Andrea Santana*, Wilbert Spooren, Dorien Nieuwenhuijsen and Ted J.M. Sanders

Do Spanish causal connectives vary in subjectivity? What crowdsourcing data reveal about native speakers' preferences

Abstract: Language users have preferences for the connectives they choose to express causal relations. These choices may depend on the subjectivity involved in the relation. Dutch connectives illustrate this situation clearly: *want* ('since/for') is preferred typically for expressing subjective relations and *omdat* ('because') for objective ones. While various corpus-based studies have revealed a similar pattern in other languages, little attention has been paid to Spanish from this perspective. Recent corpus-based studies analyzed the connectives *porque* ('because'), *ya que* ('since') and *puesto que* ('given that') using two different methods of analysis. Surprisingly, the findings did not coincide with the previous literature on Spanish connectives, and the semantic profile of such connectives in terms of subjectivity remained unclear. The current study again aims to investigate whether these connectives show systematic variation in terms of subjectivity, using crowdsourcing experimentation. Results show that Spanish native speakers prefer *puesto que* over *porque* to express subjective relations. However, no statistically significant difference was observed between *porque* and *ya que*. This study offers a better understanding of Spanish connectives in terms of subjectivity. Furthermore, it contributes to the assessment of the use of crowdsourcing as a useful and reliable method to elucidate the meaning and use of connectives.

Keywords: coherence relations, causality, subjectivity, Spanish connectives, crowdsourcing

*Corresponding author: Andrea Santana, Instituto de Literatura y Ciencias del Lenguaje (ILCL), Pontificia Universidad Católica de Valparaíso (PUCV), Av. El Bosque 1290, 2530388, Viña del Mar, Chile, E-mail: andrea.santana@pucv.cl
Wilbert Spooren, Department of Modern Languages and Cultures, Radboud Universiteit, Postbus 9103, NL-6500HD, Nijmegen, The Netherlands, E-mail: w.spooren@let.ru.nl
Dorien Nieuwenhuijsen and Ted J.M. Sanders, Department of Languages, Literature and Communication, Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK, Utrecht, The Netherlands, E-mail: d.nieuwenhuijsen@uu.nl (D. Nieuwenhuijsen), T.J.M.Sanders@uu.nl (T.J.M. Sanders)

1 Introduction

Fundamental notions in human cognition can be expressed at the level of discourse. Human language provides its speakers with linguistic means to express these notions. Connectives such as *because*, *since* and *so* are good examples of these means; they express causality, specifically, they signal that a causal relation can be established between two discourse segments.

Another important notion expressed in discourse is subjectivity. This is a cognitive principle that allows us to distinguish, among other things, between objective and subjective causal relations, depending on the presence of a Subject of Consciousness (SoC). This SoC is responsible for the causal relation constructed between two or more discourse segments (Pander Maat and Sanders 2001). Consider these examples:

(1)   *Crops died because there is a water shortage.*
(2)   *People from cold countries must feel annoyed with these heatwaves. They are not used to these temperatures.*

(1) illustrates an objective causal relation since there is no SoC involved in the causal construction; the drought, which is a physical fact, affected the crops in such a way that the crops died. By contrast, (2) shows a subjective causal relation, because there is a SoC – a speaker who is concluding about people's feelings. Causal relations can be expressed with or without connectives: (1) is signaled by *because*. However, both (1) and (2) are clear cases of causal relations, which demonstrates that the presence of connectives is not a prerequisite for expressing causal relations.

This distinction in terms of subjectivity can also be extended to the use of causal connectives. In several languages, causal connectives specialize in expressing either subjective or objective relations. Indeed, various corpus-based studies provide evidence for this view. For instance, Dutch backward connectives (first consequent Q, then antecedent P) like *want* ('since/for') and *aangezien* ('since') are used to express subjective relations, whereas *omdat* and *doordat* ('because') are preferred to express objective relations (Sanders and Spooren 2015). In German, backward connectives *denn* and *da* ('because') tend to express subjective relations and *weil* ('because'), objective relations (Wegener 2000). In other typologically less related languages, a comparable distinction is detected. For example, in French the backward connectives *car* ('for') and *puisque* ('since') are preferred to express subjective relations, while *parce que* ('because') is used to express objective relations (Zufferey 2012). In Mandarin Chinese, the forward connective *kějiàn* ('so/therefore') is used to express subjective relations, whereas *yushi* ('so/therefore') and *yīn'er* ('as a result') are used in objective relations (Li et al. 2013).

These findings suggest that a distinction in terms of subjectivity might be systematic across different languages. However, in languages such as English, subjectivity is hardly systematically encoded in causal connectives. For example, *because* and *so* are general connectives that express both types of causal relations (Knott and Sanders 1998). This situation leads us to question whether there are other languages that resemble English in this respect. Spanish is an interesting candidate because it has been understudied from a cognitive perspective. Most of the existing studies in Spanish connectives are mainly descriptive and do not focus on the possible systematic differences in subjectivity.

Recent corpus-based studies that aim to identify systematic variation of causal connectives in terms of subjectivity provide evidence that differs from the ideas in earlier literature on Spanish causal connectives (Santana et al. 2017, 2018). In the current study, we add to these insights by using a different method to explore subjectivity in Spanish connective categorization – an experimental method that forces language users to make choices reflecting their preferences on Spanish causal connectives. Specifically, we use crowdsourcing to recruit native Spanish speakers and to expose them to experimental tasks. Two crowdsourcing experiments are implemented to answer the following research question: *Do Spanish connectives show a systematic variation in terms of subjectivity in an analysis of native speakers' preferences for causal constructions?* Based on the literature, our hypotheses are that *porque* is a general connective used to express both objective and subjective relations, whereas *ya que* and *puesto que* are specific connectives used mainly to express subjective relations. We aim to determine whether the lexicon of Spanish backward causal connectives varies systematically in terms of subjectivity, or whether it resembles the English lexicon and does not show a systematic variation.

This paper first gives an overview of the specific backward causal connectives considered in this study. Then it describes crowdsourcing and the method used in the implementation of two crowdsourcing experiments, and it presents the results and discussion. The paper ends with a general discussion and conclusions.

2 Spanish causal connectives

In Spanish, three common connectives express backward causality: *porque* ('because'), *ya que* ('since') and *puesto que* ('given that') (Santana et al. 2017). Previous studies focusing on the description of causal connectives (Blackwell 2016; Goethals 2010) suggest that the connective *porque* is a general connective used to express both subjective and objective relations. However, Santana et al. (2017), who implemented automatic analyses, found that *porque* tends to occur in subjective environments (i.e., in a context containing many subjective words). Santana et al. (2018), who used

manual text analyses, showed that *porque* was not associated with any subjectivity feature. Therefore, its profile in terms of subjectivity remains unclear.

Existing literature on *ya que* (Borzi and Detges 2011; Goethals 2002; Pit et al. 1996) indicates that this connective is used for specific purposes and might be used preferentially for expressing subjective relations. Nevertheless, this connective did not occur in subjective environments in Santana et al. (2017), nor was it associated with any subjectivity features in Santana et al. (2018), which raises the question whether it is a specific connective that expresses subjective relations.

