

Signaling of Causal Coherence Relations in Spanish: Variety, Functionality, and Specificity in Academic Contexts

Romualdo Ibáñez

ROMUALDO.IBANEZ@PUCV.CL

*Pontificia Universidad Católica de Valparaíso
El Bosque 1290, Viña del Mar, Chile*

Fernando Moncada

FMONCADA@UBIOBIO.CL

*Universidad del Bío Bío
Av. Brasil 1180, Chillán, Chile*

Benjamín Cárcamo

BENJAMIN.CARCAMO@PUCV.CL

*Pontificia Universidad Católica de Valparaíso
El Bosque 1290, Viña del Mar, Chile*

Valentina Marín

VALENTINA.MARIN.B @MAIL.PUCV.CL

*Pontificia Universidad Católica de Valparaíso
El Bosque 1290, Viña del Mar, Chile*

Editor: Manfred Stede

Submitted 01/2019; Accepted 01/2020; Published online 03/2020

Abstract

While recent studies on Spanish have shown that some causal connectives specialize in expressing certain types of causal relations, others have revealed that causal relations may be signaled by a variety of linguistic devices. Given that we were interested not only in specificity and variety, but also in the functionality of causal connective expressions, our objective in the present study was threefold. First, to identify the variety of connective expressions (connectives and cue phrases) used to signal causal relations in Spanish. Second, to determine whether a relationship of specificity exists between connective expressions and particular types of causal relations. Third, to describe the functionality of those connective expressions. We analyzed a corpus of 2,514 causal coherence relations previously annotated and identified in a corpus of academic texts. 41 different linguistic devices used to signal causal relations were identified. These devices were grouped into two main functional classes: connectives and cue phrases. Regarding the functionality of the signals, we found that 8 of the most frequent connective expressions were used to signal different relations. As for specificity, in terms of syntactic categories, it was observed that various conjunctions and conjunctive adverbs specialize in signaling specific relations.

Keywords: Causal coherence relations, connectives, cue phrases, Spanish

1 Introduction

Based on the assumption that there is no one-to-one mapping between connective expressions and coherence relations, scholars have devoted considerable attention to the description of coherence relation signaling. This interest has given rise to studies that explore patterns of signaling from different theoretical and methodological approaches. From an approach that assumes coherence relations as explicit or implicit, studies have shown that some causal connectives specialize in expressing certain types of causal relations. This specialization is directly related to the notion of specificity proposed by Spooren (1997), who claims that a coherence relation is said to be specified, when it is marked by a connective that is prototypically used to encode its meaning, and underspecified when it is marked by one that is not. The notion of specificity has also been explored by scholars interested in causal connectives, who refer to them as *specified* causal connectives or *underspecified* causal connectives, depending on the degree of subjectivity a connective is involved with (Li et al., 2017). For example, the Dutch connective *dus* ('so') is frequently used in epistemic relations (Stukker et al., 2009) while *omdat* ('because') is predominant in volitional content ones (Sanders et al., 2012); according to this perspective, both can be considered specified, as they specify the degree of subjectivity the relations are involved with. This pattern of use has also been observed in other languages. The French *car* ('for') and *puisque* ('since') are mostly used in epistemic relations while *parce que*, in content ones (Anscombe & Ducrot, 1983; Degand & Pander Maat, 2003; Zufferey, 2010). In German, *weil* ('because') is predominant in content relations whereas *denn* ('because') is typical in epistemic ones (Stukker & Sanders, 2012). In Spanish, *porque* ('because'), *ya que* ('since') and *puesto que* ('given that') are typically used to signal causal coherence relations (Montolio, 2001); and *puesto que* ('given that') is preferred over *porque* ('because') to express subjective relations (Santana et al., 2018). A pattern of use different from specificity is that of polyfunctionality, which means that a connective can be used to signal different types of coherence relations (Fischer, 2006; Blackwell, 2016). In Spanish, connectives *porque* ('because') and *debido a* ('due to') may be considered polyfunctional, as they may signal either subjective or objective causal coherence relations (Santana et al., 2017; Santana et al., 2018; Cárcamo, 2019).

From a different approach, other researchers question the existence of implicit coherence relations (Taboada, 2009; Taboada & Das, 2013; Das, 2014; Das & Taboada, 2017). So, they focus on the variety of linguistic devices that may signal coherence relations, beyond connectives (or discourse markers). Studies that take this approach have shown that, in Spanish, the same coherence relation (concession) can be signaled by 18 different types of signals, including connectives (*pero*/'but'); other connective expressions, such as prepositional phrases (*para*+ NP, *Con*+NP); and a variety of linguistic devices, such as gerunds (*siendo*/'being'), and impersonal clauses (*bien es cierto que*/'it is well known that') (Taboada & Gómez-González, 2012). With regard to causal coherence relations in Spanish, it has been observed, not only that most of them (97%) are signaled, but also, that they may be signaled by linguistic devices other than connective expressions, such as lexical items (psychological verbs, such as *enojarse*/'to get angry'), non-finite verbs (*Miguel fue castigado por **llegar** tarde a casa*/'Miguel was punished for **arriving** home late'), and genre structure (Objective or Purpose coherence relations in the third rhetorical move of an abstract) (Duque, 2014).

Studies from these two approaches have shed light on important issues regarding the signaling of causal coherence relations in Spanish; however, there are still other aspects that have been understudied. One of them is the degree of specificity of the relation between particular types of coherence relations and connective expressions other than connectives (or discourse markers). Another aspect is the phenomenon of functionality of causal connective expressions, i.e. the possibility that a connective expression may signal different types of causal relations (Fischer, 2006; Redeker & Gruber, 2014). Finally, it also seems necessary to gather information regarding the interaction between connective expressions and coherence relations in different

contexts of use. To go further into the description of these phenomena would not only widen our understanding of the way coherence relations are signaled in Spanish, it would also contribute to the delineation of the way connective expressions and causal coherence relations interact across languages.

In order to account not only for specificity and variety, but also for functionality in the signaling of causal coherence relations in Spanish, we take an integrative approach. Hence, our objective is threefold. First, to identify the variety of connective expressions (connectives and cue phrases) used to signal causal relations in Spanish. Second, to determine whether a relationship of specificity between connective expressions and particular types of causal relations exists. Third, to describe the functionality of those connective expressions. To achieve our objectives, we analyzed a corpus of 2,514 causal coherence relations previously identified by Ibáñez et al. (2015). We added signaling information in what could be understood as a process of annotation upon annotation (Taboada & Das, 2013). Given that the analysis was carried out in causal coherence relations identified in texts belonging to different academic genres used in university programs of Biology and Law, this study provides information regarding the signaling of causal coherence relations in academic contexts.

The article is organized as follows. First, we provide an introduction to the concept of causal coherence relations. Second, we present our conception of the relation between causal coherence relations and different connective expressions. Third, we present the methods. Fourth, we present and discuss the results. Finally, we discuss the implications of those results and provide the conclusion.

