treebank[3] is a good illustration of this (Cunha et al., 2011). This raises the possibility to compare the subjectivity profile of explicit relations to that of implicit relations. Now that the resources are available, we think it is very important to analyze implicit relations in order to better understand the notion of subjectivity. Therefore, the present study aims to fill this research gap by exploratively formulating the second research question: *What differences can be identified between explicit (with connectives) and implicit (no connectives) causal relations in terms of subjectivity?*

A third aspect that this study investigates is the influence of context in meaning and use of causal relations and their connectives. The issue of whether the semantic profiles of causal connectives can be characterized in terms of subjectivity as a stable characteristic, or whether this subjectivity is a context-dependent characteristic has become a debate within the study of connectives. Some studies have suggested that there is a relationship between context and type of connective or type of relation. For instance, Sanders et al., (1993) demonstrated that language users recognize objective versus subjective relations[4] easily when appropriate communicative contexts were provided. Sanders (1997) showed that the interpretation of ambiguous cases of coherence relations was strongly influenced by their descriptive versus argumentative context. Furthermore, a corpus study showed how objective relations were predominant in informative texts, whereas subjective relations were predominant in expressive and persuasive texts. Zufferey (2012) revealed that the distribution of French connectives varies according to different modalities of texts. In written texts, *car* ('because') is more often used for epistemic relations, whereas *parce que* ('because') prefers content relations. However, the situation varies dramatically in spoken texts since *car* is absent and *parce que* is used significantly more often in speech-act and epistemic relations. Zufferey et al., (2017) also provided empirical evidence that register is a distinguishing factor between connectives *car* and *parce que*. A final example is our previous study, in which we identified a significant relationship between the use of Spanish causal connectives and text type (informative versus persuasive/argumentative texts) in journalistic texts and between the use of Spanish causal connectives, text type (informative and persuasive/argumentative texts) and domain (Education and Psychology) in academic texts (Santana et al., 2017).

Other studies, however, have demonstrated how connectives have a robust semantic-pragmatic profile in terms of subjectivity, irrespective of the context in which they occur. For example, Sanders and Spooren (2015) found that Dutch connectives *omdat* ('because') and *want* ('since/for') showed a clearly different pattern, irrespective of the media used in the research, i.e. written texts, conversations and chat interactions. Zufferey (2012) also identified that the semantic-pragmatic profile of the French connective *puisque* ('since') was clearly subjective, and stable across written and spoken data. Apparently, the context is relevant in some cases depending on the text types and the connectives that are analyzed. This situation leads us to the third research question of this paper: *What is the relationship between contexts and the meaning and use of causal coherence relations?* Specifically, we will analyze the subjectivity of coherence relations in two different contexts[5]: journalistic and academic. These contexts differ in their audiences. While journalistic contexts are oriented to a broad audience (van Dijk, 1988; Waugh, 1995), academic contexts are more specialized, since they address a specific audience: academics (Bhatia, 2002, 2004; Hyland, 2009; Swales, 1990; Silver, 2006). We expect to identify whether

---

[3] Available from http://corpus.iingen.unam.mx/rst/corpus.html

[4] These relations are named "semantic" and "pragmatic" relations in the original papers (see Sanders et al., 1992 and 1993)

[5] On the one hand, the term "context" adopted in this paper refers to the source from which different text types were extracted to analyze connectives and coherence relations. On the other hand, the term "text type" refers to the classical distinction between informative and persuasive/argumentative texts.

the different contexts influence the meaning and use of causal coherence relations, that is, whether the frequency of causal subjective and objective relations depends on the context in which these coherence relations are produced. Our goal is to elucidate whether the realization of subjectivity in Spanish coherence relations is an inherent property of certain relations and their preferred connectives, or whether this is a matter of contextual dependency. In comparison with our previous study (Santana et al., 2017), we believe that a manual analysis will allow us to examine this issue with greater depth.

This paper is organized into four sections. First, we describe the analytical model of subjectivity applied in the manual analysis. Secondly, we present the method carried out in this study along with the results of the inter-rater agreement tests. Third, we present the results obtained after conducting different statistical analyses. Finally, the paper ends with the main discussion and conclusions.

2. An analytical model of subjectivity

In this paper, we analyze backwards causal relations in Spanish and we make use of the analytical model of subjectivity proposed by Sanders & Spooren (2015), which decomposes the general notion of subjectivity in its most relevant components. This model unites the subjectivity features introduced in earlier approaches: subjectivity is analyzed in terms of three key features. First, the reference to the speaker (Lyons, 1977; Traugott, 1995) and the implicit presence of the speaker (Langacker, 1990) are recognized as relevant components of subjectivity. For this reason, Modality of the Q-segment (the consequent in a coherence relation) and the Presence of the SoC are central variables of this analytical model. Secondly, considering that our analysis focuses on causal coherence relations, the nature of coherence relations should also be identified as a relevant aspect in the distinction of subjectivity (Sweetser, 1990; Sanders et al., 1992, 1993). Consequently, Domain is another essential variable included in this model, which allows us to classify four types of causal relations. Finally, assuming that the presence of the SoC plays a relevant role on the distinction of causal relations and that different mental spaces are linked to this SoC (Sanders et al. 2012), we estimate that the distinction between the author/speaker-subjectivity and character-subjectivity is also important to identify subjectivity. Thus, the Identity of the SoC is the fourth variable included in our analytical model. Previous studies have used all or some of these components to study coherence relations and connectives in other languages (Li, et al. 2013; Degand & Pander Maat, 2003; Sanders & Spooren, 2009, 2015; Spooren, et al. 2010). The current study does not propose a new operationalization of subjectivity. Table 1 presents the variables and the respective subjectivity values of this analytical model:

Variables	+……………….... Subjectivity values…………………………-		
Domain	Epistemic/Speech-act	Volitional content	Non-volitional content
Modality (Q)	Judgement/Speech-act	Mental fact	Physical fact
Presence of the SoC	Implicit	Explicit	Absent
Identity of the SoC	Author	Current speaker	Character

Table 1. The analytical model for subjectivity.

Every variable involves different indicators of subjectivity that represent two or more values. These distinctions correspond to different degrees of subjectivity (Traugott, 1995; Langacker, 1990).

2.1. Domain

The category domain operationalizes subjectivity in terms of the nature of causal relations (Sweetser, 1990; Sanders et al., 1992, 1993). It distinguishes between four types of causal relations in terms of domains (Sweetser, 1990). The content domain concerns causal relations that

occur in the physical world; this means that one event causes another in the "real world". This domain is divided into two subtypes: the volitional content domain that involves the presence of a SoC who performs an intentional act and the non-volitional content domain that does not implicate a SoC at all (Stukker, Sanders & Verhagen, 2008). The epistemic domain concerns causal relations in which a SoC is involved, to the effect that the events are related by the speaker's reasoning. Finally, the speech-act domain concerns causal relations in which the events are related by the speaker who performs a speech act.

As was mentioned previously, these relations indicate different degrees of subjectivity depending on the presence of the SoC: the non-volitional content is assumed as the least subjective domain since there is no SoC in the construction of the causal relation; volitional content is more subjective than non-volitional content because there is a SoC who performs an intentional act involved in the causal relation; epistemic domain is considered even more subjective because there is a SoC who is reasoning, inferencing, concluding, giving an opinion in the causal relation, and the speech-act domain is also a subjective domain because there is a SoC corresponding to the speaker who is performing the speech act involved in the causal relation. Thus, causal relations can be ordered from least subjective to most subjective, as follows:

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

The interpretation of these four domains can be facilitated by using a paraphrase test (Sanders, 1997; Sanders & Spooren, 2015; Li et al., 2013) which is presented in Table 2 (P and Q correspond to the antecedent and the consequent of the causal relation, respectively):

Domain	Paraphrase
Non-volitional content	P leads to the physical fact/mental fact that Q, and no intention is involved in Q
Volitional content	P leads to the intentional physical act/mental act that Q
Speech-act	P leads to question/advice/command/promise that Q
Epistemic	P leads to the claim/decision/inference/conclusion that Q

Table 2. Paraphrase test used in the analysis of Domain.

Applying this paraphrase test, example (4) is a non-volitional content relation because the avalanche leads to the physical fact that the road was blocked, and no intention is involved in this causal relation. Moreover, the interpretation of this meaning relation does not require reference to a specific SoC, which is absent. (5) is interpreted as a volitional content relation since feeling tired leads to the intentional physical act of going home. There is a SoC involved in the causal relation, the speaker, who made an intentional decision. (6) is a speech-act relation because the fact that the neighbors are not at home leads to the request to turn up the radio. The speaker motivates her request by reference to the absence of the neighbors. Finally, (7) is classified as an epistemic relation because the observation that the baby is crying leads to the speaker's conclusion that the baby is hungry[6].

(4) There had been an avalanche at Roger's pass. As a result, the road was blocked.
(5) I went home because I felt tired.
(6) Why don't you turn up the radio? The neighbors are not at home.
(7) The baby must be hungry, because it is crying.

---

[6] Example (4) is extracted from Pander Maat and Sanders (2001), (5) and (7) are extracted from Pander Maat and Sanders (2006), and (6) is extracted from Sanders and Spooren (2007).