Literature on *puesto que* (Pit et al. 1996; Santos Río 2003) also suggests this is a specific subjective connective. However, Santana et al. (2017) indicate that this connective might be associated both with subjective and objective relations since the segments preceding this connective contained relatively many subjective words, whereas the segments following this connective contained a relatively low percentage of subjective words. Furthermore, *puesto que* was associated with a low number of objective features in Santana et al. (2018), which leads us to infer that it is not used to express objective relations. Still, this tendency was observed in only one of the analyzed contexts (academic discourse) and for only two subjectivity features (the presence of SoC and the identity of SoC).[1] Therefore, further research is needed to provide evidence about the subjectivity profile of this connective.

In sum, the semantic-pragmatic profile of these Spanish causal connectives in terms of subjectivity remains unclear and the evidence provided by corpus-based studies differs from the ideas in earlier literature. Why do corpus results not coincide with the theoretical profiles of these connectives suggested in the literature? A possible explanation might be that Spanish connectives do not systematically encode subjectivity. Accordingly, the hypothesis that causal connectives show a systematic variation in terms of subjectivity across different languages must be reconsidered; in this respect, English would not be a mere exception since Spanish would not follow such a pattern of subjectivity either. Another feasible explanation could be that the differences of meaning and use among Spanish connectives are so subtle that they cannot be observed with automatic analyses as applied in Santana et al. (2017). A final likely reason could be that using manual analyses (Santana et al. 2018) provides only a limited understanding of Spanish connectives, since the results are based on the judgment of a small number of annotators, which might reflect the idiosyncrasies and/or the specialized knowledge of annotators instead of the meaning and use of connectives. These three explanations raise the question concerning the extent to which the methods used were appropriate and sufficiently effective to explore subjectivity in Spanish connective categorization. The use of other methods can lead

---

1 In Santana et al. (2018), 'context' refers to the source from which text types were extracted to analyze connectives and coherence relations. These were academic and journalistic contexts.

us to more precise findings. Crowdsourcing has proved to be a promising adequate method. The following section moves on to describe in greater detail this method.

3 Method

3.1 Crowdsourcing

A reasonable approach to determine whether the lexicon of Spanish backward causal connectives varies systematically in terms of subjectivity is to directly investigate Spanish native speakers' intuitions on the meaning and use of these connectives. Particularly, in experimental set-ups, speakers are asked to perform production tasks in which they are forced to choose between different options. Previous studies have provided crucial information on the meaning and use of coherence relations and connectives using such methods (see Sanders et al. 1992). Currently, an increasing number of efficient and sophisticated methods and tools are available to execute experimental tasks. *Crowdsourcing* is one such tool (Scholman and Demberg 2017a). It is suitable for investigating forced choices that reveal language users' preferences.

Crowdsourcing is defined as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" (Howe 2008: 99). It enables researchers to recruit participants through the web to perform tasks, which are rewarded with a low-value payment.

One advantage of crowdsourcing is that it allows researchers to conduct information retrieval experiments extremely fast and that it is an economical alternative (Alonso and Baeza-Yates 2011). However, among its drawbacks is the potential lack of commitment from participants, who might complete the tasks as quickly as possible to increase their remunerations (Sayeed et al. 2010). To avoid this situation, it is important to encourage participants to demonstrate a good performance by conditioning their participation in future experiments or by offering financial bonuses. Furthermore, crowdsourcing poses demands on the experimental design and setup, since participants are not experts in the tasks they need to perform. Consequently, tasks must be designed meticulously, and must provide clear instructions to ensure that participants understand them (Alonso and Mizzaro 2012).

Notwithstanding these restrictions, crowdsourcing has proved to be a promising method to elicit discourse interpretations and perform annotation tasks of discourse coherence relations and connectives (Scholman and Demberg 2017a, 2017b). There may be three reasons for this increasing success. Firstly, large amounts of data can be coded in a short time since it is easy to recruit subjects that only need a minimum

of instructions to complete a task. This is an important improvement because getting expert annotators via traditional annotation methods requires time, resources and effort, especially if additional and specific training is required.

Secondly, crowdsourcing can provide more insights into the meaning and use of coherence relations and connectives since subjects are forced to be precise and make decisions using their language knowledge. Researchers ensure that the collected data represent the average reader's interpretation, which is not biased by specialized knowledge of specific theoretical approaches. Consequently, the validity of the data increases since the obtained information corresponds to what really matters – the meaning and use of coherence relations and connectives. In the same vein, Krippendorff (2004: 428) states that the more coders participate in the process and the more common they are, the more likely the reliability of data can be ensured.

Lastly, crowdsourcing can be a more exhaustive method than traditional manual annotations. Collecting several responses for the same item reflects the meaning between two segments or the meaning encoded in a connective in a better way than a classification given by a single annotator. This is what Scholman and Demberg (2017a, 2017b) call a probability distribution over relation senses. Subjects choose the most fitting connective and the patterns in the choices might reveal semantic-pragmatic properties of the connectives involved.

These reasons lead us to consider crowdsourcing as an adequate method to identify the semantic-pragmatic profile of Spanish causal connectives in terms of subjectivity. What follows is a description of two experiments that aim to reach such a purpose.

3.2 Crowdsourcing experiments

Because similar methodologies were used in both experiments, we first present the information for both experiments and then specify details in which they differ.

Experiment 1 aims to distinguish the meaning and use of *porque* and *ya que*, and Experiment 2 focuses on the meaning and use of *porque* and *puesto que*. We used the connective insertion paradigm and the drag and drop interface of LingoTurk, which have been applied in similar experiments (Demberg et al. 2015; Scholman and Demberg 2017a).[2] LingoTurk is an open-source, crowdsourcing client/server system that aims at facilitating psycholinguistic experimentation. It enables user-friendly local hosting of experiments, and it is compatible with Prolific (P), which is a crowdsourcing platform focusing on non-US-centric clientele (Pusse et al. 2016).[3] LingoTurk enabled us to design, construct and apply

---

2 LingoTurk is available in https://github.com/FlorianPusse/Lingoturk.

3 Prolific is available in https://www.prolific.ac/.

experiments and P allowed us to administrate them and recruit people according to their backgrounds (age, language, education, etc.). This combination is beneficial for the experimental control. LingoTurk keeps track of experimental conditions and exclusions, whereas P keeps track of stimuli and responses.

3.2.1 Participants

Forty-three native Spanish speakers participated in Experiment 1, and 44 in Experiment 2. They were recruited via P and reimbursed for their participation (£3,00 per completed task). Their education level ranged from undergraduate degrees to doctoral degrees.

3.2.2 Materials and design

The experimental texts for each experiment consisted of 32 relations: 16 *porque* and 16 *ya que* relations for Experiment 1, and 16 *porque* and 16 *puesto que* relations for Experiment 2. The *porque* items were identical in both experiments. All relations corresponded to subjective and objective causal relations. Thus, a 2 × 2 experimental design with four different conditions was created for each experiment.

Sixteen fillers, which were identical for both experiments, were included. These consisted of clear cases of concessive and instantiation relations: eight items with *sin embargo* ('however') and eight items with *por ejemplo* ('for example'). These relations were chosen because they cannot be confused with causal relations. Thus, the total number of items for each experiment was 48.