2 Causal Coherence Relations

In the last four decades, several proposals have been developed for the description and classification of the different existent types of coherence relations (Mann & Thompson, 1988; Carlson & Marcu, 2001; Sanders et al., 1992; Hovy et al., 1992; Polanyi et al., 2004, among many others). In spite of the fact that such proposals vary considerably in terms of the types and number of relations or the specificity of the groupings, all of them include the group of causal relations. The predominance of such relations can be explained because of the importance that causality has not only in human cognition (Salmon, 1997; Sanders, 2005; Noordman & De Blijzer, 2000), but also in text organization (Meyer, 2000). In fact, it is claimed that all languages have specific means to express causality (Pander Maat & Sanders, 2000). For the Spanish language, extensive studies have been conducted to describe those means, paying special attention, among other aspects, to the encoding of causality in syntactic structures (Galán, 1999; Gutiérrez Ordoñez, 2002), the use of explicative causal conjunctions (Goethals, 2002), the relation between causativity, agentivity and transitivity (Gozalo Gómez, 2004) and the interaction between punctuation markers and connectives in causal constructions (Figueras, 2000) (for an excellent overview, see Arroyo, 2017).

Important attempts have been made to describe and classify the coherence relations that encode causality and to explain their system and use. One approach to describing coherence relations, particularly causal ones, is the Cognitive approach to Coherence Relations (CCR) (Sanders, et al., 1992). In CCR, causality is defined as the implicational meaning that can be inferred between consecutive discourse segments. Therefore, causality is not restricted exclusively to those causal relations that connect two events in the physical world (*fact 1 leads to fact 2*), but it is also present in cases where one event leads to a conclusion (*based on fact 1, someone concludes X*), where one event occurs given certain circumstances (*if X, then Y*), when someone performs an action to reach a purpose (*action X is done to achieve goal Y*), and even in cases where a final state/result is not expected (*X, however Y*). Therefore, relations typically

referred to as condition, purpose, or concession are considered causal in CCR and other frameworks based on it.

In its original version, CCR presents a set of four basic cognitive primitives that can be used to organize different types of coherence relations (mostly causal) that language users infer between two or more segments in a text: BASIC OPERATION, SOURCE OF COHERENCE, ORDER OF SEGMENTS and POLARITY. **BASIC OPERATION** indicates the strength of the semantic link between segment 1 (S1) and segment 2 (S2), and it has two values: Causal and Additive. A relation is *Causal* if there is an implication relation ($P \rightarrow Q$) between the two segments, in which P is antecedent and Q is consequent. On the contrary, in *Additive* relations the only relation that can be inferred between the segments is conjunction ($P \& Q$). **SOURCE OF COHERENCE** refers to the nature of the link established between the segments. It is *Semantic*, if the link is established at the level of the propositional content; or *Pragmatic*, if it is established at the level of illocutionary meaning. In further developments these values have been reformulated. *Semantic* relations have been referred to as *Content* relations while *Pragmatic* relations have been subdivided into *Speech Acts* (links motivated by illocutionary force) and *Epistemic* ones (connections that involve logical reasoning and inferences) (Spooren & Sanders, 2008). **ORDER OF SEGMENTS** accounts for the correspondence between each discourse segment and its role as either antecedent (P) or consequent (Q) in the coherence relation. Thus, there is *Basic Order* when P corresponds to the S1 and Q, to the S2. Conversely, the order is *Non-Basic* when the opposite sequence is present. Provided this criterion reflects two different orders in which P and Q can be presented in the connected discourse segments, it distinguishes cause-consequence relations (e.g. *S1. As a result, S2*) from consequence-cause relations (*S1 because S2*) or claim-argument (*S1 since S2*) from argument-claim (*S1 therefore S2*) relations. Finally, **POLARITY** distinguishes between Negative and Positive relations. A *Negative* relation holds when the relation between S1 and S2 involves the negation of the propositional content of one of the segments, while a *Positive* relation holds when there is no such a negation. Negative relations also involve the violation of the expectations generated by P, whereas in Positive relations, Q is in line with what can be expected according to P. For instance, in *Peter studied hard during the semester and he passed the course*, Polarity is positive since passing a course (Q) is something one can expect for someone who has studied hard (P). On the contrary, in *Peter studied hard during the semester but he failed the course*, Polarity is negative because failing a course (Q) is not a typical (or expected) result for someone who has studied hard (P). Since some negative relations involve an implication operation (as described above), they are considered as causal in CCR, different from cases like *Peter studied hard during the semester but I did not* (where there is no implication).

Let us consider examples (1), (2), and (3) to illustrate the classification of causal relations according to the CCR approach.

- (1) [En la regulación alostérica, el efector se combina con la enzima en un lugar diferente del centro activo, denominado centro alostérico.] s1 Por ello, [ocurre una modificación en el centro activo de la enzima.] s2

[‘In allosteric regulation, the effector combines with the enzyme in a different place of the active center, named allosteric center.’] s1 Because of that, [a modification in the enzyme active center occurs’]. s2

- (2) [Los neurotransmisores regulan la transmisión de impulsos nerviosos.]s1 Por lo tanto, [desempeñan un papel fundamental en el funcionamiento del sistema nervioso.] s2

[‘Neurotransmitters regulate the transmission of nervous impulses.’] s1 Therefore, [they play a crucial role in the functioning of the nervous system.’] s2

- (3) [Los cloroplastos transfirieron menos DNA, en comparación con las mitocondrias.] s1 Por lo tanto, [es posible que la endobiosis de las mitocondrias haya ocurrido antes de la endobiosis que originó los cloroplastos.] s2

[‘Chloroplasts transferred less DNA compared to mitocondria.] s1 Therefore, [it is possible that the endobiosis of mitocondria had occurred before the endobiosis that created chloroplasts.’] s2

In (1), (2) and (3), the Basic Operation is *Causal* since a relation of implication can be deduced between S1 and S2. In all cases, the Polarity is *Positive* since S2 does not imply a negation of S1. The Order of Segments is *Basic*, given that S1 operates as the antecedent (P) and segment 2, as the consequent (Q). One may conclude, then, that these examples reflect the same type of causality and, therefore, that they could be classified under the same label. However, the difference between these examples can be identified by taking into account their Source of Coherence. According to Spooren and Sanders (2008), in (1) a *Content* relation holds since the states of affairs described in S1 and S2 occur in the physical world. On the contrary, in (2) a *Speech Act* relation holds given that S2 is a claim by the author based on evidence presented in S1. In (3), the relation is *Epistemic* since S2 is an inference made by the author based on the evidence provided in S1. Therefore, according to the CCR approach, (1) is a case of Non-volitional cause, while (2) is Evaluation, and (3) Interpretation.

CCR has become an important discourse annotation scheme. It has been used in several studies covering languages such as German (Pit, 2003), Dutch (Stukker, 2005; Spooren & Sanders, 2008), Mandarin Chinese (Li et al., 2013; Wei, 2018) and recently, Spanish (Santana et al., 2018). Since the original 1992 proposal, several modifications and updates have been proposed for the interpretation and operationalization of the original primitives (see Hoek, 2018; Hoek et al., 2019 for a state-of-the-art in CCR). For instance, a further distinction has been made in Source of Coherence, regarding the degree of *Subjectivity* causal relations may encode. The notion of Subjectivity refers to the degree to which a reasoning entity or Subject of Consciousness (SoC) is involved in the construal of the coherence relation (Degand & Pander Maat, 2003; Pander Maat & Degand, 2001; Pander Maat & Sanders, 2000, 2001). The more present an SoC is in the construction of the relation, the more subjective the relation is (Epistemic relations). Hence, if the SoC is absent, a relation will be objective (Non-volitional content relations). Following those principles, (1) can be classified as an objective relation because the causal link is established between two events of the physical world, without the intervention of an SoC. On the other hand, (2) and (3) are considered subjective since the causal link is mediated by the participation of an SoC.