2.2. Modality (Q)

Another possible indicator of subjectivity is in the modality of the segments connected. In our analysis we consider the modality of backward causal relations, i.e. relations in which the consequent (Q) corresponds to the first segment whereas the antecedent (P) to the second segment. Particularly, we focus on the first segment because the Q is the place where subjectivity can be most manifest in backward relations.

The modality or propositional attitude can be of different types (Sanders et al, 1992, 1993; Sanders & Spooren, 2015). In this study modality has four values: physical fact, mental fact, speech-act and judgement. The modality is a physical fact if the segment Q describes events or states that take place in the real world and can be observed in the physical world (4); it is a mental fact if the segment Q depicts mental states such as personal feelings, mental processes, or psychological activities (5); the modality is a speech-act if the segment Q is a general question, a rhetorical question, or an imperative construction (6) and finally the modality is a judgement if the segment Q presents an opinion, claim, conclusion or deduction (7).

These types of modality (Q) show different degrees of subjectivity, depending on the complexity involved in the causal relation (Sanders et al., 2012): physical fact indicates the lowest degree of subjectivity because it does not involve a mental process of a SoC. Mental fact is more subjective than physical fact since it involves a mental state of a SoC, but it is less subjective than speech-act and epistemic because these indicators require complex actions that must be carried out by a SoC like commanding an action, asking a question, formulating an opinion or drawing a conclusion, for instance. Thus, indicators of modality can be ordered from least subjective to most subjective, as follows:

Physical fact < Mental fact < Speech act/Judgement

The interpretation of these indicators of modality (Q) can be facilitated by using a paraphrase test, which is presented in Table 3:

Modality (Q)	Paraphrase
Physical fact	The fact/event/situation that Q
Mental fact	The decision/thinking/analysis/feeling that Q
Speech-act	Speaker questions/recommends/promises/commands here and now that Q
Judgement	SoC comes here and now to the conclusion that Q, and an evaluation is involved in Q

Table 3. Paraphrase test used in the analysis of Modality (Q).

Given this paraphrase test and the previous examples, the modality (Q) in example (4) is a physical fact since the fact that the road was blocked is observable in the real world. No mental process is carried out by a speaker. In (5), modality (Q) is mental fact because the volitional action of going home was a decision ("I decided to go home"). Therefore, a mental process carried by a speaker was necessary to make that decision. In (6), the modality (Q) is speech-act because the speaker asks a question here and now ("I ask you here and now to turn up the radio"). So, a mental process carried out by a speaker was mandatory to ask the action to another speaker. Finally, the modality (Q) in (7) is judgement because the claim that the baby is hungry is a conclusion in the here and now ("I here and now conclude that hunger must be the reason").

2.3. Presence of the SoC

The presence of the SoC distinguishes whether a SoC, if present, is explicitly referred to (Lyons, 1977; Langacker, 1990; Traugott, 1995). This variable has three values, which express different

degrees of subjectivity depending on the linguistic reference to the SoC: Absent SoC, Implicit SoC and Explicit SoC. Absent SoC is the least subjective because there is no SoC involved in the causal relation. Consider again the examples (4-7), repeated below. In (4) there is no SoC because there is no human intention or motivation in the causal relation. In (5), the presence of the SoC is explicit since there is an explicit reference to a SoC who decided to go home (the first-person pronoun *I*). In (6) and (7), the presence of the SoC is implicit because there is a SoC that performs a speech act and draws a conclusion, respectively, but there is no linguistic signal that provides information about these SoCs.

(4) There had been an avalanche at Roger's pass. As a result, the road was blocked.
(5) I went home because I felt tired.
(6) Why don't you turn up the radio? The neighbors are not at home.
(7) The baby is hungry, because it is crying.

Implicit SoCs are the most subjective indicator because there is a SoC involved in the causal relation but as there is no linguistic signal referring to that SoC, it is not part of the utterance and remains off-stage. Explicit SoC is more subjective than absent SoC because there is a SoC involved, but the explicit reference puts the SoC on-stage (Langacker, 1990), making it objective to a certain degree. Consequently, indicators of the presence of the SoC can be ordered from least subjective to most subjective, as follows:

Absent SoC < Explicit SoC < Implicit SoC

2.4. Identity of the SoC

The identity of the SoC is a variable that has three values: character, main author/speaker or current speaker (Sanders et al., 2012). The identity of the SoC is character when the SoC is someone other than the author/speaker. It is main author/speaker when it is the main speaker who establishes the causal relation. Finally, it is current speaker when the main speaker is presenting a causal relation for which s(he) is not responsible; generally, a secondary speaker is quoted. Those cases in which no SoC is identified have been coded as non-applicable (N/A).

The identity of the SoC in (8) is character because the basketball player is the actor who volitionally apologized to the authorities. It is a third person actor. In (9), the identity of the SoC is the author because there is a SoC who is offering his/her opinion and it is the main speaker. (S)he is establishing the causal connection between both segments. Finally, the identity of the SoC in (10) is current speaker 'Primo Levi' because an author (in this case a quoted speaker, the public prosecutor) is quoting the causal construction made by Primo Levi and quotation marks indicate what he said.

(8) The basketball player Federico Kammerichs, from Pamesa Valencia, apologized to the authorities of the Spanish club because last Friday he returned to Argentina without permission.
(9) There should not be "classes" in a democracy since the title per se implies inequality.
(10) The public prosecutor has pointed out that "those words were not pronounced by a common person and they have a decisive influence". "I remind you what Primo Levi wrote: 'We have a responsibility while we live. We must answer for everything we write, word for word because every word leaves a mark," the public prosecutor has said to a downcast De Luca.

The different degrees of subjectivity of these indicators are associated with the distance between the SoC and the main speaker: character is the least subjective identity because it corresponds to the responsible actor for the causal relation, who differs from the main speaker;

the current speaker is more subjective than character since it corresponds to a SoC who constructs the causal relation, but it is less subjective than the author because the SoC is referred to explicitly by the main speaker; author is the most subjective indicator because it corresponds to the SoC who constructs the causal relation and is the main speaker of the causal relation. In this way, indicators of the identity of the SoC can be ordered from least subjective to most subjective, as follows:

N/A < Character < Current speaker < Author

We have now identified the four subjectivity features with their indicators, which constitute the analytical model that has been applied in the manual analysis of coherence relations. In this analysis, each variable is considered a separate variable, even though various empirical studies using this model (Li et al., 2013; Spooren et al, 2010; Sanders & Spooren, 2015) have revealed a correlation between some variables, such as presence of the SoC and domain of the relations. For instance, non-volitional relations typically do not have a SoC. Still, we think it is important to analyze corpora in terms of these variables separately, because it is very well possible that subjectivity is linguistically expressed differently in one language than in another. In the section that follows, the method used in this study will be described.

3. Method
Different methodological steps were carried out: sampling of the corpus, data analysis and processing and analyses establishing inter-rater agreement. In this section, this information will be described in detail.

3.1. Corpus
Six data sets with backward causal coherence relations were constructed, four of them corresponding to causal explicit relations and the other two to causal implicit relations. The sets for explicit relations contain relations marked by the three connectives of interest: *porque* ('because'), *ya que* ('since') and *puesto que* ('given that'), which were extracted from different corpora[7]. The first and second set correspond to examples selected randomly from the academic and journalistic corpus of Spanish constructed in our previous corpus-based study (Santana et al., 2017); specifically, the fragments were extracted from essays, research articles and textbooks in Education and Psychology, and from editorials and news, respectively. The third and fourth set contain examples selected randomly from the CORPES XXI[8], which is a freely available online corpus created by the Royal Spanish Academy (RAE in Spanish). We decided to use this resource instead of others because it enabled us to select the required number of cases for each connective. For example, in the case of RST Spanish Treebank, we explored the distribution of those relations that could be considered as causal relations (*cause*, *justification*, *result*, *motivation* and *evidence*) and we noticed that the frequencies of the connectives of interest were restricted to few cases. Moreover, CORPES XXI allowed us to consult cases of these connectives using different search parameters. The search parameter for the third set was 'academic texts'; the fragments stem mainly from textbooks, research articles, essays, proceedings, theses and research reports from different disciplines. For the fourth set, the search parameters were 'news', 'editorials' and 'opinion articles'; the resulting fragments stem from these journalistic text types. Table 4 shows the distribution of these explicit relations:

---

[7] The random selections for the first, second, third and fourth set of causal coherence relations were carried out using a sequence generator available from http:www.random.org/

[8] Available from http://web.frl.es/CORPES/view/inicioExterno.view

Connective	Academic sub-corpus	Journalistic sub-corpus	Academic CORPES XXI	Journalistic CORPES XXI	Total
Porque ('because')	30	30	30	30	120
Ya que ('since')	30	30	30	30	120
Puesto que ('given that')	30	30	30	30	120
Total	90	90	90	90	360

Table 4. Distribution of explicit causal relations selected from each corpus.