A portion of the experimental items stemmed from the Journalistic Sub-Corpus of Spanish (Santana et al. 2018), and another part was extracted from CORPES XXI.[4] This corpus allowed us to search for cases considering text type as a parameter (news), and it enabled us to fulfill the established requirements of this study: 1) to select enough cases of subjective and objective relations since the Journalistic Sub-Corpus of Spanish contained mostly subjective relations; 2) to select unambiguous subjective and objective cases considering that some cases identified in the Journalistic Sub-Corpus of Spanish presented more than one indicator of subjectivity (i.e., relations containing subjective and objective indicators simultaneously); and 3) to obtain true examples that do not contain in their co-textual information the same connectives that were used in the filler condition or phrases with similar meanings (e.g., *pero* ('but'), *a modo de ejemplo* ('as an example')). Thus, we excluded elements that could influence the decision of the participants.

---

4 CORPES XXI is freely available in https://webfrl.rae.es/CORPES/view/inicioExterno.view.

The cases selected from the Journalistic Sub-Corpora of Spanish correspond to cases extracted from news texts, and the requirements mentioned above were applied as well. Then, all cases from both corpora consisted of subjective relations containing only subjective indicators (speech-act/epistemic in domain, speech-act/judgement in modality of Q-segment, implicit in presence of SoC and author/current speaker in identity of SoC) and objective relations containing only objective indicators (volitional/non-volitional in domain, physical/mental facts in modality, explicit/absent in presence of SoC and character/N/A in identity of SoC). Furthermore, the same analytical model used in Santana et al. (2018) was applied to select the experimental cases from CORPES XXI. The filler items were also extracted from this online corpus.

Text passages consisted of previous co-text, which was the clause preceding the causal relation; segment 1, which was the Q (consequence) clause in the causal relation; a green box, which was the place where participants had to drag the connective; segment 2, which was the P (cause) clause adjacent to segment 1 in the causal relation; and finally, posterior co-text, which was the clause following the causal relation established between segments 1 and 2. The two segments corresponding to the causal relation were displayed in black font, while the co-text sentences were in grey font. Moreover, to avoid participants being influenced by the original punctuation markers, slashes (//) were added before and after each connective and participants were told that these slashes could be interpreted as the presence or absence of punctuation markers (for a general perspective on experimental items, see the Appendix).

3.2.3 Procedure

Experiments were hosted on LingoTurk (Pusse et al. 2016). First, the participants read the instructions and accepted a consent form. Next, they were presented with the experimental interface, which contained a summary of the instructions, the predefined list of connectives, and the text passage. The participants were asked to choose the connective that best reflected the connection between the two segments (black sentences) but considering the co-textual information (grey sentences). The task was to insert a connective by dragging and dropping an option from the list of connectives to the green box. The participants could drag and drop connectives as many times they wanted, but once they had submitted their response they could not go back. Choosing a connective from the candidate list was mandatory to advance to the next item. In case the participants considered that no connective fitted the text, they had the option of clicking *ninguno de estos* ('none of these') and type a linking phrase more appropriate for the text. If they changed their mind, they could go back and select a connective from the candidate list of connectives. The inclusion of the choice *ninguno de estos* pursued to corroborate if the

Figure 1: The experimental interface.

participants preferred the predefined connectives; to obtain information about other connectives that may be considered as preferences and to avoid any possible blinds. Figure 1 illustrates the experimental interface for Experiment 1, which is identical to Experiment 2 except for the items presented.

The order of presentation of items was randomized to prevent order effects. The average completion time was 23:42 min in Experiment 1 and 29:29 min in Experiment 2.

3.2.4 Statistical procedure

The results of both experiments were modeled using a Linear Mixed-Effect Regression model (LMER; Baayen et al. 2008). We performed binomial mixed effects regression models using the *lme*4 package within the statistical software R (Bates and Sarkar 2007; R Development Core Team 2008). All models had a maximal random effect structure (Barr et al. 2013).

The factor for subjectivity (objective/subjective) was centered and the 2-level factor for connectives (*porque* and *ya que* for Experiment 1, and *porque* and *puesto que* for Experiment 2) was deviation coded. The significance of fixed effects was evaluated by performing likelihood ratio tests, in which the fit of a model containing the fixed effect for each condition is compared to another model without it but which is otherwise identical in random effects structure. The answers corresponding to non-causal connectives in the experimental items (*por ejemplo*, *sin embargo* and *other*) were not considered in the analyses.[5] They were chosen

---

5 '*Other*' corresponds to those connectives that were inserted manually by the participants.

relatively rarely (13.6% in Experiment 1 and 12.3% in Experiment 2) and no clear pattern was identified.[6]

4 Results

This section presents the information separately for each experiment.

4.1 Experiment 1

This experiment aimed to identify whether *porque* and *ya que* show a systematic variation in terms of subjectivity. Based on previous literature, our hypothesis was that *porque* is a general connective that expresses objective and subjective relations, whereas *ya que* is a specific connective used mainly to express subjective relations.

Prior to the analyses, data of three participants were removed because their completion times were short (<13 min for 48 passages of at least four sentences each) and showed incongruity on filler items with other participants. Similar studies adopted the same strategy (Scholman and Demberg 2017a, 2017b). The following analyses considered a total of 1920 observations, corresponding to the answers of 40 participants.

4.1.1 The validity of the method

By performing crowdsourcing experimentation, we are interested in identifying patterns in the choices of participants, which might reveal the meaning and use of the connectives. To evaluate whether the method was effective, we first focus on the overall results. Figure 2 shows the general distribution of inserted connectives per relation type. Data reveal a clear preference for inserting the connective *sin embargo* in concessive relations (85%) and *por ejemplo* in instantiation relations (75%).

---

6 To determine the effect of non-causal insertions, we ran a Multinomial Logistic Mixed Effects model using the IBM SPSS Statistics 24 software. As multinomial response variables we considered *porque*, *ya que* and *other* for Experiment 1, and *porque*, *puesto que* and *other* for Experiment 2). The results from the Multinomial analysis were similar to those of the binomial analysis reported here; all effects remained significant.
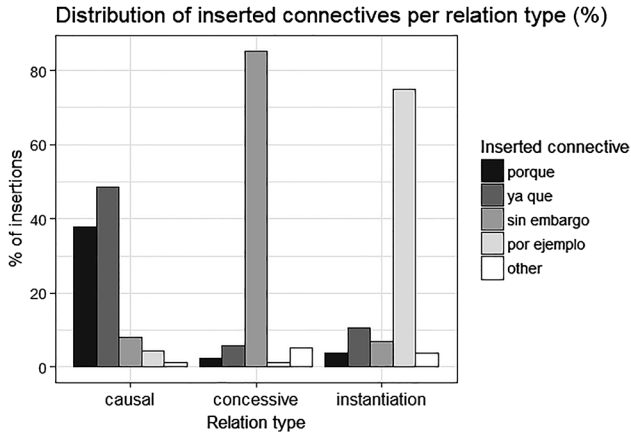
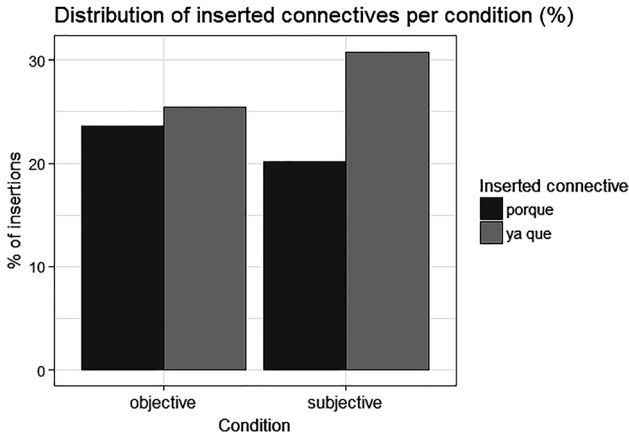**Figure 2:** General distribution of inserted connectives.