CCR’s original primitives and later reformulations have greatly influenced other approaches to coherence relations. Such is the case of the top-down bottom-up approach developed by Ibáñez et al. (2015) and Ibáñez et al. (2019) for the Spanish language. This approach has been used to annotate corpora of academic genres and school textbooks of different disciplines, allowing the identification of types of relations that vary in their frequency depending on the discipline. For instance, when analyzing a corpus of academic genres used in Law and Biology, Ibáñez, et al. (2015) identified that causal relations that involve obligation (as (4) below) were highly frequent in Legal texts but were not present in the texts from Biology. These relations, labeled as Condition-Obligation, differ from the other conditional relations of the taxonomy in that the consequent (Q) is not an action that an agent performs voluntarily (Condition-Action) nor is a state that results when a condition is met (Condition-Event) but is an action that an agent is required to do (See Appendix).

- (4) [Si el ciudadano no respeta la norma,] s₁ [debe someterse a juicio.] s₂

['If the citizen shall not respect the law,'] s₁ ['he or she must be judged.'] s₂

As described in Section 3.1, the study reported in this paper was based on the taxonomy of Ibáñez et al. (2015), and it was conducted upon their corpus of causal coherence relations extracted from academic genres.

2.1 Causal coherence relations and connective expressions

Despite the considerable amount of research on coherence relations and their signals, there is still no general consensus on some fundamental issues. The first has to do with the nature of the signal. To some researchers, coherence relations are explicit, when they are signaled by a connective expression that indicates the link between the discourse segments, and implicit, when there is no connective expression involved (Knott & Dale, 1994; Meyer & Webber, 2013; Fraser, 2009; van der Vliet & Redeker, 2014). Others claim that there are no implicit coherence relations, given that coherence relations may also be signaled by a range of linguistic devices whose primary function is not to mark coherence relations, such as non-finite verbs, genre-structure, punctuation, lexical items, or even sentence mood (Taboada, 2009; Taboada & Das, 2013; Das & Taboada, 2017; Das, 2014). Another issue in which there seems to be no consensus has to do with the names given to connective expressions. In the literature, they are referred to as *cue phrases* (Knott & Dale, 1994; Knott, 1996; Knott & Sanders, 1998), *discourse markers* (Schiffrin, 1987; Portolés, 1998; Zorraquino & Portolés, 1999; Taboada, 2006; Redeker & Gruber, 2014), *pragmatic markers* (Fraser, 1999, 2009), *discourse operators* (Redeker, 1990, 1991), *discourse particles* (Fischer, 2006; Aijmer, 2002; Briz et al., 2008), *connectives* (Pons, 1998; Degand & Pander Maat, 2003; Van der vliet & Redeker 2014; Evers-Vermeul et al., 2017), among other names (Fraser, 2009). There is also lack of consensus regarding the way connective expressions are conceived. To some researchers (Redeker & Gruber, 2014), connectives are the same as discourse markers, while to others (Portolés, 1993, 1998; Zorraquino & Portolés, 1999; Duque, 2014), the former correspond to a subgroup of the latter. Simultaneously, some other researchers (Fraser, 1996, 2009) understand discourse markers as part of pragmatic markers, while others (Knott & Dale, 1994; Knott, 1996) use cue phrase as a covering term for all kinds of expressions that signal coherence relations, including discourse markers.

Despite the terminological and conceptual heterogeneity (Das 2014), there is consensus on the fact that connective expressions comprise a functional class that signals the coherence relation that holds between two discourse segments. Though, there is no one-to-one link between connective expressions and the type of coherence relation they signal. In fact, sometimes, a type of coherence relation may be signaled by different connective expressions (*Condition* relation may be signalled by *if*, *unless*, *when*, and *since*), showing lack of specificity (Spooren, 1997), while some other times, the same connective expression may signal different coherence relations – a phenomenon identified as ambiguity by some researchers (Stede, 2014) and as a polyfunctional profile of use by others (Fischer, 2006; Redeker & Gruber, 2014) (connective *but* may signal relations of *Contrast*, *Concession* and *Antithesis*).

As a functional class, connective expressions are drawn from different syntactic classes, such as subordinating and coordinating conjunctions, complex prepositions, and conjunctive adverbs (Fraser, 1999, 2009; Redeker & Gruber, 2014). They can be one word (such as adverbs) or multi-word expressions (such as complex prepositions) (Stede, 2014; Crible & Cuenca, 2017). They may be fixed expressions, such as conjunctive adverbs, or less frozen expressions, such as prepositional phrases or other unsystematic constructions (Das, 2014; Duque, 2014). Descriptions of connective expressions in Spanish treat multi-word fixed expressions as locutions (Pons, 1998; Zorraquino & Portolés, 1999; Galán, 1999; Real Academia Española, 2019). Hence, and

depending on the syntactic category the multi-word fixed expression is associated to, it may be classified as prepositional location (*a pesar de*/'in spite of'), adverbial location (*sin embargo*/'however'), or conjunctive location (*puesto que*/'given that') (Pavón Lucero, 1999).

In the present study, two main types of causal connective expressions are distinguished: *causal connectives* (CC), conceived as one-word or multi-word invariable expressions, whose main function is to signal the causal coherence relation that holds between discourse segments; and *cue phrases* (CP), which comprise those less frozen expressions that signal the causal coherence relation that holds between discourse segments, and may allow for syntactic modification. While the former are drawn from syntactic categories, such as conjunction (*porque*/'because') – including conjunctive location (*ya que*/'given that') –, conjunctive adverb (*sin embargo*/'however'), and complex preposition (*a pesar de*/'in spite of'), the latter are usually expressed by prepositional phrases (*por esa razón*/'for that reason') and other unsystematic constructions (*preposition + infinitive*).

| | Coherence relation | Frequency |
|----|----------------------|-----------|
| 1 | Cause effect | 228 |
| 2 | Effect causes | 33 |
| 3 | Action reason | 69 |
| 4 | Reason action | 59 |
| 5 | Act purpose | 112 |
| 6 | Purpose act | 83 |
| 7 | Claim argument | 446 |
| 8 | Argument claim | 311 |
| 9 | Condition event | 374 |
| 10 | Event condition | 264 |
| 11 | Condition obligation | 128 |
| 12 | Obligation condition | 7 |
| 13 | Basic contrast | 281 |
| 14 | Non basic contrast | 70 |
| 15 | Evidence deduction | 47 |
| 16 | Deduction evidence | 2 |
| | Total | 2,514 |

Table 1. Types and instances of causal coherence relations of the corpus.

3 Methods

In order to account for variety, functionality, and specificity of causal connective expressions, we followed the approach adopted by Taboada and Das (2013). We added a new layer of information (causal connective expressions) to a corpus already annotated for coherence relations (Ibáñez et al., 2015). The original corpus consisted of 27 complete exemplars (762,737 words) of academic genres (Textbook, Disciplinary Text, and Research Article), written in Spanish and used in undergraduate programs of Law and Biology.