Table 4 shows an identical frequency of examples for every connective. However, originally it was not possible to extract enough cases for *ya que* ('since') and *puesto que* ('given that') from the academic and journalistic corpus of Spanish[9] (Santana et al., 2017) since their frequencies did not reach 30. Given the purpose of describing the semantic-pragmatic profile of Spanish connectives, it was necessary to obtain the same number of examples per connective. To achieve this purpose, new texts were added to the corpus considering the same sources settled in its original construction: one essay of Psychology, one essay of Education, one research article of Psychology, ten editorials and ten news texts.

A specific segmentation was considered for each selected case: previous context (C1), which is the clause[10] that precedes the causal relation; segment 1 (S1), which is the clause that functions as Q (consequence) in the causal relation; the causal connective (*porque,* 'because', *ya que* 'since'*, puesto que* 'given that'); segment 2 (S2), which is the clause adjacent to S1 and functions as P (cause) in the causal relation; finally, posterior context (C2), which is the clause that follows the causal relation established between S1 and S2. Analysts analyzed exclusively the causal relation that exists between S1 and S2; C1 and C2 were presented just to provide more information to the analysts, which could lead to a better classification. Fragment (11) illustrates the examples included in the data sets of explicit relations:

(11) **(C1)** Desde 2012 cuando se superaron los tres millones de peregrinos, Arabia Saudí ha limitado el número de asistentes debido tanto a las controvertidas obras que se realizan en la Gran Mezquita de La Meca como al temor a epidemias. **(S1)** Pero se trata de una medida temporal, **ya que (S2)** el objetivo de los trabajos es ampliar la superficie de la aljama en 400.000 metros cuadrados, **(C2)** para que pueda acoger hasta 2,2 millones de fieles a un tiempo.
'**(C1)** Since 2012 when there were more than three million of pilgrims, Saudi Arabia has limited the number of attendees because of the controversial works performed in the Grand Mosque of Mecca and the fear of epidemics. **(S1)** But it is a temporary measure, **since (S2)** the purpose of the works is to expand the area of the mosque in 400,000 square meters **(C2)**, so that it can accommodate up to 2.2 million of faithful at a time.'

As can be noticed in (11), the clause was considered as the minimal unit of analysis for S1 and S2. However, we also considered the propositional content, this is the meaning that underlies every segment, and which allows us to construct a mental representation of the causal relation. Thus, the selected examples vary in terms of clauses, some of them containing only one clause in

---

[9] The frequencies of backwards cases of *porque* ('because'), ya que ('since') and *puesto que* ('given that') in academic texts were 98, 106 and 31, respectively (273,359 words); and in journalistic texts, the frequencies were 79, 22 and 14, respectively (175,466).

[10] 'Clause' was defined here as "a structure that contains a predicate headed by a main verb".

S1 and S2, respectively, as in (11), and others containing more than one clause in S1 or S2. In (12), S1 consists of a complex of hierarchically ordered clauses (indicated by lowercase 'c'), but *ya que* ('since') is connecting two independent and adjacent segments, which was the main criterion assumed for all the selected examples:

(12) **(S1)** Llama la atención la forma [c1 como lograron [c2 ascender en la sociedad colombiana [c3 a pesar de no encontrar un marco institucional [c4 que favoreciera su integración debido a políticas restrictivas hacia los extranjeros [c5 cuya procedencia no fuera europea c5] c4] c3] c2] c1], **ya que** **(S2)** la postura gubernamental era recibir a manos llenas la influencia del viejo continente a través de sus hijos. **(C2)** Sin embargo las leyes, aunque abundantes en este campo, resultaban ineficientes a pesar de haberse redactado varios proyectos, pues las cifras de inmigrantes europeos que decidieron establecerse en el país durante la emigración masiva fueron mínimas por el poco atractivo económico que revestía la nación.

'**(S1)** It is striking [c1 how they managed [c2 to ascend in Colombian society [c3 in spite of the fact that they did not find an institutional framework [c4 that favored their integration due to restrictive policies towards foreigners [c5 whose origin was not European c5] c4] c3] c2] c1], **since** **(S2)** the governmental position was to receive generously the influence of the old continent through their children. **(C2)** However, the laws, although abundant in this field, were inefficient in spite of the fact that several projects had been drafted, since the numbers of European immigrants who decided to settle in the country during the mass emigration were minimal because of the lack of economic attractiveness of the nation.'

Regarding the sets for implicit relations, the cases were extracted from the academic and journalistic corpus of Spanish constructed in our previous corpus-based study (Santana et al., 2017). To this effect, academic and journalistic text types were randomly selected from the corpus and examined to identify the implicit causal relations. This process was repeated until 60 implicit relations were identified for each context, which was identical to the number collected by each connective in the explicit sets (see Table 4). Table 5 shows the distribution of these implicit relations for every corpus[11]:

	Academic sub-corpus	**Journalistic sub-corpus**	**Total**
Implicit causal relations	60	60	120

Table 5. Distribution of implicit causal relations selected from each corpus.

Implicit relations were segmented in the same way as explicit relations, with the exception that these relations did not contain a connective linking S1 and S2. Fragment (13) illustrates the examples included in the data sets of implicit relations:

(13) **(C1)** 3.2. Concepto de cine de animación
    (S1) El cine es una ilusión. **(S2)** Las imágenes en movimiento no existen. **(C2)** Se necesita la existencia de 24 imágenes fijas por segundo para crear esa sensación de movimiento.
    '**(C1)** 3.2. Concept of animation cinema

---

[11] 84 academic texts (180,955 words) and 215 journalistic texts (100,647 words) were analyzed to identify 120 examples.

(S1) Cinema is an illusion. **(S2)** Moving images do not exist. **(C2)** The existence of 24 still images per second is required to create that feeling of movement.

3.2. Manual analysis

The manual text analysis consisted of coding the data sets described in Section 3.1 by evaluating the variables of the analytical model explained in Section 2. This task was carried out initially by three native speakers of Spanish, one of the authors of this paper and two collaborators. All of them work in the field of discourse studies at different universities. The analysis proceeded in three phases. The first was the 'warming-up phase' aimed at training the analysts. In this phase, a protocol of analysis was presented (see Appendix 1), containing information about the purpose of the analysis, explaining the categories of the analytical model and the procedures that should be carried out. The analysts also examined a set of examples (20 cases per each one) in order to practice with the categories and discuss problematic cases. These examples were extracted from the same corpora used for the creation of the data sets. Four cases corresponded to implicit relations, six to relations marked by *porque* ('because') relations, five to relations marked by *ya que* ('since'), and five to relations marked by *puesto que* ('given that'). Table 6 illustrates one of these examples:

Item	Example	Domain	Modality	Presence of the SoC	Identity of the SoC	Observations
17	**(C1)** Ciertamente, el concepto de deconstrucción no era empleado por Aguilar indistintamente; como se anotó con anterioridad, en la experiencia con las obras debía sugerirse el modulador adecuado para su apreciación y comprensión. **(S1)** Por otra parte, las connotaciones particulares que el concepto cobró en el discurso crítico de Aguilar están íntimamente relacionadas con la especificidad de las obras que fueron observadas mediante esta plataforma conceptual, **puesto que (S2)** Aguilar se remitió a la deconstrucción como un concepto táctico para explicar ciertas características de orden formal. **(C2)** Existen dos artículos clave para ilustrar este fenómeno: "Construcciones privadas en la U.N.: Armar y desarmar en un lienzo" y "Salas Silva en El Museo: Un exquisito 'déjà vu'". '**(C1)** Certainly, the concept of deconstruction was not used by Aguilar interchangeably; as noted previously, in the experience with the works, the appropriate modulator should be suggested for its appreciation and understanding. **(S1)** On the other hand, the particular connotations that the concept took in the critical Aguilar's speech are intimately related to the specificity of the works that were observed through this conceptual platform **given that (S2)** Aguilar referred to deconstruction as a tactical concept to explain certain	Epistémico 'Epistemic'	Juicio? 'Judgement?'	Implícito 'Implicit'	Autor 'Author'	Dominio: en casos como este me cuesta determinar si es realmente un estado de cosas del mundo físico observable o si se trata de una interpretación, juicio, valoración, conclusión del autor. Tiendo a creer que es epistémico, ya que con el parafraseo no suena lógica una interpretación no volitiva 'Domain: in cases like this one, it is hard for me to determine if it is a state of the physical and observable world or if it is an author's interpretation, judgement, evaluation or conclusion. I tend to think that it is epistemic since by using the paraphrase test a volitional interpretation does not sound logical.'

formal characteristics. **(C2)** There are two key articles to illustrate this phenomenon: "Construcciones privadas en la U.N.: Armar y desarmar en un lienzo" and "Salas Silva en El Museo: Un exquisito 'déjà vu".'					

Table 6. Example of codebook used in the manual text analysis.

Table 6 shows the elements that were included in the codebook of the manual text analysis. In the *Example* column, the case to be analyzed was presented, in the *Domain*, *Modality*, *Presence of the SoC* and *Identity of the SoC* columns, annotators identified the values corresponding to such variables and in the *Observation* column, annotators included a comment, question or doubt related to each case. In this way, every case was discussed between annotators, and they agreed on how to analyze the examples, especially the most problematic cases.