Causal connectives are also inserted in concessive relations, but their percentages were relatively low (2.5% for *porque*, and 5.9% for *ya que*). Low frequencies were also found for the instantiation connective *por ejemplo* (1.2%) and with *other* (5.3%). Within these manual answers, the participants used different connectives, like *pero* ('but'), the most frequent connective.

Regarding instantiation relations, Figure 2 shows that other connectives different from *por ejemplo* were inserted. *Ya que* was the second preference (10.6%) and *sin embargo*, the third preference (6.8%), whereas *porque* and *other* reached the same percentage, which was relatively low (3.7% each). Among the manual answers corresponding to *other*, the participants appear to use various connectives (*como* ('like'), *en el que* ('in which'), *por lo que* ('so'), *con lo cual* ('wherewith'), *inclusive* ('even'), *a propósito* ('by the way'), *con* ('with'), *para* ('for'), *de* ['of']).

These results allow us to conclude that the method was effective: the predominant connectives inserted in the filler conditions were those we expected: *sin embargo* for concessive relations and *por ejemplo* for instantiation relations. Therefore, we deduce that the participants executed the task seriously.

4.1.2 Distribution of connectives in causal relations

Having evaluated the method, let us now turn to causal relations. In Figure 2, the distribution of the inserted connectives *porque* and *ya que* was quite similar (37.8 and 48.5%, respectively). The participants also used other connectives instead of the causal ones: *sin embargo*, which was the third most frequently used connective

Figure 3: Distribution of inserted connectives in causal relations.

(8%), *por ejemplo* and *other,* which reached relatively low percentages (1.2 and 5.3%, respectively). Regarding the manual answers corresponding to *other*, no clear pattern was identified, the participants used various connectives, even commas and blank spaces.

Regarding the distribution of *porque* and *ya que* per condition (objective and subjective relations), of a total of 1920, 1105 insertions in the causal condition involved either *porque* or *ya que*. Of these 1105 insertions, 563 involved insertions in the subjective condition and 542 in the objective condition. Of the 563 subjective insertions, 223 were choices for *porque* (20.1%) and 340 for *ya que* (30.7%). Of the 542 objective insertions, 261 were choices for *porque* (23.6%) and 281 for *ya que* (25.4%). Figure 3 shows the distribution of *porque* and *ya que* per condition. It can be observed that both connectives are used in both conditions; in the objective condition the distribution is similar between both connectives, whereas in the subjective condition, there is a preference for *ya que*.

4.1.3 Statistical analyses

The answers of the 32 experimental items were modeled using a binomial linear mixed-effect regression model (see Section 3.2.4). The model included intercepts for each participant and item and random slopes for each participant. The final model is shown in Table 1. The condition of subjectivity did not have a significant effect on the use of the connectives *porque* and *ya que* ($\beta = 0.355$, SE $= 0.267$, $z = 1.333$, $p = 0.189$). This indicates that the participants did not choose *ya que* more often for subjective items than for objective items.

Table 1: Linear mixed-effect regression model fit by maximum likelihood.

Fixed effects	Estimates	SE	z	p
Intercept	0.326	0.152	2.148	
Subjectivity	0.355	0.267	1.333	0.189

Random effects	Variance	Standard deviation		
Participant	0.200	0.448		
Item	0.385	0.621		

4.1.4 Discussion

The results show that the distribution of the inserted connectives *porque* and *ya que* for causal relations are comparable, which indicates that the two connectives are used to express both subjective and objective causality. *Ya que* showed a slight tendency of being more preferred than *porque* to express subjective relations. However, no statistically significant difference was identified.

According to these results, it can be inferred that *porque* and *ya que* do not encode whether the relation is subjective or objective; they fit perfectly well in both types of relations. This situation alludes to the phenomenon of underspecification: the semantics of the connective that is used to indicate the link does not fully match the semantics of the relation that is intended by the speaker/writer (Spooren 1997). Underspecification occurs when a connective does not have a specific meaning profile; this makes it very suitable to various types of relations. In this case, *porque* and *ya que* seem to be underspecified for subjectivity: both mark backward causal relations, but neither of them specifies whether they mark a subjective or objective relation.

These results were the reason to revise segments 1 and 2 of the experimental items with *porque* and *ya que*, in order to identify other elements that may convey subjective meanings. This analysis showed that 86% of the subjective elements occur in subjective relations (36 occurrences) and only 14% in objective relations (6 occurrences), which indeed suggests a predominance of subjective elements in subjective relations. These elements consisted of evaluative adjectives that imply the author's evaluation of reality (e.g., *fácil* ['easy']) (21 occurrences in subjective items and 5 in objective items); adverbs of manner or quantity that emphasize the meaning of evaluative adjectives (e.g., *francamente ridículo* ['frankly ridiculous']) (5 occurrences in subjective items and 1 in objective items); and categorical statements and denials in present tense that may reflect the author's opinion (e.g., *El resultado es alentador* ['the result is encouraging']) (10 occurrences; all of them in subjective items).

Elements expressing a low degree of subjectivity were also identified: 76% (43 occurrences) was associated with objective relations and 22% (12 occurrences) with subjective relations, which corroborates the previous findings. The elements were sentences in past tense, which referred to events occurring in the real world (e.g., *se disculpó con las autoridades* ['he apologized to the authorities']) (30 occurrences in objective items and 12 in subjective items) and the presence of a third person (singular or plural) who carries out an intentional action causing another event in the real world (e.g., *Uno de sus dueños, Luis Hernández, negó las acusaciones* ['One of its owners, Luis Hernández, denied the accusations']), (13 occurrences; all of them in objective items). Although these findings are the result of a quite general, somewhat superficial analysis, they illustrate how subjectivity may be encoded in other linguistic elements and not necessarily in causal connectives.

Another finding is that the participants used *sin embargo*, which commonly marks a concessive relation, as the third preference in causal relations. This suggests that causal and concessive relations tend to be confused because both relations involve an implicational meaning (P→Q) and only differ in their polarity (König and Siemund 2000). This finding is consistent with previous results (Scholman and Demberg 2017a)

A final methodological conclusion is that crowdsourcing was effective. The connectives inserted in the filler conditions were those expected, which allows us to deduce that the participants performed the task seriously.
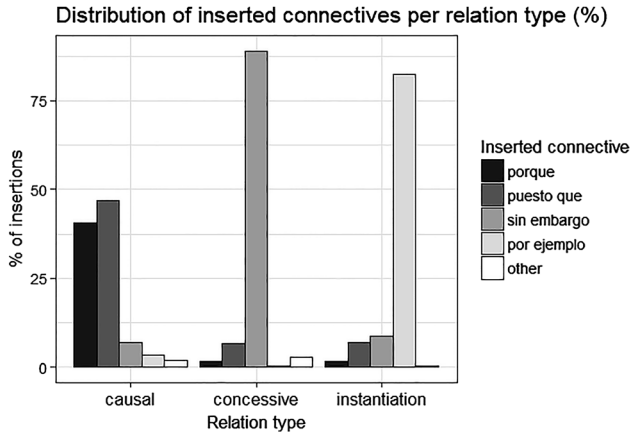
4.2 Experiment 2

This experiment aims to find out whether *porque* and *puesto que* show a systematic variation in terms of subjectivity. Following previous literature, we hypothesized that *porque* is a general connective that expresses objective and subjective relations, whereas *puesto que* is a specific connective used mainly to express subjective relations.