3.1 Corpus

The corpus of the present study corresponds to the 2,514 causal coherence relations identified in the original corpus (Ibáñez et al., 2015), which are classified into sixteen types, as shown above in Table 1.

3.2 The annotation of connective expressions

The annotation of connective expressions was carried out following two sequential steps: *identification* and *classification*.

Identification: this procedure started by distinguishing explicit from implicit relations based on whether or not they were marked, following some of Taboada and Das (2013) and Das (2014) conditions for considering an expression to be a signal: 1. The scope of the function of a signal is a single discourse sequence comprising adjacent discourse segments in a relation. 2. Signals mark relations that hold between two discourse segments. 3. Signals constitute a functional class of lexical expressions drawn from different syntactic classes. As shown in (5), the two discourse segments (in square brackets) are connected by a Claim-Argument relation. The relation is explicit since it is signaled by the connective *ya que*.

- (5) [Tales juicios no son apropiados] _{s1}, ya que [no resuelven el asunto de manera definitiva.]_{s2}

[‘Such trials are not appropriate,’] _{s1} since [‘they do not definitively solve the issue.’] _{s2}

Classification: In order to classify the connective expressions identified in our corpus, an extensive bibliographical review of different proposals of Spanish causal connective expressions was carried out (Pavón Lucero, 1999; Domínguez García, 2007; Martí, 2008; Martínez, 1997; Montolio, 2001; Portolés, 1998; Pons, 1998; Zorraquino & Portolés, 1999; Galán, 1999; Real Academia Española, 2019). Most of these proposals classify connectives as functional categories, according to the type of coherence relation they signal (*concessive marker/connective*, *contrastive connective/marker*, *additive connective/marker*, etc.). Given that this study is driven by the assumption that there is no one-to-one correspondence between signals and types of relations, and that, as functional expressions, they are drawn from different syntactic classes, causal connective expressions were also classified according to syntactic classes (Pons, 1998; Real Academia Española, 2019). Hence, for the functional class of causal connective, we used the categories of conjunction, conjunctive locution, conjunctive adverb, and complex preposition, while for the functional class of cue phrase, we used the categories of prepositional phrases and constructions (*preposition+infinitive*).

3.3 Procedure

Due to the number and variety of forms and, mainly, to their use in combination, classification of causal signals is neither simple nor objective. For that reason, in order to ensure inter rater agreement, a two-coders-discuss strategy was used (Spooren & Degand, 2010). After a training period of one month, two pairs of coders (all postgraduate students in Linguistics) were assigned the 2,514 causal relations. Each pair was asked to identify and classify the causal connective expressions in the whole corpus, according to the categories selected for the study. First, causal connective expressions were classified independently by the two coders in each pair. After that, in case of disagreement, differences were discussed until agreement was reached by the pair. Subsequently, the results of both pairs were compared. The agreement index was $k = 0.78$, which can be interpreted in this context as substantial (Landis & Koch, 1977; Spooren & Degand, 2010).

4 Results

In this section, and based on the analysis performed to our corpus of academic texts, we first present the variety of connective expressions used to signal causal coherence relations in Spanish. Then, we show the functionality of the most frequent connective expressions, and, finally, the degree of specificity of the relation between those signals and particular causal relations.

4.1 Variety of connective expressions

The first objective of the present study was to identify the variety of connective expressions used to signal causal relations in Spanish. Out of the 2,514 causal relations analyzed, 2,284 (92%) were signaled, while only 230 were not (8%), which shows the predominance of explicit causal relations in the corpus. This is in line with prior research that has established that Spanish coherence relations tend to be signaled in some manner most of the time (Taboada & Das, 2013; Duque, 2014). The analysis also revealed the existence of 41 different linguistic devices, which were grouped into two functional classes (causal connectives and cue phrases) and six syntactic categories, as shown in Table 2.

Table 2 shows the wide variety of linguistic devices used to signal causal relations in the corpus of study. It can be observed that out of the two functional classes, *connective* encompasses most cases of causal coherence relations signaling. By far, the most frequent syntactic class observed is *conjunction*, which constitutes on its own 51.72% of all the linguistic devices identified in the corpus. This is significant considering conjunction only includes 9 of the devices used. The second highest frequency can be attributed to conjunctive adverb, which constitutes 14.93%. As for cue phrase, the linguistic devices are very similarly distributed between prepositional phrases and preposition + infinitive construction, which constitute 7.97% and 9.16% respectively. Based on this data, it can be suggested that causal coherence relations are prototypically signaled by conjunction in Spanish written texts, in the academic contexts of Biology and Law. These findings contribute to the study of signaling devices in Spanish, considering most of the research conducted on this matter is diachronic rather than synchronic (García-Cervigón, 2006; Herrero, 1999; Zagona, 2002) or purely theoretical (Cid, 2002) rather than empirical.

| Functional class | Syntactic category | Occurrence | Percentage | Linguistic devices |
|------------------|--|------------|------------|---|
| Connective | Conjunction | 1174 | 51.72% | Si, aunque, porque, como, pero, pues, cuando, mientras, así |
| | Conjunctive locution | 322 | 14.19% | ya que, para que, puesto que, siempre que, de modo que, a menos que, aun cuando, salvo que, dado que, con tal que |
| | Conjunctive adverb | 339 | 14.93% | Sin embargo, por lo tanto, si bien, por consiguiente, en consecuencia, por cuanto, no obstante, una vez, entonces |
| | Complex preposition | 46 | 2.03% | a pesar de, a fin de, en caso de |
| Cue phrase | Prepositional phrase | 181 | 7.97% | Por eso, por lo que, por ello, en este caso, por lo mismo, por lo cual, por esa razón |
| | Preposition + infinitive constructions | 208 | 9.16% | Para + infinitivo, por + infinitivo, al + infinitive |

Table 2. Connective expressions, classification and occurrence.

4.2 Functionality of connective expressions

In order to describe the functionality of the causal connective expressions, the 10 most frequent linguistic devices identified in our corpus (which represent 68.3% of the signaled relations) were examined. The results are shown in Table 3.

As it can be observed, 9 of the 10 most frequent linguistic devices belong to the category of connective and one to cue phrase. Our data confirms that most linguistic devices are polyfunctional, since they signal various types of causal relations, ranging from two (*aunque*/'although') to five (*porque*/'because'), even though there are others that only signal one relation such as *pero* ('but') and *sin embargo* ('however'). Table 3 shows that the conjunction *si* ('if'), the most frequent in the corpus, allows the signaling of four types of relations that involve conditionality. Over half of the cases in which *si* is used correspond to Condition-event (53%), a relation that follows a Basic Order (antecedent-consequent) and whose consequent depicts a