The second phase was the preliminary manual analysis of data sets, which focused on applying the analytical model and clarifying doubts about the categories. This phase was considered important because it is assumed that the quality of coding improves over time (Spooren & Degand, 2010). Thus, we ensured the correct interpretation of the variables and the codebook before carrying out the official analysis. In this phase, the type of analysis that was carried out is called 'partial overlap coding', which implies that the sample is cut up in several subparts of which some are double coded, while others are coded by only one analyst (Spooren & Degand, 2010; Li, et al., 2013). Considering that we counted on three analysts and our corpus contained 480 examples distributed in explicit and implicit relations, 15% of the sample was coded by analyst 1 (72 cases); another 15% was coded by analyst 2 (72 cases); and finally, all the examples evaluated by analyst 1 and 2 (144 cases) were coded by analyst 3. These examples were similar to those illustrated in Table 6. After carrying out this preliminary round of analysis, several disagreements were identified between annotators, conflicting cases were discussed and criteria for the official analysis were agreed on. This discussion was carried out separately between the analysts, so the examples analyzed by analysts 1 and 2 were not shared between them; analyst 3 was the only one who had access to all the examples.

The third phase corresponded to the final manual analysis of the data. Here we used the same type of analysis 'partial overlap coding' but unlike the second phase, the analyses were done with two rather than three analysts. In this way, one analyst examined 15% of the sample (72 cases), which corresponded to different examples to those given in the preliminary round of analyses and the other analyst examined all the examples of the sample (480 cases). These examples were comparable to those presented in the previous phases (see Table 6). In total, considering all phases of the analysis process (the warming-up phase, the preliminary manual analysis, and the final manual analysis of the data), three rounds of annotations were carried out. Different, ambiguous and clear examples were discussed in the first and second phases, so the annotator who analyzed most of the sample (480 cases) in the third phase, followed the criteria established as a result of previous annotations. Therefore, annotations in the third phase do not show as much variation as the annotations carried out in the previous phases. This method ensured that at least 15% of the data was double-coded, it also enabled us to use the information of double-coded data in order to conduct statistical analyses and it made possible to enhance the reliability of data coded by one analyst.

3.3. Inter-rater agreement
The level of inter-rater agreement was calculated taking into account the double-coded data collected in the third phase of our analysis, which correspond to 72 cases of the sample (15% of the total analyzed corpus). This selection consisted of 36 cases extracted from the academic context and 36 cases from the journalistic context. All of them were extracted randomly.

Regarding the cases of the academic context: five essays of Education, two essays of Psychology, six research articles of Education, three research articles of Psychology, four textbooks of Education and four textbooks of Psychology were extracted from the academic corpus of Spanish (Santana et al., 2017) and four research articles, three textbooks, two essays, one proceeding, one research report and one thesis were obtained from CORPES XXI. In relation to the cases of the journalistic context: twelve editorials and eight news were extracted from the journalistic corpus of Spanish (Santana et al., 2017) and seven editorials, five news and four opinion articles were collected from CORPES XXI. Table 7 shows the inter-rater agreement:

Variable	Percentage of agreement	Cohen´s Kappa	N Cases
Domain	90.3%	0.68	72
Modality (Q)	65.3%	0.42	72
Presence of the SoC	93.1%	0.84	72
Identity of the SoC	94.4%	0.85	72

Table 7. Inter-rater agreement per variable among annotators.

It can be seen from the data in Table 7 that the percentage of agreement was over 90% in three of the variables. Moreover, the indicator of Cohen's Kappa shows that the agreement was excellent for the identity of the SoC and for the presence of the SoC (following the interpretation rules for Kappa suggested by Landis and Koch, 1977). In the case of the domain, the level of agreement can be labeled substantial. The agreement was low in case of modality (Q), which reveals that it is the most controversial category of the analytical model. For further detail, Tables 8, 9, 10 and 11 provide the confusion matrices of each subjectivity feature[12]:

Domain	Analyst 3			
Analyst 1	Epistemic	Speech-act	Volitional	Non-volitional
Epistemic	56	3	1	0
Non-volitional	1	0	0	1
Volitional	2	0	8	0

Table 8. Confusion matrix of Domain.

Modality (Q)	Analyst 3			
Analyst 1	Judgement	Speech-act	Mental fact	Physical fact
Judgement	30	0	0	0
Speech-act	2	3	0	0
Mental fact	17	0	13	0
Physical fact	0	0	6	1

Table 9. Confusion matrix of Modality (Q).

---

[12] Matrices are calculated on the basis of the total number of double-coded data, i.e. 72 cases of the sample (15% of the total analyzed corpus).

Presence of the SoC	Analyst 3		
Analyst 1	Implicit	Explicit	Absent
Implicit	48	2	0
Explicit	3	18	0
Absent	0	0	1

Table 10. Confusion matrix of the Presence of the SoC.

Identity of the SoC	Analyst 3			
Analyst 1	Author	Current speaker	Character	N/A
Author	54	0	0	0
Current speaker	1	1	0	0
Character	1	2	12	0
N/A	0	0	0	1

Table 11. Confusion matrix of the Identity of the SoC.

Regarding the disagreements in modality (Q), most of them corresponded to cases that were classified as judgement by one analyst and mental fact by the other (see Table 9). This situation illustrates the difficulty of distinguishing these indicators, especially because both involve mental processes that are not directly observable in the external world.

Once we obtained the inter-rater agreement, statistical analyses were conducted. To investigate the distribution of subjectivity features, log-linear analyses were used specifically to identify associations and interactions between variables (subjectivity, linguistic marking, context), which in case of three-way interactions were followed up by chi-square tests (following the strategy suggested by Field, 2012). In the section that follows, this data will be presented in detail.

4. Results

The present study aims to investigate whether Spanish connectives show a systematic variation in terms of subjectivity in a manual corpus analysis, whether explicit and implicit relations differ with respect to subjectivity, and whether these relationships depend on context. As was mentioned in Section 1, we assumed academic and journalistic as sources of context. In this section, the results are reported for each subjectivity feature and the observed data correspond to the analyses carried out by one of the annotators who analyzed all the 480 cases in the corpus.

As indicated previously, log-linear analyses were carried out with the purpose of identifying associations and interactions between variables (subjectivity, linguistic marking, context). To achieve this aim, it was necessary to collapse some categories. This decision was made as a solution to the problem of having more than 20% of cells with expected frequencies less than 5, which violates one of the assumptions in the log-linear analysis (see Field, 2012). Grouping was done taking into account categories that have similar degrees of subjectivity. Table 12 illustrates the organization of categories:

Variables	Subjective	Objective
Domain	Epistemic/Speech-act	Volitional content/Non-volitional content
Modality (Q)	Judgement/Speech-act	Mental fact/Physical fact
Presence of the SoC	Implicit	Explicit/Absent
Identity of the SoC	Author/Current speaker	Character/N/A

Table 12. Organization of categories for the statistical analyses.

Table 12 shows that in the case of the domain, speech-act and epistemic relations were grouped together as subjective relations, whereas volitional and non-volitional as objective. In the case of modality of the Q-segment, speech-act and judgement were gathered as subjective and mental fact and physical fact as objective indicators. In the case of the presence of the SoC, implicit SoC was considered subjective and explicit SoC and absent SoC were grouped as objective indicators. Finally, in the case of the identity of the SoC, author and current speaker were grouped as subjective cases, whereas character and N/A were considered as objective ones.

4.1. Domain

The log-linear analyses produced a final model that retained all effects. The fit of the model was perfect (χ^2 (0) = 0; p=1). This indicated that the highest order interaction of Domain * Context * Linguistic marking was significant (χ^2 (3) = 15.41; p <.01). To interpret this interaction separate chi-square tests were calculated between Domain and Linguistic marking for each source of context. For academic context, the association between Domain and Linguistic marking did not reach significance (χ^2 (3) = 6.49; p = .09). For journalistic context, the association was significant (χ^2 (3) = 16.71; p <.01). Analysis of the standardized residuals showed that the association is mainly caused by the relatively low number of objective indicators (volitional and non-volitional) for implicit relations, compared to the three types of explicit relations. The data are summarized in Table 13:

Domain	Subjective	Objective
Linguistic marking	**Academic context**	
Porque ('because')	46 (-0.7)	14 (1.6)
Ya que ('since')	51 (0.0)	9 (0.0)
Puesto que ('given that')	56 (0.7)	4 (-1.7)
Implicit	50 (-0.1)	10 (0.2)
	Journalistic context	
Porque ('because')	43 (-0.8)	17 (1.7)
Ya que ('since')	45 (-0.5)	15 (1.1)
Puesto que ('given that')	48 (-0.1)	12 (0.2)
Implicit	59 (1.5)	1 (-3.1)

Note: Subjective and objective correspond to the total number of the collapsed categories in the variable Domain, speech-act/epistemic and volitional/non-volitional, respectively.

Table 13. Frequencies (and standardized residuals) of Linguistic marking by Domain for each Source of Context (academic, journalistic).

Overall[13], 84.5% of the 240 relations in academic context were subjective (speech-act/epistemic) and 15.4% were objective (volitional/non-volitional). A similar predominance of subjective relations was found in journalistic context, 81.2% subjective (speech-act/epistemic) and 18.7% objective (volitional/non-volitional).