Similarly to Experiment 1, the data of four participants were removed because of short completion times (<5 min for 48 passages of at least four sentences each) and incongruity on filler items with other participants. The following analyses considered a total of 1920 observations, corresponding to the answers of 40 participants.

4.2.1 Method validity

As in Experiment 1, we first evaluated the effectiveness of the method (Figure 4 provides the general distribution of inserted connectives per relation type).

Figure 4: General distribution of inserted connectives.

Regarding concession and instantiation relations, this experiment replicates Experiment 1 in that *sin embargo* and *por ejemplo* are clearly predominant in each relation (88.7 and 82.5%, respectively).

Figure 4 shows that in the concessive relations, causal connectives were also inserted, but their percentages were relatively low (1.5% for *porque*, and 6.5% for *puesto que*). The same is observed with the instantiation connective *por ejemplo* (0.3%) and with *other* (2.8%). With regard to the choice of *other*, there was no clear pattern; the participants used different connectives, commas and blank spaces.

With respect to instantiation relations, Figure 4 shows that other connectives different from *por ejemplo* were also used. *Sin embargo* and *puesto que* were the other preferences (8.7 and 6.8%, respectively), whereas *porque* and *other* reached the lowest percentages (1.5 and 0.3%, respectively). The latter percentage corresponds to only one manual answer (a full stop).

These results, once again, allow us to infer that the method was effective. The predominant connectives inserted in the filler conditions were those that we expected, which lead us to conclude that the participants executed the task seriously.

4.2.2 Distribution of connectives in causal relations

Figure 4 reveals that the distribution of the inserted connectives *porque* and *puesto que* reached similar percentages (40.7 and 46.8%, respectively). Among the other connectives used in causal relations, we observed the same pattern identified in Experiment 1. The concessive connective *sin embargo* was the third most frequently used connective (6.9%), so once again, this experiment replicates

Distribution of inserted connectives per condition (%)
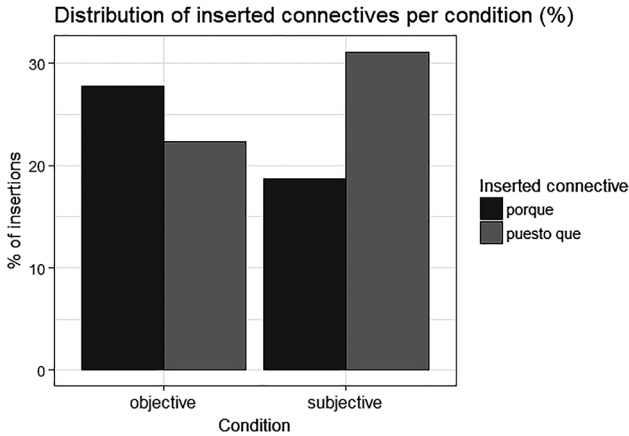
Figure 5: Distribution of inserted connectives in causal relations.

previous results (Scholman and Demberg 2017a). Regarding *por ejemplo* and *other,* the percentages were also relatively low (3.5 and 1.8%, respectively). Regarding this latter option, the most frequently used manual answers were *ya que* and *debido a que* ('due to'), which are also causal connectives.

Concerning the distribution of *porque* and *puesto que* per condition (objective and subjective relations), out of a total of 1920, 1122 insertions in the causal condition involved either *porque* or *puesto que*. Of these 1122 insertions, 559 involved insertions in the subjective condition and 563 in the objective condition. Of the 559 subjective insertions, 210 were choices for *porque* (18.7%) and 349 for *puesto que* (31.1%). Of the 563 objective insertions, 312 were choices for *porque* (27.8%) and 251 for *puesto que* (22.3%). Figure 5 shows the distribution of *porque* and *puesto que* per condition. Both connectives were also used in both conditions (cf. Figure 3); in the objective condition the distribution is similar between both connectives, whereas in the subjective condition there is a preference for *puesto que*.

4.2.3 Statistical analyses

The answers of the 32 experimental items were modeled using binomial linear mixed-effect regression models (see Section 3.2.4). The model included intercepts for each participant and each item and random slopes for each participant. Table 2 shows the final model. Unlike in Experiment 1, the condition of subjectivity did have a significant effect on connective choice (β = 0.783, SE = 0.276, z = 2.840, $p < 0.01$). The odds ratio indicates that if the relation is subjective, then the choice for *puesto que* is 2.065 times more likely than the choice for *porque*.

Table 2: Linear mixed-effect regression model fit by maximum likelihood.

Fixed effects	Estimates	SE	z	p
Intercept	0.185	0.156	1.187	
Subjectivity	0.783	0.276	2.840	<0.01

Random effects	Variance	Standard deviation		
Participant	0.068	0.261		
Item	0.454	0.674		

4.2.4 Discussion

The results show that the distribution of the inserted connectives *porque* and *puesto que* is comparable in objective relations. However, in subjective relations, *puesto que* tends to be used more than *porque*, and this difference is statistically significant.

These results allow us to conclude that *puesto que* is a specific subjective connective. The participants showed a (small) preference for this connective over *porque* to express subjective relations, which shows that *puesto que* fits better than *porque* to express this kind of relation. However, it is important to bear in mind that both connectives were used in both subjective and objective relations, and in objective relations, the distribution was similar.

These results can be interpreted considering the prototype structure theory (Stukker and Sanders 2012), which proposes two hypotheses. First, it is expected that causal categories are prototypes and have a prototype structure. This prototype structure implies that causal connectives specializing in expressing subjective causal relations are used in objective causal relations less frequently than in subjective relations. Similarly, connectives specializing in expressing objective causal relations are used in subjective causal relations less frequently than in objective relations. In other words, the prototype structure theory states that causal connectives that are specialized in expressing a type of causal relation are prototypically used to express that specific type of relation more than any other type. The obtained results seem to support this hypothesis to a certain degree in the sense that the participants demonstrated a preference for using *puesto que* over *porque* to express subjective causality. However, the results seem to clearly support the corollary of the prototype structure theory: non-prototypical instantiations of causal categories may be rare but they do occur. That is, it should be expected that, in certain contexts, cases show up in which presumably 'objective' causal connectives are used to express subjective relations, and cases in which presumably 'subjective' causal connectives are used to express objective relations. This is exactly what was

found in Experiment 2: *puesto que* is a subjective connective that was also used in objective relations. These occurrences, which are rarer than those in subjective relations, correspond to the non-prototypical meaning and use of *puesto que*. This demonstrates, as would be expected on the basis of a prototype analysis, that the distinction between subjective and objective connectives is not always a clear-cut issue (see Sanders and Spooren 2013).