| Connective expressions | Functional class | Frequency | % | Coherence relations signalled |
|---|------------------|-----------|------|--|
| <i>Si</i> ('if') | Connective | 434 | 19 | Event-condition (25%) Condition-event (53%) Condition-obligation (21%) Obligation-condition (1%) |
| <i>Porque</i> ('because') | Connective | 200 | 8.75 | Action-reason (14%) Claim-argument (74%) Argument-claim (2%) Deduction-evidence (1%) Effect-Cause (9%) |
| <i>Para</i> + <i>infinitivo</i> ('to + infinitive') | cue phrase | 171 | 7.48 | Act-purpose (49%) Purpose-act (44%) Event-condition (7%) |
| <i>Cuando</i> ('when') | Connective | 137 | 5.99 | Event-condition (40%) Condition-obligation (6%) Condition-event (54%) |
| <i>Pues</i> ('since') | Connective | 133 | 5.82 | Action-reason (8%) Claim-argument (77%) Effect-Cause (2%) Argument-claim (13%) |
| <i>Ya que</i> ('since') | Connective | 125 | 5.47 | Action-reason (8%) Argument-claim (3%) Claim-argument (86%) Effect-Cause (3%) |
| <i>Sin embargo</i> ('however') | Connective | 102 | 4.46 | Basic contrast (100%) |
| <i>Pero</i> ('but') | Connective | 90 | 3.94 | Basic contrast (100%) |
| <i>Aunque</i> ('although') | Connective | 85 | 3.72 | Basic Contrast (51%) Non basic contrast (49%) |
| <i>Por lo tanto</i> ('therefore') | Connective | 83 | 3.63 | Argument-claim (72%) Cause-effect (18%) Evidence-deduction (6%) Reason-action (4%) |
| | | 1,560 | 68.3 | |

Table 3. 10 Most frequent causal connective expressions and the relations they signal.

situation occurring in the physical world without intentionality. The same conjunction is also used in Event-Condition (25%), a relation that differs from the previous one only in terms of the order in which the segments are presented (consequent-antecedent). This conjunction frequently also signals relations whose consequent represents an obligatory situation (Condition-obligation, 21%). Another interesting profile of use is observed in the conjunction *porque* ('because'), since it is a device that signals the widest variety of relations (5). Among them, it signals relations mediated by intentions (Action-reason, 14%), relations that involve speakers' stance (Claim-argument (74%) and Argument-claim (2%)), and relations that express causality in the physical world (Effect-cause, 9%).

4.3 Specificity of Connective expressions

In order to determine whether a relation of specificity between connective expressions and types of causal relations exists, the chi-square goodness of fit test (χ^2) was used. This statistical test assesses the distribution of categorical data and compares it against a specific proportion by which observed and expected frequencies are examined. A χ^2 was run for each linguistic device based on the coherence relations they signaled. In all instances an even hypothesized distribution was assumed for each test. This decision was methodologically important since it allowed us to differentiate the expectations of each of the devices. When a connective was found to signal different coherence relations, it was assumed as the null hypothesis for the chi-square goodness of fit test that it would not show any preference toward a particular coherence relation. In those cases where the null hypothesis was rejected, adjusted standardized residuals were used to complement the significance of the statistical test. Those cases where the residuals were higher than 3.0 were considered to represent specificity, since this number has been considered to reflect whether the number of observations is significantly larger than expected (Agresti, 2018). Following this procedure, several connective expressions were found to signal specific types of coherence relations, as shown in Table 4.

It is worth mentioning that there are two connectives (*pero*/'but' and *sin embargo*/'however') that are used exclusively in Basic Contrast relations. It should be noted that in (Ibáñez et al., 2015), and similar to CCR (although with another label), this type of relation is regarded as causal negative because a relation of implication holds between the connected discourse segments. Such implication, in turn, involves the generation of expectations, which in negative relations are not fulfilled. In other words, in negative relations, like Basic Contrast, the consequent (Q) is not in line with the expectations triggered by the antecedent (P), as illustrated in (6) and (7) below. Given that *pero* ('but') and *sin embargo* ('however') are used only in Basic Contrast relations, both connectives could be regarded as having a profile of specificity towards Basic Contrast.

- (6) En general, las células de los cultivos primarios mueren después de un cierto número de mitosis (50 a 100 mitosis). Pero a veces algunas células experimentan mutación y se hacen inmortales.

'In general, cells of primary cultures die after a certain number of mitosis (50 to 100 mitosis). But sometimes some cells experience mutation and become immortal.'
(Basic Contrast/Biology/Handbook)

- (7) El sistema nos parece mucho más adelantado que el de la República Argentina y el de Brasil. Sin embargo, su fundamento es inadmisibile y utilitario.

'The system seems to us to be much more advanced than that of Argentina's and Brazil's. However, the argument is inadmissible and utilitarian.'
(Basic Contrast/Law/Handbook)

| Connective expressions | Observed frequency | Expected frequency | Chi-square goodness of fit test | Specific coherence relation |
|---|--------------------|--------------------|-----------------------------------|------------------------------------|
| <i>Cuando</i> (‘when’) | 71 55 | 34 34 | $(\chi^2(3) = 100.635, p = <.05)$ | Condition-Event Event-Condition |
| <i>Para + inf.</i> (‘to+ infinitive’) | 84 75 | 43 43 | $(\chi^2(3) = 128.485, p = <.05)$ | Act-Purpose Purpose-Act |
| <i>Por lo tanto</i> (‘therefore’) | 59 | 17 | $(\chi^2(4) = 142.361, p = <.05)$ | Argument-Claim |
| <i>Porque</i> (‘because’) | 147 | 40 | $(\chi^2(4) = 369.350, p = <.05)$ | Claim-Argument |
| <i>Pues</i> (‘since’) | 102 | 27 | $(\chi^2(4) = 273.729, p = <.05)$ | Claim-Argument |
| <i>Si</i> (‘if’) | 229 | 109 | $\chi^2(3) = 237.318, p = <.05$ | Condition-Event |
| <i>Ya que</i> (‘since’) | 108 | 31 | $(\chi^2(3) = 252.248, p = <.05)$ | Claim-Argument |

Table 4. Observed and expected frequency of the most frequent connective expressions across the coherence relations they signal.¹

Among those connectives used to signal more than one relation, a clear pattern of specificity was found for *si* (‘if’). This conjunction signals various types of conditional relations, but, as shown in Table 4, it is predominantly used in Condition-Event. This relation is established when the (non) occurrence of one or more events determine the (non) occurrence of others, as illustrated in (8) and (9).

¹ *Sin embargo* (‘however’) and *pero* (‘but’) have not been included in Table 4 because all observed instances are Basic contrast. *Aunque* (‘although’) has not been included because it shows an even distribution between Basic contrast (N=43) and Non-basic contrast (N=42) ($\chi^2(1) = .012, p = 0.914$).

- (8) Si un aparato mitótico aislado se trata con estas sustancias, su estructura se pierde rápidamente.

‘If an isolated mitotic apparatus is treated with these substances, its structure is lost quickly.’

(Condition-Event/Biology/Handbook)

- (9) Si los tribunales nacionales carecen de ella, la ley fija su falta de competencia.

‘If national courts do not have it, the law fixes its lack of competence.’

(Condition-Event/Law/Handbook)

Another clear case of specificity was found for *porque* (‘because’) and *pues* (‘since’), and *ya que* (‘since’). As mentioned above, these connectives signal a wide repertoire of causal relations: *porque* signals 5 types of relations, and *pues* and *ya que*, 4 (see Table 3). Interestingly, Table 4 shows that in spite of such polyfunctionality, these connectives show specificity for the same relation: Claim-Argument. This causal relation, as illustrated in (10), (11) and (12), involves the statement of an opinion or judgment, which is expressed in the first segment of the causal relation and functions as the consequent (Q). The evidence or arguments for such a claim is presented in the second segment of the causal relation and functions as the antecedent (P).