4.2. Modality (Q)

The log-linear analyses also produced a final model that retained all effects. The fit of the model was perfect (χ^2 (0) = 0; p=1). This indicated that the highest order interaction of Modality (Q) * Context * Linguistic marking was significant (χ^2 (3) = 9.71; p <.01). To interpret this interaction separate chi-square tests were calculated between Modality (Q) and Linguistic marking for each source of context. For academic context, the association between Modality (Q) and Linguistic

---

[13] All percentages presented in Section 4 are calculated on the basis of the total number of relations in each source of context (240 relations in academic contexts and 240 relations in journalistic contexts).

marking did not reach significance (χ^2 (3) = 2.87; p = .41). For journalistic context, the association was significant (χ^2 (3) = 14.41; p <.01). Analysis of the standardized residuals showed that the association is mainly caused by the low number of objective indicators (mental fact and physical fact) for the implicit cases, compared to the three types of explicit cases. The data are summarized in Table 14:

Modality (Q)	Subjective	Objective
Linguistic marking	**Academic context**	
Porque ('because')	40 (-0.7)	20 (1.1)
Ya que ('since')	45 (0.1)	15 (-0.1)
Puesto que ('given that')	48 (0.5)	12 (-0.9)
Implicit	45 (0.1)	15 (-0.1)
	Journalistic context	
Porque ('because')	40 (-0.6)	20 (0.9)
Ya que ('since')	39 (-0.7)	21 (1.2)
Puesto que ('given that')	41 (-0.4)	19 (0.7)
Implicit	55 (1.7)	5 (-2.8)

Note: Subjective and objective correspond to the total number of the collapsed categories in the variable Modality (Q), speech-act /judgement and mental/physical facts, respectively.

Table 14. Frequencies (and standardized residuals) of Linguistic marking by Modality (Q) for each Source of Context (academic, journalistic).

Overall, 74.1% of the 240 Q-segments in academic context had a subjective modality (speech-act/judgement) and 25.8% had an objective modality (mental/physical facts). The same predominance of subjective modalities was also found in journalistic context, 72.9% subjective modality, 27.0% objective modality.

4.3. Presence of the SoC

Similar to Domain and Modality (Q), the log-linear analyses produced a final model that retained all effects. The fit of the model was perfect (χ^2 (0) = 0; p=1). This indicated that the highest order interaction of Presence of the SoC * Context * Linguistic marking was significant (χ^2 (3) = 27.93; p <.01). To interpret this interaction separate chi-square tests were calculated between Presence of the SoC and Linguistic marking for each source of context. For academic context, the association between Presence of the SoC and Linguistic marking was significant (χ^2 (3) = 9.16; p <.01). For journalistic context, the association was also significant (χ^2 (3) = 29.46; p <.01). Analysis of the standardized residuals showed that the association in academic context is mainly caused by the relatively low number of objective cases (explicit and absent SoCs) for explicit relations marked with *puesto que* ('given that'), compared to the explicit relations marked by the other connectives and implicit relations. In the case of journalistic context, the association is mainly caused by the high number of subjective indicators (implicit SoC) and the low number of objective indicators (explicit and absent SoCs) for implicit relations, compared to the three types of explicit relations. The data are summarized in Table 15:

Presence of the SoC	Subjective	Objective
Linguistic marking	**Academic context**	
Porque ('because')	44 (-0.6)	16 (1.2)
Ya que ('since')	46 (-0.3)	14 (0.6)
Puesto que ('given that')	56 (1.2)	4 (-2.3)
Implicit	46 (-0.3)	14 (0.6)

	Journalistic context	
Porque ('because')	35 (-0.1)	25 (1.5)
Ya que ('since')	34 (-1.2)	26 (1.7)
Puesto que ('given that')	39 (-0.4)	21 (0.6)
Implicit	58 (2.6)	2 (-3.8)

Note: Subjective and objective correspond to the total number of the collapsed categories in the variable Presence of the SoC, implicit SoC and explicit/absent SoC, respectively.

Table 15. Frequencies (and standardized residuals) of Linguistic marking by Presence of the SoC for each Source of Context (academic, journalistic).

Overall, 80.0% of the 240 relations in academic context had a subjective SoC (implicit SoC) and 20.0% were objective (explicit/absent SoC). The percentages for journalistic context were 69.2% subjective and 30.8% objective.

4.4. Identity of the SoC

As with the previous variables, the log-linear analyses produced a final model that retained all effects. The fit of the model was perfect (χ^2 (0) = 0; p=1). This indicated that the highest order interaction of Identity of the SoC * Context * Linguistic marking was significant (χ^2 (3) = 19.94; p <.01). To interpret this interaction, separate chi-square tests were calculated between Identity of the SoC and Linguistic marking, for each source of context. For academic context, the association between Identity of the SoC and Linguistic marking was significant (χ^2 (3) = 10.67; p <.01). For journalistic context, the association was also significant (χ^2 (3) = 21.69; p <.01). Analysis of the standardized residuals showed that the association in academic context is mainly caused by the low number of objective cases (character and N/A) for explicit relations marked with *puesto que* ('given that'), compared to the explicit relations marked by the other connectives and implicit relations. In journalistic contexts, the association is mainly caused by the high number of subjective indicators (author and current speaker) and the low number of objective ones (character and N/A) for implicit relations, compared to the three types of explicit relations. The data are summarized in Table 16:

Identity of the SoC	Subjective	Objective
Linguistic marking	**Academic context**	
Porque ('because')	49 (-0.4)	11 (1.1)
Ya que ('since')	48 (-0.6)	12 (1.4)
Puesto que ('given that')	59 (1.0)	1 (-2.5)
Implicit	52 (0.0)	8 (0.0)
	Journalistic context	
Porque ('because')	38 (-0.1)	22 (1.7)
Ya que ('since')	40 (-0.7)	20 (1.2)
Puesto que ('given that')	43 (-0.3)	17 (0.4)
Implicit	58 (2.0)	2 (-3.4)

Note: Subjective and objective correspond to the total number of the collapsed categories in the variable Identity of the SoC, author/current speaker and character/N/A, respectively.

Table 16. Frequencies (and standardized residuals) of Linguistic marking by Identity of the SoC for each Source of Context (academic, journalistic).

Regarding the total percentages, 86.6% corresponded to subjective cases (author/current speaker) and 13.3% to objective ones (character/N/A) in academic context. In the case of

journalistic context, 74.5% corresponded to subjective indicators (author/current speaker) and 34.0% to objective indicators (character/N/A).

5. Discussion and conclusions

The current study aimed to achieve a better understanding of subjectivity in Spanish causal relations and connectives by carrying out manual text analyses. Evidence of several corpus-based studies in different languages (Pander Maat & Degand, 2001; Pander Maat & Sanders, 2000, 2001; Degand, 2001; Verhagen, 2005; Pit, 2006, 2007; Stukker & Sanders, 2009; Sanders & Spooren, 2015; Günthner, 1993; Keller, 1995; Wegener, 2000; Pit, 2007; Stukker & Sanders, 2012; Degand & Pander Maat, 2003; Zufferey, 2012; Li, et al., 2013; Wei, et al., 2018) have demonstrated that causal connectives follow a specific pattern in terms of subjectivity: some causal connectives are mainly used to express subjective meaning, whereas others are used to express objective meanings. A priori, Spanish might be expected to follow such a pattern. However, our previous corpus-based study (Santana et al., 2017), in which automatic analyses were carried out, did not reveal this. This situation led us to question whether the results were due to the method of automatic analysis of subjectivity. That is why we decided to use another method to shed light on this matter: manual analyses of subjectivity. Local contexts of Spanish causal explicit relations marked by the connectives *porque* ('because'), *ya que* ('since') and *puesto que* ('given that') were analyzed in order to investigate whether these connectives show a systematic variation in terms of subjectivity. Based on previous literature of Spanish connectives (Borzi and Detges 2011; Goethals, 2002, 2010; Pit et al., 1996; Blackwell, 2016), we expected to find that *porque* ('because') is a general connective used to express subjective and objective relations, whereas *ya que* ('since') and *puesto que* ('given that') are specific connectives used to express subjective meanings.

A second issue concerned implicit relations. Most available studies on coherence relations have investigated the subjectivity of explicit relations; hardly any investigations exist on the subjectivity of implicit relations. Therefore, the local contexts of implicit relations were examined in order to identify differences between this type of relations and explicit relations. Our purpose was to explore whether subjectivity occurs differently in the absence of connectives. These relations were analyzed in different contexts (academic and journalistic) in order to identify whether the subjectivity of these relations depends on the context in which they occur or whether this subjectivity is stable across these contexts. Log-linear analyses were conducted to identify associations and interactions between variables (subjectivity, linguistic marking, context) and chi-square tests were applied to obtain a more fined-grained analysis of the interaction between variables.