These results also coincide with the pattern observed in other languages. The connective that revealed a preference was specific for subjective causality, which is the category that has manifested greater consistency in other languages. Specific subjective connectives like *want* in Dutch, *denn* in German and *car* in French have shown a clearer pragmatic-semantic profile than their objective counterpart *omdat*, *da* and *parce que*. In this respect, future research might explore whether *puesto que* demonstrates the same profile across different genres.

For *porque*, this experiment replicates the results obtained in Experiment 1. This connective is used to express both subjective and objective relations. Therefore, *porque* can be considered an underspecified connective in terms of subjectivity and has a semantic profile similar to the English connective *because*.

Finally, the method used in this experiment demonstrated to be effective.

5 General discussion and conclusions

Two crowdsourcing experiments were implemented to answer the following research question: *Do Spanish connectives show a systematic variation in terms of subjectivity in an analysis of native speakers' preferences for causal constructions?* The results showed that the distribution of *porque* and *ya que* was similar between objective and subjective relations, and no statistically significant difference was identified. Regarding *porque* and *puesto que,* the participants preferred *puesto que* to express subjective relations, and this difference was statistically significant.

The hypothesis that *puesto que* is a specific subjective connective is borne out. In this respect, our results coincide with previous studies (Pit et al. 1996; Santos Río 2003) suggesting that *puesto que* has a more specific profile. Furthermore, our results are congruent with results obtained in Santana et al. (2018), in which manual analyses revealed that *puesto que* was associated with a low number of objective features.

The hypothesis that *porque* is a general connective used to express both objective and subjective relations is also borne out. Both experiments revealed the underspecified nature of this connective in terms of subjectivity. These results are in line with previous literature (Blackwell 2016; Goethals 2010) and are also congruent with the results of Santana et al. (2018), in which *porque* demonstrated not to be associated with any subjectivity feature.

The hypothesis that *ya que* is a specific subjective connective is rejected. No significant differences were found in the participants' preferences. This result corroborates the results obtained in Santana et al. (2018), in which *ya que* was not associated with any subjectivity feature. It also accords with the results of Santana et Al. (2017), in which *ya que* did not co-occur with high percentages of subjective words in any of the text types analyzed. Nonetheless, this result does not coincide with previous literature (Borzi and Detges 2011; Goethals 2002), which suggests that this connective is used for specific purposes.

A possible explanation for this discrepancy might be that the research methods affected the results. Where claims in the existing literature were mainly based on qualitative analyses by linguists, our inferences are based on the statistical evaluation of responses produced by naïve informants. It may be possible, therefore, that qualitative analyses overinterpreted the differences that in our quantitative study did not prove to be large enough to be statistically significant.

Additionally, in comparison to all existing studies, the data in the current study were obtained very differently. Our data are the answers of participants who were forced to make decisions about connectives within experimental settings, whereas previous studies collected data from annotations of specific corpora (Borzi and Detges 2011; Goethals 2002). Consequently, the participants' answers reflect lay people's 'theory-less' intuitions about the meaning and use of Spanish connectives. By contrast, the corpus-based annotations are the result of the specialized knowledge of linguistically trained annotators. We can only speculate about how this differs from non-trained persons, but there is a possibility that such a training influences the categorizations, depending on the theoretical framework adopted in the analyses.

Taken together, these findings allow us to conclude that the Spanish language has a repertoire of causal connectives with different subjectivity profiles. On the one hand, general connectives such as *porque* and *ya que* express both objective and subjective relations. On the other hand, Spanish also has a specific subjective connective: *puesto que*. Furthermore, we might conclude that there are more similarities between *porque* and *ya que* than between *porque* and *puesto que*. Indeed, their frequencies would be evidence of this similarity. Santana et al. (2017) revealed that *porque* and *ya que* were the most frequent connectives in news texts and in academic text types, whereas *puesto que* was less frequent than these connectives in the same contexts.

In conclusion, the answer to our research question must reflect the twofold nature described above: the Spanish causality lexicon includes general and specific connectives in terms of subjectivity. Moreover, the systematic categorization of causal connectives in terms of subjectivity in Spanish is less strong than in several other languages. Only one connective shows specificity in terms of

subjectivity, contrary to other languages, which have at least two specific connectives. In this respect, the lexicon of Spanish backward causal connectives resembles the English lexicon; both languages do not show an evident systematic variation in terms of subjectivity.

What are the theoretical implications of these results? Do Spanish and English not express subjectivity at the discourse level? Clearly, they do. Corpus studies show that subjective coherence relations occur in both languages (Cunha et al. 2011; Prasad et al. 2004). In Spanish, there is a body of research concerning subjectivity (Aliaga and Bustos 2002; Cornillie 2007, among others). Furthermore, in Santana et Al. (2018) the occurrence of subjective relations constituted more than 75% of the analyzed relations. Therefore, subjectivity cannot be ignored. At the same time, the current findings raise the necessity to reconsider the hypothesis that causal connectives of all languages show a systematic variation in terms of subjectivity (Stukker and Sanders 2012). The results confirm that subjectivity is encoded differently across languages, and that it is not necessarily encoded at the discourse level. This suggests a relationship between the subjectivity expressed in connectives and the subjectivity expressed elsewhere in the discourse. There might even be a trade-off between subjectivity indicators in the discourse segments and that expressed in connectives. For instance, when speakers use modal expressions, it may no longer be necessary to use a subjective causal connective. Conversely, when discourse segments do not contain elements expressing subjectivity, speakers are likely to use a specific subjective connective to express the causal relation in order to make themselves clear. For this reason, further research on other linguistic elements in the context of Spanish connectives is strongly recommended. Recent studies in other languages have assumed such an approach and have offered new insights into the use of linguistic elements expressing subjectivity associated with causal connectives (Levshina and Degand 2017; Wei et al. 2020).

Against this backdrop, we conclude that the systematization of causal connectives in terms of subjectivity is a matter of degrees rather than a dichotomy. In one extreme, some languages show a clear distinction between specific connectives that mark subjective and objective causality prototypically. This would be the case with the Dutch language, which is a clear exponent of such systematization. Mandarin Chinese is another language that might be placed at this extreme end since it also has a wide repertoire of specific connectives in terms of subjectivity. In the middle of this continuum, German and French also reveal such systematization, but to a lesser degree than Dutch and Mandarin Chinese do. In German, there is one connective that has revealed some inconsistency across text types and studies (*weil* ['because']). In French, some connectives are subject to a grammaticalization process: objective connectives are 'becoming' subjective, specifically

in spoken discourse (*parce que*, ['because']) (Zufferey et al. 2018). Despite this, both languages display connectives that demonstrate a prototypical pattern of subjectivity. Finally, at the other end of this continuum, there are languages like Spanish and English in which the systematic categorization of the lexicon of causal connectives is much less evident than in other languages.

Another contribution of the present study is to provide empirical evidence that confirms the effectiveness of crowdsourcing. It proved to be a useful, effective and reliable method to elucidate the meaning and use of Spanish connectives in terms of subjectivity. By asking for many responses from native speakers for the same item, we obtained more exhaustive information about the meaning and use of the analyzed connectives than the information obtained in previous corpus-based studies (Santana et al. 2017, 2018). This approach allowed us to reach a more precise categorization of Spanish causal connectives, since the data reflect the interpretation of native speakers' preferences.