- (10) Será necesaria la prueba indirecta, porque el hecho no está presente o ha dejado de existir.

‘The indirect evidence will be required because the fact is not present or has ceased to exist.’

(Claim-Argument/Law/Disciplinary Text)

- (11) No es sólo una simple mezcla de estas sustancias, pues el protoplasma tiene una organización muy compleja.

‘It is not just a simple mixture of substances, since the protoplasm has a very complex organization.’

(Claim-argument/Biology/Disciplinary text)

- (12) La glicolisis es un proceso poco eficiente, ya que de las 690 Kcal mol presentes en la glucosa, apenas 20 son aprovechadas.

‘Glycolysis is not an efficient process since out of the 690 Kcal present in glucose, only 20 are used.’

(Claim-argument/Biology/Disciplinary text)

Another profile of specificity was found for *cuando* (‘when’). This connective, usually used as a signal of temporality in Spanish, is also frequently used for causal relations, particularly in those that involve conditionality. Although *cuando* is used to signal 3 types of relations (see Table 3), it shows a preference for two: Condition-event and Event-condition. These relations are used to present a state (P), whose occurrence results in the occurrence of another (Q), as shown in (13), or to present a state (Q), whose occurrence results from the occurrence of another (P), as shown in (14).

- (13) La presión predatoria tiende a aumentar cuando crece la población de presas.

‘Predatory pressure tends to increase when prey population grows.’
(Event-Condition/Biology/Disciplinary text)

- (14) Cuando son colocadas en una solución hipertónica, las células disminuyen de volumen.

‘When placed in a hypertonic solution, cells decrease their volume’.
(Condition-Event/Biology/Disciplinary text)

In addition, while conducting the analysis, it was noticed that the connective expressions also show patterns of specificity towards subjective or objective causal relations when considering the nature of the coherence relations they predominantly signal. For example, following Ibáñez et al. (2015), (15) is a case of Condition-Event, a relation whose Source of Coherence is non-volitional. In cases like this, the link between the antecedent (P) and the consequent (Q) is established in the physical world and there is no volition or obligation involved in the state of affairs expressed in Q. Therefore, it can be regarded as an objective relation. On the contrary, in (16) an Argument-Claim relation holds, where the Source of Coherence is Speech Act. In cases like this, Q corresponds to a claim used by the speaker with the support of the argument presented in P. Therefore, the speaker is responsible for the construal of the causal relation. Thus, (16) is considered a subjective relation.

- (15) Cuando las células musculares o hepáticas son expuestas a la hormona adrenalina, hay un aumento en el contenido intracelular de cAMP.

‘When muscular or hepatic cells are exposed to the adrenaline hormone, there is an increase in the intracellular content of cAMP.’
(Condition-event/Biology/Handbook)

- (16) Ya que gobernar y ser gobernado es algo diferente, hemos de suponer que la moderación de los gobernantes no es idéntica a la moderación de los gobernados.

‘Since to govern is different from being governed, we have to expect that the moderation of the leaders is not identical to the moderation of those being governed.’
(Argument-claim/Law/Disciplinary text)

Following this line of reasoning, *cuando* (‘when’), *pero* (‘but’), *sin embargo* (‘however’) and *aunque* (‘although’) can be regarded as showing a preference for objective relations since they predominantly signal objective relations such as Condition-Event and Basic Contrast. On the other hand, *por lo tanto* (‘therefore’), *porque* (‘because’), *pues* (‘as’), and *ya que* (‘since’) may be considered as subjective since they predominantly signal subjective relations such as Argument-Claim and Claim-Argument (see Table 4).

The results reported up to this point reveal that regardless of the fact that some connectives show a polyfunctional profile (i.e., they signal different types of relations), they feature a clear pattern of specificity. These findings lead us to suggest that, at least in two academic contexts (Biology and Law), some causal connectives are highly specific for particular relations. From another perspective, it could be argued that certain causal relations are predominantly signaled by specific connective expressions. Therefore, further studies could focus on specific sets or subsets of causal relations to determine whether such relations show a preference for certain connectives.

5 Discussion and Conclusions

The relation between coherence relations and connectives has been understudied in Spanish. In fact, most research has traditionally been aimed at providing comprehensive descriptions and grammar-oriented classifications of connectives, without considering coherence relations (Gili Gaya, 1961; Portolés, 1993, 1998; Zorraquino & Portolés 1999, Fuentes 1987, 1996; Iglesias Recuero, 2000; Caravedo, 2003; Arroyo, 2017). More recent discourse-oriented studies have switched the focus, paying special attention to the relation between coherence relations and connectives. In this scenario, it is possible to identify, at least, two types of studies on explicit causal coherence relations: those interested in describing the prototypical use of causal connectives (Cao et al., 2016; Santana et al., 2017; Santana et al., 2018) and those intended to account for the variety of linguistic devices that may signal causal relations (Taboada & Gómez-González, 2012; Duque, 2014). To the best of our knowledge, there have not been previous studies combining both approaches. Therefore, the current corpus-based study aimed not only to describe the variety of linguistic devices used to signal causal relations in the Spanish language, but also to explore their functionality and specificity. In line with Taboada and Das' suggestions (2013), instead of starting from scratch, we re-used an already annotated corpus to which we added an extra layer (their signaling). We used a corpus of 2,514 causal coherence relations previously annotated (Ibáñez et al., 2015), which were identified in academic written texts in Spanish.

The manual analysis carried out enabled us, in the first place, to identify 41 different connective expressions used to signal causal relations. These devices were grouped into two main classes: connectives and cue phrases. Within the class of connectives, four types were distinguished: conjunctions (*si*/'if', *porque*/'because', *pues*/'since', *pero*/'but'), conjunctive locutions (*puesto que*/'given that', *ya que*/'since'), conjunctive adverbs (*sin embargo*/'however') and complex prepositions (*a pesar de*/'in spite of'). In the case of cue phrases, two types were identified: Prepositional phrases (*por esa razón*/'for that reason') and preposition + infinitive constructions (*para conseguir*/'to achieve'). An interesting finding is that out of the 10 most frequent connective expressions identified, 9 are connectives. Among these, the most frequent one is the conjunction *si* ('if'), which is coherent with the predominance of conditional relations in the corpus. This reinforces the idea that connectives are the prototypical means of marking coherence relations in Spanish. Furthermore, in line with previous studies in Spanish (Montolío, 2001; Domínguez García, 2007), our data shows that *porque* ('because') is one of the most typical ones. Actually, in our corpus it has the second highest frequency, which supports the idea that this conjunction is one of the most commonly used to express causality in Spanish (Domínguez García, 2007). Some of the other frequent conjunctions in our data, such as *ya que* ('since'), have been characterized as typical signals of causality in Spanish in previous research (Pit et al., 1996; Goethals, 2002; Duque, 2016; Santana et al., 2018). Therefore, our results contribute to provide new empirical evidence on the most typical signals used to mark causality in the Spanish language.