The first research question was: *To what extent do Spanish connectives show a systematic variation in terms of subjectivity in a manual corpus analysis?* Results of this corpus-based study showed that there is no evidence for subjectivity as a categorical principle in Spanish causal connectives. The connectives observed did not reveal a systematic variation in terms of subjectivity. *Porque* ('because') and *ya que* ('since') were not associated with any subjectivity feature, which suggests that they do not have a profile that can be characterized in terms of more or less subjectivity. *Puesto que* ('given that') was associated with a low number of objective features, which could lead us to infer that it is not used to express objective meanings. However, this tendency was observed only in academic context and for only two features: the presence of the SoC (explicit and absent SoCs) and identity of the SoC (character and N/A). Therefore, we conclude that in comparison with other languages, the distinction between connectives expressing subjective versus objective relations does not occur so evidently in Spanish.

Our second research question was: *What differences can be identified between explicit (with connectives) and implicit (no connectives) causal relations in terms of subjectivity?* The results showed that implicit relations were associated with relatively low number of objective scores

across the analyzed variables (domain, modality (Q), presence of the SoC and identity of the SoC). Moreover, implicit relations were associated with high number of subjective indicators in two variables: presence of the SoC (implicit SoC) and identity of the SoC (author/current speaker). However, these results were observed only in journalistic context. In the case of academic context, implicit relations were not associated with any subjectivity feature. Therefore, these observed data lead us to conclude that implicit relations behave differently depending on the context that they occur in. More specifically, in journalistic context they tend to be non-objective and in the case of containing a SoC, implicit relations are used preferentially to express subjective meanings.

Our third research question was: *What is the relationship between contexts and the meaning and use of causal coherence relations?* Results revealed that there was a significant three-way interaction in all the observed variables, which would indicate that context is relevant for the relation between subjectivity and linguistic marking. By splitting the results per source of context and by carrying out a more fine-grained analysis, we identified that associations between subjectivity and linguistic marking were significant in all variables in journalistic context (domain, modality (Q), presence of the SoC and identity of the SoC), whereas they were significant only in two variables in the case of academic context (presence of the SoC and identity of the SoC). These results suggest that the relevance of context in the relation between subjectivity and linguistic marking is clearer in journalistic context than in academic context.

In conclusion, the three research questions of this study received an explicit, and sometimes surprising, answer. Regarding our first research question, the analyzed Spanish connectives do not show a systematic variation in terms of subjectivity as various other languages do (such as Dutch, German, French, Mandarin Chinese). *Porque* ('because'), *ya que* ('since') and *puesto que* ('given that') do not have a clear-cut semantic profile in terms of subjectivity. This conclusion is based on solid manual text analyses considering a substantial number of relations and including statistical evaluation. The question arises why our results differ from previous studies of Spanish connectives. For *ya que* ('since') was identified by Borzi and Detges (2011) as a connective that is associated with the presence of a third singular or plural person. Goethals (2002, 2010) indicates that *ya que* ('since') identifies a speech act of justification in which a conceptualizer expresses a particular or subjective point of view, but without implying that this point of view is also the speaker's. According to these studies, we would have expected that *ya que* ('since') would be related to some indicator of presence of the SoC and identity of the SoC, but our data did not reveal that. In the case of *puesto que* ('given that') Pit et al. (1996) claim that this connective signals subjectivity co-occurring generally with a speaker in an evaluative role, which would lead us to expect that *puesto que* ('given that') would be associated to judgement in modality (Q). However, in our study, this connective was not associated with subjective indicators at all.

These discrepancies could be attributed to the fact that our sample was selected randomly considering different texts per each source of context (essays, research articles, textbooks, proceedings, theses and research reports from different disciplines for academic context, and editorials, news and opinion articles for journalistic context), whereas texts that were analyzed in previous studies were selected more narrowly. For example, the corpus used by Borzi and Detges (2011) consisted of examples extracted from journalistic articles of Argentinian newspapers and from five essays that corresponded to three Argentinian authors. Similarly, Goethals (2002) used cases of *ya que* ('since') only from the section of politics of the newspaper *El País* and Pit et al. (1996) analyzed 25 cases for *ya que* ('since') as well as for *porque* ('because') and *puesto que* ('given that').

Furthermore, our results differed from previous studies because of the analytical model that was used. Where we used an integrative approach to subjectivity that decomposes the general notion of subjectivity in its most relevant components (Lyons, 1977; Langacker, 1990; Traugott 1995; Degand & Pander Maat, 2003; Sanders & Spooren, 2009, 2015; Spooren et al., 2010),

previous studies considered some specific aspects of subjectivity. Borzi and Detges (2011) described *ya que* ('since') as polyphonic connective, similar to the French connective *puisque,* assuming the proposal of Bourcier and Ducrot (1980) and they examined the position and the quality of the information associated to these connectives; Goethals (2002) described and classified Spanish causal conjunctions within a framework that combines elements of the semiotics and speech act philosophy. Finally, Pit et. al. (1996) analyzed causal relations considering the content/epistemic domain dichotomy proposed by Sweetser (1990) and applying a schema of perspective analysis. We do believe that subjectivity was operationalized in the current study in a detailed and precise way, with a model that has revealed systematic differences in other languages (Sanders & Spooren, 2015; Li et al, 2013). As a result, the conclusion on the lexicon of causal connectives in Spanish seems valid.

Taken together, these results provide new insights into the categorization of coherence relations in terms of subjectivity. We obtained empirical evidence that shows that the phenomenon of subjectivity is encoded differently across languages. Spanish, similar to English, does not have connectives that have a clear subjectivity profile. Rather, Spanish seems to have a wider repertoire of general causal connectives, like *because* (Sweetser, 1990; Ford, 1993; Knott & Sanders, 1998)*,* which can be used to express both subjective and objective relations. In this way, subjectivity in Spanish might be conveyed through other linguistic elements. Therefore, it would be necessary to look into other specific features in the context. For example, Levshina and Degand (2017) analyzed several contextual variables to identify objective and subjective meanings expressed by English *because* on the basis of uses associated to Dutch connectives *omdat* and *want*; the presence or absence of modal verbs, polarity (positive and negative clauses), semantic class of the verbal predicate (mental or social verbs), semantic class of the subject (animate, inanimate, no subject), tense of the finite predicate and voice of the finite predicate could be promising variables to explore in order to identify whether they are associated with subjective or objective meanings in Spanish.

As they have been understudied from this type of approach, further investigation into Spanish causal connectives and coherence relations is still worthwhile. A reasonable approach might be to use experimental methods, which could increase the reliability and validity of data collected so far, with manual annotations. Crowdsourcing experimentation seems to be a promising methodology. It allows us to obtain information based on the knowledge of native speakers avoiding possible biased annotations by analysts (Pusse, et al., 2016; Scholman & Demberg, 2017). Our future work will be oriented in this direction.

Regarding our second research question, implicit relations in journalistic context tend to be non-objective, i.e. they avoid expressing objectivity, and in the case of containing a SoC, implicit relations are used to express subjective meanings. By observing the cases of journalistic context in detail, we notice that many of these relations are like examples (14) and (15):

(14)   **(C1)** El Gobierno peruano ha negado esta acusación. **(S1)** <u>Sin duda</u>, la anunciada y no concretada rendición de Humala acabó con la paciencia del Gobierno. **(S2)** En un comunicado, a media tarde de ayer, el Ministerio de Interior deploró la actitud de Humala de no deponer las armas, tal como lo había anunciado el pasado domingo **(C2)** e invitó a la población que vive en las inmediaciones de la comisaría a abandonar temporalmente sus viviendas. [NE180]
'**(C1)** The Peruvian government has denied this accusation. **(S1)** <u>Undoubtedly</u>, the announced and unspecified surrender of Humala put an end to the patience of the government. **(S2)** In an announcement, yesterday afternoon, the Ministry of Interior deplored the attitude of Humala, who did not put down weapons as he had announced the last Sunday **(C2)** and it invited the population living in the vicinity of the police station to temporarily abandon their homes.'

(15)   **(C1)** El movimiento de la Casa Blanca y las sanciones aplicadas a siete altos funcionarios venezolanos marcan un cambio significativo en la actitud estadounidense ante los graves acontecimientos que se suceden en Venezuela y que se han acelerado desde que Nicolás Maduro asumió la presidencia hace casi dos años. **(S1)** La situación política y social en el país sudamericano <u>es casi insostenible</u>. **(S2)** La ineficacia en la gestión económica, junto a la violación por parte del Gobierno de derechos básicos, han llevado al país a un estado de penuria material y de tensión social que tienen difícil justificación.
(C2) Algunos de los dirigentes más importantes de la oposición se encuentran encarcelados.
'**(C1)** The movement of the White House and the sanctions applied to seven Venezuelan senior officials mark a significant change in the US attitude towards the serious events that are happening in Venezuela and that have increased since Nicolás Maduro assumed the presidency almost two years ago. **(S1)** The political and social situation in the South American country is <u>almost unsustainable</u>. **(S2)** The inefficiency in economic management and the violation of basic rights by the government have led the country to a state of material scarcity and social tension, which are difficult to justify.
(C2) Some of the most important leaders of the opposition are imprisoned.'