We stress the importance of using various methods to investigate discourse coherence phenomena, especially when these have been understudied in a specific language, which was the case of the present study. Corpus-based studies carried out previously gave us a global perspective of subjectivity in Spanish coherence relations and connectives (Santana et al. 2017, 2018), but by adopting a more specific method such as crowdsourcing experimentation, it was possible to obtain a precise characterization of the phenomenon. In sum, this combined implementation of traditional and innovative methods provides substantial benefits: while corpus-based manual text analyses facilitated the identification of causal connectives, causal coherence relations and details of subjectivity in Spanish (Santana et al. 2017, 2018), crowdsourcing made it possible to collect more data than would have been possible using traditional methods.

An obvious route for further research is to analyze discourse functions of the connectives and their association with other linguistic elements that express degrees of subjectivity. Another promising study would concern the processing of causal connectives in Spanish. We trust that the current study serves as a preliminary step for such further investigations. Finally, we emphasize the necessity of exploring causal coherence relations and connectives in various contexts and languages as well as investigating other types of relations, such as those that are not expressed by connectives or cue phrases. Since these reflect the majority of causal coherence relations, such studies would provide important additional information about the role of causality and subjectivity in discourse coherence.

Appendix

This section presents one experimental item per each condition of both experiments.[7]

A.1: Experimental items corresponding to ya que in Experiment 1

YA QUE
OBJECTIVE

Dos jóvenes mueren en aparatoso accidente
*El accidente produjo algunas horas de congestionamiento vial, **ya que** los curiosos no paraban de observar la escena del percance.*
La imprudencia temeraria nuevamente cobra la vida de dos personas.
'Two young people died in a dramatic accident
*The accident produced some hours of traffic congestion **since** curious people did not stop watching (kept on watching) the incident scene.*
The temerarious imprudence took the life of two people again

YA QUE
SUBJECTIVE

También fue incautado un viejo taxi iraquí, pintado con los colores tradicionales para este tipo de vehículos en este país, blanco y amarillo.
*Esto resulta interesante, **ya que** parece confirmar los rumores que circulaban desde hace tiempo en Bagdad: Saddam se ha dejado crecer la barba y viaja de incógnito en un viejo taxi.*
'Also, an old Iraqi taxi, painted with the traditional colors for this type of vehicle in this country, white and yellow, was confiscated.
*This turns out interesting **since** it seems to confirm the rumors that were circulating around for a long time in Bagdad: Saddam has let grow his beard and travels incognito in an old taxi.'*

---

7 Filler items were not included in this section since objective and subjective conditions only apply for causal experimental items.

A.2: Experimental items corresponding to puesto que in Experiment 2

PUESTO QUE
OBJECTIVE

A las siete de la mañana del 1 de agosto de 1981 el 90% de los controladores aéreos que estaban llamados a trabajar se ausentó de su puesto en demanda de un aumento salarial y de reducción de jornada. El entonces presidente de EE UU, Ronald Reagan, les dio un ultimátum de 48 horas para volver a su trabajo, **puesto que** una ley de 1956 declaraba ilegal toda huelga realizada por funcionarios.
Los 11.345 controladores -el 65% del total- que no respondieron a la llamada para que regresaran a sus puestos fueron fulminantemente despedidos.
'At seven in the morning on August 1st of 1981, 90% of the air traffic controllers that were called to work were absent from their positions demanding a salary increase and a working hours reduction. The then president of the USA, Ronald Reagan, gave them an ultimatum of 48 h to go back to work **given that** a law of 1956 declared as illegal every strike carried out by employees.
The 11.345 controllers -65% of the total- that did not answer the call to go back to work were immediately dismissed.'

PUESTO QUE
SUBJECTIVE

En el 2018 las bolsas de plástico de un solo uso tienen que haber desaparecido. A la vista de los resultados, esta campaña que se puso en marcha en abril de 2011, va por muy buen camino **puesto que** se han superado los objetivos marcados para este año pasado que era reducir el consumo de bolsas de plástico un 20%.
En seis meses de campaña, se ha logrado suprimir la emisión de 138 millones de bolsas anuales en Castilla y León.
'In 2018 single-use plastic bags must have disappeared. In view of the results, this campaign that started in April 2011 is on the very right track **given that** the established objectives for this last year which was reducing the use of plastic bags in a 20% have been surpassed.
In six months of the campaign, it has been managed to suppress the emission of 138 million annual bags in Castilla and León.'

A.3: Experimental items corresponding to porque in Experiment 1 and 2

PORQUE
OBJECTIVE

Kammerichs se arrepintió
*El basquetbolista Federico Kammerichs, de Pamesa Valencia, se disculpó con las autoridades del club español **porque** el viernes último regresó a la Argentina sin permiso. "Estoy arrepentido. Demostré falta de madurez y de experiencia", reconoció el alero.*
'Kammerichs repented
*The basketball player Federico Kammerichs of Pamesa Valencia apologized to the authorities of the Spanish club **because** last Friday he went back to Argentina without permission. "I am regretful. I demonstrated a lack of maturity and experience", the forward recognized.'*

PORQUE
SUBJECTIVE

Fútbol – historia: Francisco Rivera, talento y clase con la "10"
*No todo es fácil en la vida de un futbolista, **porque** hay que tener ánimo y una gran fortaleza moral y física para entrenar. Requiere mucho sacrificio.*
'Football – history: Francisco Rivera, talent and quality with the "10"
*Not everything is easy in a football player's life **because** you must have energy and an enormous moral and physical strength to train. It requires much sacrifice.'*

References

Aliaga, Francisco & Eduardo de Bustos. 2002. Mental spaces and epistemic attitudes: On the Spanish subjunctive/indicative alternation. In Javier Gutiérrez-Rexach (ed.), *From words to discourse: Trends in Spanish semantics and pragmatics*, 135–144. Amsterdam: Elsevier.

Alonso, Omar & Ricardo Baeza-Yates. 2011. Design and implementation of relevance assessments using hcrowdsourcing. In Paul Clough, Colum Foley, Cathal Gurrin, Gareth Jones, Wessel Kraaij, Hyowon Lee & Vanessa Mudoch (eds.), *European Conference on Information Retrieval*, 153–164. Berlin: Springer.

Alonso, Omar & Stefano Mizarro. 2012. Using crowdsourcing for TREC relevance assessment. *Information Processing & Management* 48(6). 1053–1066.

Baayen, Rolf, Douglas Davidson & Douglas Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. Elsevier 59(4). 390–412.

Barr, Dale, Roger Levy, Christoph Scheepers & Harry Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.

Bates, Douglas & Deepayan Sarkar. 2007. *The lme4 package*. R package version, 2.

Blackwell, Sara. 2016. Porque in Spanish oral narratives: Semantic porque, (meta)pragmatic porque or both?. In Alessandro Capone & Jacob Mey (eds.), *Interdisciplinary studies in pragmatics, culture and society*, 615–651. Heidelberg: Springer.

Borzi, Claudia & Ulrich Detges. 2011. Ya que, un marcador polifónico [ya que, a polyphonic marker]. In Heidi Aschenberg & Óscar Loureda Lamas (eds.), *Marcadores del discurso: De la descripción a la definición [discourse markers: from description to definition]*, 263–281. Madrid: Iberoamericana Vervuert.