Regarding the functionality of the connective expressions, we analyzed the 10 most frequent ones and found that 8 of them signal more than one relation. An interesting case was observed in the conjunction *porque* ('because'). As claimed in previous studies (Montolío, 2001; Domínguez García, 2007), this conjunction is one of the most frequently used because it allows to express different types of causal relations. Our data provides evidence not only for its high frequency in language use, but also for its polyfunctionality. Specifically, our study shows that *porque* signals 5 different types of relations, ranging from those that involve volition (Action-reason) to those mediated by the speaker's stance (Claim-Argument, Argument-Claim) to those belonging to the content domain (Effect-Cause). A similar case was observed in the profile of use of *cuando* ('when'), a conjunction which, in the Spanish language, is often used as a signal for temporality (*Cuando era joven, solía viajar en motocicleta*, 'When I was young, I used to ride a motorbike').

Our data shows that *cuando* has the fourth highest frequency, which proves that its use in causal relations is not peripheral. In fact, *cuando* allows the signaling of 3 different types of conditional relations, having the same meaning as *si* ('if'). Cases like these, when the semantics of the connective used to explicit a relation does not fully match the semantics of the relation that is intended by the speaker/writer, is what Spooren (1997) calls underspecification. Previous psycholinguistic research (Spooren, 1997; Li et al., 2017) has shown that underspecified coherence relations impose different cognitive efforts compared to those relations signaled by a typical (specific) connective. Therefore, further research is needed to investigate whether such differences are also found in the Spanish language.

Regarding specificity, and based on previous literature, we expected that certain connective expressions would show a preference for specific causal relations. Our results were in line with our expectations since it was observed that various conjunctions and conjunctive adverbs specialize in signaling specific relations. For instance, *pero* ('but') and *sin embargo* ('however') are used to signal exclusively Negative causal relations (Basic Contrast). Other cases are the conjunctions *si* ('if'), which is used mostly in conditional relations (Condition-Event), and *pues* ('since'), which is used frequently in relations involving a speaker's stance (Claim-Argument). In addition, based on the causal relation they frequently signal, our data showed that certain connective expressions show a preference for subjective meanings. *Por lo tanto* ('therefore'), *porque* ('because'), *pues* ('as'), and *ya que* ('since') are mostly used in subjective relations, where the speaker is involved in the construal of the relation, for instance through the statement of a conclusion or an opinion. These findings differ from previous corpus-based studies. Santana et al (2018), for instance, found that *porque* ('because'), and *ya que* ('since') were not associated with subjective features, which led them to conclude that Spanish does not have connectives that have a clear subjectivity profile. They argued that Spanish *porque*, similar to English *because*, is used to express both subjective and objective relations (Sweetser, 1990; Knott & Sanders, 1998).

A possible explanation for this difference is how subjectivity was operationalized. Santana et al. (2018) used an analytical model that decomposes the notion of subjectivity in a series of components (Domain, Modality, Presence of the SoC and Identity of the SoC), whereas in our study the distinction between objective and subjective relations was based only on Sweetser's (1990) distinction for the Source of Coherence of the relation. Besides, since our analysis was performed on academic genres only, it could be possible that the subjective pattern shown by *porque* and *ya que* may be due to the predominantly argumentative nature of the genres they were extracted from and of the disciplines (Law). It could be that in other (non-academic) genres, with other discourse organization modes (such as narrative or descriptive), these connectives may display more objective profiles. In consequence, further research is required to extend our understanding of subjectivity in Spanish and the devices that express it. Complementary, psycholinguistic studies could provide us with new insights about the processing of connectives depending on their degree of specificity towards subjectivity. It could be that some differences may be observed in the processing and comprehension of subjective relations signaled by connectives that vary their degrees of specificity. Specifically, it could be expected, for instance, that a Claim-argument relation is processed faster when it is signaled by *ya que* than by *porque*.

The main contribution of the current study is the approach adopted to account for the signaling of causal relations in Spanish. Different from previous studies that have analyzed either how a particular relation is marked or whether a specific marker is associated to specific relations, we combined both perspectives. In addition, given that we did not construct a taxonomy of connective expressions to analyze their profiles of use, our results provide a richer picture on the ways causal relations are actually signaled and on how those connective expressions are used, at least in academic written texts. Further research, focused on spontaneous conversations, for instance, may shed new light on usage pattern of causal connective expressions in conversations, which may differ from the patterns reported here. Studies such as the ones conducted by Günthner (1993) and Keller (1995) have demonstrated that a connective like German *weil*

(‘because’) usually expresses epistemic relations in conversations, while in written texts it is used for content domain relations. Therefore, it could be possible that a typical connective like *porque* may present different usage patterns depending on modality.

Acknowledgments

The first author’s work was enabled by a grant awarded by FONDECYT (project 1160094) from the National Commission for Scientific and Technological Research (CONICYT). The second author was funded by postdoctoral grant FONDECYT 3180779 from the National Commission for Scientific and Technological Research (CONICYT). The third author was funded by doctoral grant CONICYT-PFCHA/Doctorado Nacional/2017-21170031.

Appendix. Causal coherence relations used by Ibáñez et al. (2015)

| CAUSAL COHERENCE | | | | | | |
|------------------|----------|---------------------|------------------|----------------------|----------------|--------------------|
| Order of Events | Polarity | Source of Coherence | | | | |
| | | Content | | | Speech Act | Epistemic |
| BASIC | POSITIVE | Cause-Effect | Reason-Action | Condition-Obligation | Argument-Claim | Evidence-Deduction |
| NON BASIC | POSITIVE | Effect-Cause | Action-Reason | Obligation-Condition | Claim-Argument | |
| BASIC | NEGATIVE | Basic Contrast | | | | |
| NON BASIC | NEGATIVE | Non Basic Contrast | | | | |
| BASIC | POSITIVE | Condition-Event | Condition-Action | | | |
| NON BASIC | POSITIVE | Event-Condition | | | | |
| BASIC | POSITIVE | | Purpose-Act | | | |
| NON BASIC | POSITIVE | | Act-Purpose | | | |

References

- Alan Agresti (2018). *An Introduction to Categorical Data Analysis*. New York, Wiley & Sons.
- Karin Aijmer (2002). *English discourse particles. Evidence from a corpus*. Studies in Corpus Linguistics. Amsterdam, John Benjamins.
- Jean-Claude Anscombre and Oswald Ducrot (1983). *L'argumentation dans la langue*. Brussels, Editions Mardaga.
- Ignacio Arroyo (2017). *La expresión de causa en español*. Madrid, Visor Libros.
- Sarah Blackwell (2016). *Porque* in Spanish Oral Narratives: Semantic *Porque*, (Meta) Pragmatic or Both? *Perspectives in Pragmatics, Philosophy & Psychology*, 4: 615-651.
- Antonio Briz, Salvador Pons and José Portolés (2008). Diccionario de partículas discursivas del español. In *El diccionario como puente entre las lenguas y culturas del mundo. Actas del II Congreso Internacional de Lexicografía Hispánica*. Alicante, Biblioteca Virtual Cervantes: 217-227.
- Shuyuan Cao, Iría Da Cunha and Nuria Bel (2016). An analysis of the Concession relation based on the discourse marker *aunque* in Spanish Chinese parallel corpus. *Procesamiento del Lenguaje Natural*, 56: 81-88.