In both examples we observed that there is a SoC who is evaluating a situation. The epistemic stance marker *sin duda* ('undoubtly') in example (14) and the use of the adverb *casi* ('almost') and the adjective *insostenible* ('unsustainable') in (15) reflect this SoC's presence. It could be the case that by not having any mark to express causality, implicit relations need to make even clearer the relation established between segments and consequently, these are more preferred to express subjective relations like (14) and (15), which were the predominant ones in our corpus (subjective indicators speech-act/epistemic, speech-act/judgement, implicit SoC and author/current speaker reached more than 70% in both types of contexts, academic and journalistic). In academic context similar relations to (14) and (15) were identified, but the number of objective relations was less extreme than in journalistic context, which would explain the non-significance in this context.

It is important to stress that in the current study, implicit relations were analyzed from a subjectivity perspective. Implicit relations have been understudied in discourse coherence studies (but see Taboada & Das, 2013; Das & Taboada, 2017 for recent exceptions), which have always been biased towards explicitly marked relations. Our findings provide a preliminary perspective of how subjectivity is manifest in absence of connectives, which is a step forward into the knowledge of subjectivity in discourse. We identified that subjectivity is manifested differently in implicit relations, depending on contexts. For this reason, we strongly encourage exploring implicit relations in different languages and contexts in order to compare results and obtain more insights into subjectivity. Moreover, the exploration of these relations would be an interesting issue for future studies on reading comprehension and processing information, especially because there have been studies that claim that subjective relations are more difficult to process and that connectives serve as instructions that facilitate the processing of this type of relations (Canestrelli et al., 2013).

As for our third research question, the current study allowed us to confirm that context plays an influential role in the meaning and use of coherence relations. For instance, significant relationships were identified between subjectivity, linguistic marking and context. This implies that context is a variable that should be considered in the analysis of coherence relations and connectives. Especially, in languages that have been understudied or that do not have connectives with a clear-cut profile in terms of subjectivity, context might give more insights into the meaning and use of these relations and connectives. Clear as this may be as a general tendency, we have to

acknowledge that it is not easy to interpret specific interactions between context, subjectivity and linguistic marking in a straightforward way, because we are talking about three-way interactions.

For instance, the results suggest that the relevance of context in the relation between subjectivity and marking is clearer in journalistic context than in academic context, since this association was significant for all variables in journalistic contexts, whereas it was significant only for two variables in academic contexts. However, it is hard to relate this directly to the characteristics of journalistic versus academic contexts, in general. It is tempting to argue that in journalistic texts it is necessary to communicate precisely what kind of events happened, how they occurred, who participated in those events, in order to clarify if they correspond to events of the real world or if they represent the thoughts of someone in particular. However, this intuition is not in line with the results that we found, because we found many epistemic relations in which the author is clearly present. For academic texts we found a similar relationship, but only for explicit *puesto que*-relations. This may be because authors of academic texts are more oriented to a particular audience of academic readers. They want to make a point, to argue, or to demonstrate the relevance of their work (Hyland, 2001, 2011). These characteristics are in line with the subjectivity variables that reached statistical significance in our corpus, especially for epistemic relations in the case of the connective *puesto que* ('given that') (presence of the SoC and identity of the SoC). These findings illustrate that we are still in need of further specification of the role of the context.

Among the limitations of this study, we mention the size of the corpus. As this study focused on specific contexts, it was necessary to collect texts corresponding to these contexts searching several resources in Spanish, which are limited, and in some cases, a part of them are not available freely. Therefore, future analyses, especially those that involve specialized contexts, should take this methodological issue into account.

Finally, we believe that the methodology used in this research provided us with a better understanding of subjectivity of Spanish causal relations and connectives. In comparison with automatic analyses carried out in our previous corpus-based study (Santana et al., 2017), manual analyses allowed us to explore in more depth the local contexts of Spanish coherence relations revealing interesting insights about these relations and their connectives. The strategy of analysis used in this study played a crucial role in the inter-rater agreement. The discussion of disagreements in the warming-up phase and in the preliminary manual analysis were beneficial steps in our methodology. These two rounds of annotations allowed us to clarify doubts, define the categories and indicators even more precisely, which was beneficial to the reliability because the coding can be reproduced in other research contexts. Moreover, it was favorable for the validity because the coding reached a higher quality, which now provides us with a clearer picture of causal coherence in the Spanish language.

References

Nicholas Asher and Alex Lascarides (2003). *Logics of Conversation*. Cambridge University Press, Cambridge.

Claudia Borzi and Ulrich Detges (2011). Ya que, un marcador polifónico. In H. Aschenberg and Óscar Loureda Lamas (eds.), *Marcadores del Discurso: De la Descripción a la Definición:* 263–281. Iberoamericana Vervuert, Madrid.

Danièle Bourcier and Oswald Ducrot (1980). *Les Mots du Discours*. Éditions de Minuit, Paris.

Vijay Bhatia (2002). A generic view of academic discourse. In J. Flowerdew (ed.), *Academic Discourse*: 21–39. Routledge, New York.

Vijay Bhatia (2004). *Worlds of Written Discourse: A Genre-Based View*. Continuum, London.

Sara Blackwell (2016). *Porque* in Spanish Oral Narratives: Semantic *Porque*, (Meta) Pragmatic *Porque* or Both? In A. Capone and J.L. Mey (Eds.), *Interdisciplinary Studies in Pragmatics, Culture and Society*: 615–651. Springer, Heidelberg.

Antonio Briz (1998). *El Español Coloquial en la Conversación: Esbozo de Pragmagramática*. Ariel, Barcelona.

Antonio Briz (2000). *¿Cómo se Comenta un Texto Coloquial?* Ariel, Barcelona.

Anneloes Canestrelli, Willem Mak and 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.

Lynn Carlson, Daniel Marcu and Mary Ellen Okurowski (2003). Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In J. van Kuppevelt and R. Smith (Eds.), *Current Directions in Discourse and Dialogue*: 85-122. Kluwer, Dordrecht.

Manuel Casado Velarde (1991). Los operadores discursivos es decir, esto es, o sea ya saber en español actual: valores de lengua y funciones textuales. *Lingüística Española Actual*, 13(1): 87–116.

Iria da Cuhna, Juan Manuel Torres-Moreno and Gerardo Sierra (2011). On the development of the RST Spanish Treebank. In *Proceedings of the V Linguistic Annotation Workshop (LAW)* (pp. 1-10). Stroudsburg, PA, USA: Association for Computational Linguistics.

Debopam Das and Maite Taboada (2017). RST Signalling corpus: A corpus of signals of coherence relations. *Language Resources & Evaluation*, 52(1):149-184.

Liesbeth Degand (2001). *Form and Function of Causation: A Theoretical and Empirical Investigation of Causal Constructions in Dutch*. Peeters Publishers, Leuven.

Liesbeth Degand and Henk Pander Maat (2003). A contrastive study of Dutch and French causal connectives on the speaker involvement scale. In A. Verhagen and J. van de Weijer (eds.), *Usage-Based Approaches to Dutch. Lexicon, Grammar, Discourse:*175–199. LOT, Utrecht.

Noemí Domínguez García (2007). *Conectores Discursivos en Textos Argumentativos Breves*. Arco Libros, Madrid.

Andy Field, Jeremy Miles and Zoe Field (2012). *Discovering Statistics Using R*. SAGE Publications Ltd., London.

Celia Ford (1993). *Grammar in Interaction: Adverbial clauses in American English Conversations*. Cambridge University Press, Cambridge.

Catalina Fuentes (2009). *Diccionario de Conectores y Operadores del Español*. Arco Libros, Madrid.

Carmen Galán Rodríguez (1999). La subordinación causal y final. In I. Bosque and V. Demonte (eds.), *Gramática Descriptiva de la Lengua Española*: 3597-3642. Espasa Calpe, Madrid.

María del Pilar Garcés Gómez (2014). *Diacronía de los Marcadores Discursivos y Representación en un Diccionario Histórico* (Anexos de revista de lexicografía, 28). Universidade da Coruña, A Coruña.

María del Camino Garrido Rodríguez (2004). *Conectores Contraargumentativos en la Conversación Coloquial.* Secretariado de Publicaciones y Medios Audiovisuales de la Universidad de León, León.

Patrick Goethals (2002). *Las Conjunciones Causales Explicativas Españolas Como, Ya que, Pues y Porque: Un Estudio Semiótico-lingüístico.* Peeters Publishers, Leuven.

Patrick Goethals (2010). A multi-layered approach to speech events: The case of Spanish justificational conjunctions. *Journal of Pragmatics*, 42(8): 2204–2218.

Susanne Günthner (1993). «... weil—man kann es ja wissenschaftlich untersuchen». Diskurspragmatische Aspekte der Wortstellung in WEIL-Sätzen. *Linguistische Berichte*, 143: 37–59.

Ken Hyland (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20(3): 207–226.

Ken Hyland (2009). *Academic Discourse: English in a Global Context.* Continuum, London.

Ken Hyland (2011). Disciplines and discourses: Social interactions in the construction of knowledge. In D. Starke-Meyerring, A. Paré, N. Artemeva, M. Horne and L. Yousoubova (eds.), *Writing in the Knowledge Society*: 193-214. Parlor Press and The WAC Clearinghouse, West Lafayette, Indiana.