Cornillie, Bert. 2007. *Evidentiality and epistemic modality in Spanish (semi-)auxiliaries: A cognitive-functional approach*. Berlin: Mouton de Gruyter.

Cunha, Iria da, Juan-Manuel Torres-Moreno & Gerardo Sierra. 2011. On the development of the RST Spanish Treebank. *Paper presented at the 5th law workshop (LAW), Portland, Oregon, USA*, 23–24 June.

Demberg, Vera, Hannah Rohde, Merel Scholman, Chris Cummins & Emily Nicolet. 2015. "On the one hand" as a cue to in the comprehension of discourse structure. *Paper presented at the first action conference of textlink structure discourse in multilingual Europe*. Belgium: Université Catholique de Louvain, 25–28 January.

Goethals, Patrick. 2002. *Las conjunciones causales explicativas españolas como, ya que, pues y porque: Un estudio semiótico-lingüístico [the Spanish explicative causal conjunctions como, ya que, pues and porque: A semiotic-linguistic study]*. Leuven: Peeters Publishers.

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

Howe, Jeff. 2008. *Crowdsourcing: Why the power of the crowd is driving the future of business*. New York: Crown Publishing Group.

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

König, Ekkehard & Peter Siemund. 2000. Causal and concessive clauses: Formal and semantic relations. In Elizabeth Couper-Kuhlen & Bernd Kortmann (eds.), *Cause – condition – concession – contrast*, 341–360. Berlin: Mouton de Gruyter.

Krippendorff, Klaus. 2004. Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research* 30(3). 411–433.

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

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

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

Pit, Mirna, Jacqueline Hulst & Henk Pander Maat. 1996. Subjectiviteit en de Spaanse connectieven porque, ya que en puesto que [subjectivity and the Spanish connectives porque, ya que and puesto que]. *Gramma/TTT* 5(3). 221–240.

Prasad, Rashmi, Eleni Miltsakaki, Aravind Joshi & Bonnie Webber. 2004. Annotation and data mining of the Penn discourse TreeBank. *Paper presented at the 2004 ACL workshop on discourse annotation, Barcelona, Spain*, 25–26 July.

Pusse, Florian, Asad Sayeed & Vera Demberg. 2016. LingoTurk: Managing crowdsourced tasks for psycholinguistics. *Paper presented at the NAACL HLT 2016, the 2016 conference of the North American chapter of the association for computational linguistics*. San Diego, California, USA: Human Language Technologies, 12–17 June.

R Development Core Team. 2008. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.

Sanders, Ted & Wilbert Spooren. 2013. Exceptions to the rules: A qualitative analysis of backward causal connectives in Dutch naturalistic discourse. *Text & Talk* 33(3). 377–398.

Sanders, Ted & 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.

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

Santana, Andrea, Dorien Nieuwenhuijsen, Wilbert Spooren & 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. https://doi.org/10.4000/discours.9307.

Santana, Andrea, Wilbert Spooren, Dorien Nieuwenhuijsen & Ted Sanders. 2018. Subjectivity in Spanish discourse: Explicit and implicit causal relations in different text types. *Dialogue & Discourse* 9(1). 163—191.

Santos Río, Luis. 2003. *Diccionario de Partículas [dictionary of particles]*. Salamanca: Luso-Española de Ediciones.

Sayeed, Asad, Timothy Meyer, Hieu Nguyen, Olivia Buzek & Amy Weinberg. 2010. Crowdsourcing the evaluation of a domain-adapted named entity recognition system. *Paper presented at the human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, Los Angeles, California, USA, 2–4 June.

Scholman, Merel & Vera Demberg. 2017a. Examples and specifications that prove a point: Identifying elaborative and argumentative discourse relations. *Dialogue & Discourse* 8(2). 56–83.

Scholman, Merel & Vera Demberg. 2017b. Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task. *Paper presented at the 11th linguistic annotation workshop (LAW)*, Valencia, Spain, 3 April.

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

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

Wegener, Heide. 2000. Da, denn und weil – der Kampf der Konjunktionen. Zur grammatikalisierung im kausalen Bereich [da, denn and weil – the battle of conjunctions. For a grammaticalization in the area of causality]. In Rolf Thieroff, Matthias Tamrat, Nanna Fuhrhop & Oliver Teuber (eds.), *Deutsche grammatik in theorie und praxis.[German grammar in theory and praxis.]*, 69–82. Berlin: Mouton de Gruyter.

Wei, Yipu, Jacqueline Evers-Vermeul & Ted Sanders. 2020. The use of perspective markers and connectives in expressing subjectivity: Evidence from collocational analyses. *Discourse & Dialogue* 11(1). 62–88.

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

Zufferey, Sandrine, Pim Mak, Sara Verbrugge & Ted Sanders. 2018. Usage and processing of the French causal connectives 'car' and 'parce que'. *Journal of French Language Studies* 28(1). 85–112.

Bionotes

Andrea Santana
Instituto de Literatura y Ciencias del Lenguaje (ILCL), Pontificia Universidad Católica de Valparaíso (PUCV), Av. El Bosque 1290, 2530388, Viña del Mar, Chile
andrea.santana@pucv.cl
Andrea Santana received her PhD in Linguistics at the Utrecht Institute of Linguistics-OTS (The Netherlands) and is currently assistant professor at the Pontificia Universidad Católica de Valparaíso (Chile). Her research interests include discourse coherence, discourse comprehension, language processing, and specialized discourse. Her most recent book-length publication is: *Is porque more like because or like omdat?* (2019, LOT).

Wilbert Spooren
Department of Modern Languages and Cultures, Radboud Universiteit, Postbus 9103, NL-6500HD, Nijmegen, The Netherlands
w.spooren@let.ru.nl
Wilbert Spooren is professor in Discourse Studies of Dutch at the Faculty of Arts, Radboud University Nijmegen. As a member of the Center for Language Studies, he is interested in the role of language in optimizing communication in different genres. More specifically he investigates how people rely on structural signals in planning or interpreting a discourse, and how this depends on the genre of the discourse.

Dorien Nieuwenhuijsen
Department of Languages, Literature and Communication, Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK, Utrecht, The Netherlands
d.nieuwenhuijsen@uu.nl
Dorien Nieuwenhuijsen received her PhD from the University of Groningen (The Netherlands), and is currently associate professor at Utrecht University (The Netherlands). Her research interests include the historical development of Spanish syntax, and variation and change in Peninsular and Latin American Spanish. She is particularly interested in the opportunities digital corpora offer for the study of historical linguistics.

Ted J.M. Sanders
Department of Languages, Literature and Communication, Utrecht Institute of Linguistics OTS, Utrecht University, Trans 10, 3512 JK, Utrecht, The Netherlands
T.J.M.Sanders@uu.nl
Ted J.M. Sanders is a Full Professor of Discourse Studies at the Department of Languages, Literature and Communication and the Utrecht institute of Linguistics OTS, Faculty of Humanities at Utrecht University. He published widely on discourse processing and representation, and on the (cross-) linguistics of coherence. He has a strong interest in improving communication through comprehensible texts. He has (co)supervised 23 PhD students. He is currently Vice-Dean of the Faculty, responsible for the *Graduate School of Humanities*.