Rudi Keller (1995). The epistemic weil. In D. Stein and S. Wright (eds.), *Subjectivity and Subjectivisation: Linguistic Perspectives*: 16–30. Cambridge University Press, Cambridge.

Alistair Knott and Ted Sanders (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, 30(2): 135-175.

Richard Landis and Gary Kosh (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1): 159-174.

Ronald Langacker (1990). Subjectification. *Cognitive Linguistics*, 1(1): 5-38.

Natalia Levshina and Liesbeth Degand (2017). Just because: In search of objective criteria of subjectivity expressed by causal connectives. *Dialogue & Discourse,* 8(1): 132-150.

Fang Li, Jacqueline Evers-Vermeul and Ted Sanders (2013). Subjectivity and result marking in Mandarin. *Chinese Language & Discourse*, 4(1): 74–119.

Ziheng Lin, Min-Yen Kan and Hwee Tou Ng (2009). Recognizing implicit discourse relations in the Penn Discourse Treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)* (pp. 343–351). Stroudsburg, PA, USA: Association for Computational Linguistics.

John Lyons (1977). *Semantics.* Cambridge University Press, Cambridge.

Manuel Martí Sánchez (2008). *Los Marcadores en Español L/E: Conectores Discursivos y Operadores Pragmáticos.* Arco Libros, Madrid.

María Antonia Martín Zorraquino and Estrella Montolío (1998). *Marcadores del Discurso.* Ariel, Barcelona.

Roser Martínez (1997). *Conectando Texto: Guía para el Uso Efectivo de Elementos Conectores en Castellano.* Octaedro, Barcelona.

Estrella Montolío (2001). *Conectores de la Lengua Escrita.* Ariel, Barcelona.

Henk Pander Maat and Ted Sanders (2000). Domains of use or subjectivity? The distribution of three Dutch causal connectives explained. In E. Couper-Kuhlen and B. Kortmann (eds.), *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives* (Topics in English Linguistics, 33): 57-82. Mouton de Gruyter, Berlin, Boston.

Henk Pander Maat and Ted Sanders (2001). Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics*, 12(3): 247–274.

Henk Pander Maat and Liesbeth Degand (2001). Scaling causal relations and connectives in terms of speaker involvement. *Cognitive Linguistics*, 12(3): 211–246.

Henk Pander Maat and Ted Sanders (2006). Connectives in text. In K. Brown, A.H. Anderson, L. Bauer, M. Berns, G. Hirst and J. Miller (eds.), *Encyclopedia of Language and Linguistics*: 33–41. Elsevier, Amsterdam.

Mirna Pit (2006). Determining subjectivity in text: The case of backward causal connectives in Dutch. *Discourse Processes*, 41(2): 151–174.

Mirna Pit (2007). Cross-linguistic analyses of backward causal connectives in Dutch, German and French. *Languages in Contrast*, 7(1): 53–82.

Mirna Pit, Jacqueline Hulst and Henk Pander Maat (1996). Subjectiviteit en de Spaanse connectieven porque, ya que en puesto que. *Gramma/TTT*, 5(3): 221–240.

Salvador Pons (1998). *Conexión y Conectores: Estudio de su Relación en el Registro Informal de la Lengua* (Cuadernos de filología, anejo no. 27). Universitat de València, Facultat de Filología, Departamento de Filología Española (Lengua Española), Valencia.

Salvador Pons (2000). Los conectores. In A. Briz (ed.), *¿Cómo se Comenta un Texto Coloquial?*: 193–220. Ariel, Barcelona.

José Portolés Lázaro (2001). *Marcadores del Discurso*. Ariel, Barcelona.

Florian Pusse, Asad Sayeed and Vera Demberg (2016). LingoTurk: managing crowdsourced tasks for psycholinguistics. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. http://aclweb.org/anthology/N/N16/N16-3012.pdf.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind K. Joshi and Bonnie L. Webber (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference of Language Resources and Evaluation (LREC).* https://www.seas.upenn.edu/~pdtb/papers/pdtb-lrec08.pdf

Ted Sanders (1997). Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes*, 24(1): 119–147.

Ted Sanders and Wilbert Spooren (2007). Discourse and text structure. In D. Geeraerts and H. Cuykens (eds.), *Handbook of Cognitive Linguistics:* 916–943. Oxford University Press, Oxford.

Ted Sanders and Wilbert Spooren (2009). Causal categories in discourse – Converging evidence from language in use. In T. Sanders and E. Sweetser (eds.), *Linguistic Categories of Causality in Discourse*: 205–246. Mouton de Gruyter, Berlin.

Ted Sanders and Wilbert Spooren (2015). Causality and subjectivity in discourse: The meaning and use of causal connectives in spontaneous conversation, chat interactions and written text. *Linguistics*, 53(1): 53–92.

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

Ted Sanders, Wilbert Spooren and Leo Noordman (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 4(2): 93–133.

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

Andrea Santana, Dorien Nieuwenhuijsen, Wilbert Spooren and Ted Sanders (2017). Causality and subjectivity in Spanish connectives: Exploring the use of automatic subjectivity analyses in various text types. *Discours. Revue de Linguistique, Psycholinguistique et Informatique*, 20.

Luis Santos Río (2003). *Diccionario de Partículas*. Luso-Española de Ediciones, Salamanca.

Marc Silver (2006). *Language Across Disciplines: Towards a Critical Reading of Contemporary Academic Discourse*. BrownWalker Press, Florida.

Merel Scholman and Vera Demberg (2017). Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task. In *Proceedings of the*

XI Linguistic Annotation Workshop (LAW) (pp. 24-34). Stroudsburg, PA, USA: Association for Computational Linguistics.

Wilbert Spooren (1997). The processing of underspecified coherence relations. *Discourse Processes*, 24(1):149-168.

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

Wilbert Spooren, Ted Sanders, Mike Huiskes and Liesbeth Degand (2010). Subjectivity and causality: A corpus study of spoken language. In J. Newman and S. Rice (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*: 241–255. University of Chicago Press, Chicago.

Ninke Stukker and Ted Sanders (2009). Another('s) perspective on subjectivity in causal connectives: A usage-based analysis of volitional causal relations. *Discours. Revue de Linguistique, Psycholinguistique et Informatique*, 4.

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

Ninke Stukker, Ted Sanders and Arie Verhagen (2008). Causality in verbs and in discourse connectives: Converging evidence of cross-level parallels in Dutch linguistic categorization. *Journal of Pragmatics*, 40(7): 1296–1322.

John Swales (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press, New York.

Eve Sweetser (1990). *From Etymology to Pragmatics: The Mind-body Metaphor in Semantic Structure and Semantic Change*. Cambridge University Press, Cambridge.

Maite Taboada (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics,* 38(4): 567–592.

Maite Taboada (2009). Implicit and Explicit Coherence Relations. In J. Renkema (Ed.), *Discourse, of Course*: 127–140. John Benjamins, Amsterdam.

Maite Taboada and Debopam Das (2013). Annotation upon annotation: Adding signalling information to a corpus of discourse relations. *Dialogue and Discourse*, 4(2): 249–281.

Teun van Dijk (1998). *News As Discourse*. Lawrance Erlbaum Associates, Inc., Publishers, New Jersey.

Elizabeth Traugott (1995). Subjectification in grammaticalisation. In S.Wright & D. Stein (Eds.), *Subjectivity and Subjectivisation: Linguistic perspectives*: 31–54. Cambridge University Press, Cambridge.

Nancy Vázquez Veiga (2002). Diccionario de colocaciones y marcadores del español: esbozo de una entrada de un marcador discursivo. *Communication in IV Congreso de Lingüística General* (2459–2472). Servicio de Publicaciones de la Universidad de Cádiz, Cádiz-Alcalá de Henares.

Agustín Vera (1984). En torno a la causalidad (aproximación a los fenómenos recursivos-causales a la luz de una teoría de base prototípica). *Anales de La Universidad de Murcia*, 42(1–2): 31–50.

Arie Verhagen (2005). *Constructions of Intersubjectivity: Discourse, Syntax, and Cognition*. Oxford University Press, Oxford.

Linda Waugh (1995). Reported speech in journalistic discourse: The relation of function and text. *Text - Interdisciplinary Journal for the Study of Discourse,* 15(1): 129–173.

Yipu Wei, Ted Sanders, Jacqueline Evers-Vermeul and Willem Mak (2018). *Causal connectives and perspective markers in Chinese: The encoding and processing of subjectivity in discourse*. LOT, Utrecht.

Heide Wegener (2000). Da, denn und weil – der Kampf der Konjunktionen. Zur Grammatikalisierung im kausalen Bereich. In R. Thieroff, M. Tamrat, N. Fuhrhop and O. Teuber (eds.), *Deutsche Grammatik in Theorie und Praxis:* 69–82. Mouton de Gruyter, Berlin, Boston.

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

Sandrine Zufferey, Willem Mak, Ted Sanders and Sara Verbrugge (2017). Usage and processing of the French causal connectives *car* and *parce que*. *Journal of French Language Studies*, 28(1): 85-112.