- Rocío Caravedo (2003). Principios del cambio lingüístico. Una contribución sincrónica a la lingüística histórica. *Revista de filología española* 83(1): 39-62.
- Benjamín Cárcamo (2019). Subjectivity in Spanish causal connectives: differentiating *porque*, *ya que* and *debido a que*. *Spanish in Context* 16(1): 51-76.
- Lynn Carlson and Daniel Marcu (2001). Discourse tagging reference manual. *ISI Technical Report ISI-TR-545*: 54-56.
- Manuel Cid (2002). Las conjunciones coordinantes del español actual desde el punto de vista funcional. *Boletín de Lingüística* 18: 49-70.
- Ludivine Crible and María José Cuenca (2017). Discourse markers in speech: characteristics and challenges for corpus annotation. *Dialogue & Discourse* 8(2): 149-166.
- Debopam Das (2014). *Signaling of coherence relations in discourse*. PhD Thesis, Simon Fraser University, Burnaby.
- 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 and Henk Pander Maat (2003). A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In A. Verhagen & J. van de Weijer (Eds.), *Usage Based Approaches to Dutch*: 75-199. Utrecht: LOT.
- María Domínguez García (2007). *Conectores discursivos en textos argumentativos breves*. Madrid, Arco/Libros.
- Eladio Duque (2014). Signaling causal coherence relations. *Discourse Studies* 16(1): 25-46.
- Eladio Duque (2016). *Las relaciones del discurso*. Madrid, Arco/Libros.
- Jacqueline Evers-Vermeul, Jet Hoek and Merel Scholman (2017). On temporality in discourse annotation: Theoretical and practical considerations. *Dialogue & Discourse* 8(2): 1-20.
- Carolina Figueras Solanilla (2000). Puntuación e interpretación de las expresiones causales en el texto escrito. In J. de Bustos Tovar et al (eds.), *Lengua, discurso, texto: I simposio internacional de análisis del discurso*: 281-295. Madrid, Visor Libros.
- Kerstin Fischer (Ed.). (2006). *Approaches to discourse particles*. Amsterdam, Elsevier.
- Bruce Fraser (1996). Pragmatic markers. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)* 6(2): 167-190.
- Bruce Fraser (1999). What are discourse markers? *Journal of Pragmatics* 31(7): 931-952.
- Bruce Fraser (2009). An account of discourse markers. *International Review of Pragmatics* 1(2): 293-320.
- Catalina Fuentes (1996). *Aproximación a la Estructura del Texto*. Málaga, Ágora.
- Catalina Fuentes (1987). *Enlaces extraoracionales*. Sevilla, Universida de Sevilla.
- Carmen Galán (1999). La subordinación causal y final. In V. Demonte and I. Bosque, *Gramática descriptiva la lengua española*: 3597-3642. Madrid, Espasa Calpe.
- Alberto García-Cervigón (2006). *El grupo del nombre en la analogía de la GRAE, 1771-1917*. Madrid, Editorial Complutense.
- Samuel Gili Gaya (1961). *Curso superior de sintaxis española*. Barcelona, Bibliografs/a.
- Patrick Goethals (2002). *Las conjunciones causales explicativas en castellano. Un estudio semiótico-lingüista*. Leeuven, Peeters.
- Paula Gozalo Gómez (2004). *La expresión de la causa en castellano*. Madrid, Universidad Autónoma de Madrid.
- Susanne Günthner (1993). Weil – Man kann es ja wissenschaftlich untersuchen: Diskurspragmatische Aspekte der Wortstellung in Weil-Sätzen. *Linguistische Berichte* 143: 37-55.
- Salvador Gutiérrez Ordóñez (2002). «Causales», Boletín de la Real Academia Española. In S. Gutiérrez Ordóñez (Ed.). *Forma y sentido en sintaxis*: 100-208. Madrid, Arco Libros.
- Francisco Javier Herrero (1999). Sobre la evolución de las oraciones y conjunciones adversativas. *Revista de Filología Española* 79(3): 292-328.

- Jet Hoek (2018). *Making sense of discourse: On discourse segmentation and the linguistic marking of coherence relations*. PhD thesis, Utrecht University. Utrecht: LOT.
- Jet Hoek, Jacqueline Evers-Vermeul and Ted Sanders (2019). Using the Cognitive Approach to Coherence Relations for Discourse Annotation. *Dialogue & Discourse* 10(2): 1-33.
- Eduard Hovy, Julia Lavid, Ema Maier, Vibhu Mittal and Cecile Paris (1992). Employing knowledge resources in a new text planner architecture. In *Aspects of automated natural language generation*, pages 57-72. Springer, Berlin, Heidelberg.
- Romualdo Ibáñez, Fernando Moncada and Benjamín Cárcamo (2019). Coherence Relations in Primary School Textbooks: Variation across School Subjects. *Discourse Processes*, 56(8), 764-785.
- Romualdo Ibáñez, Fernando Moncada and Andrea Santana (2015). Variación disciplinar en el discurso académico de la Biología y del Derecho: un estudio a partir de las relaciones de coherencia. *Onomázein* (32): 101-131.
- Silvia Iglesias Recuero (2000). La evolución del “pues” como marcador discursivo hasta el siglo XV. *Boletín de la Real Academia Española* 280(80): 209-308.
- Rudi, Keller (1995). The epistemic Weil. In D. Stein and S. Wright (eds.), *Subjectivity and Subjectivisation: Linguistic Perspectives*. 16-30. Cambridge: Cambridge University Press.
- Alistair Knott (1996). *A data-driven methodology for motivating a set of coherence relations*. PhD Thesis, University of Edinburgh.
- Alistair Knott and Robert Dale (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes* 18(1): 35-62.
- 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 Koch (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 33(2): 363-374.
- Fang Li, Jacqueline Evers-Vermeul and Ted Sanders (2013). Subjectivity and result marking in Mandarin. *Chinese Language & Discourse* 4(1): 74-119.
- Fang Li, Willem Mak, Jacqueline Evers-Vermeul and Ted Sanders (2017). On the online effects of subjectivity encoded in causal connectives. *Review of Cognitive Linguistics* 15(1): 34-57.
- William Mann and Sandra Thompson (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse* 8(3): 243-281.
- Manuel Martí Sánchez (2008). Los marcadores en español L/E: *conectores discursivos y operadores pragmáticos*. Madrid, Arco/Libros.
- Roser Martínez (1997). *Conectando texto: guía para el uso efectivo de elementos conectores en castellano*. Barcelona, Octaedro.
- Paul Meyer (2000). The relevance of causality. In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause, condition, concession and contrast: Cognitive and discourse perspectives*: 9-34. Berlin, Mouton de Gruyter.
- Thomas Meyer and Bonnie Webber (2013). Implication of discourse connectives in (machine) translation. In *Proceedings of the Workshop on Discourse in Machine Translation*, pages 19-26, Sofia.
- Estrella Montolio (2001). *Conectores de la lengua escrita. Contraargumentativos, consecutivos, aditivos y organizadores de la información*. Barcelona, Ariel.
- Leo Noordman and Femke De Blijzer (2000). On the processing of causal relations. In E. Couper-Kuhlen & B. Kortmann (Eds.), *Cause, condition, concession and contrast: Cognitive and discourse perspectives*: 35-56. Berlin, Mouton de Gruyter.

- 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 (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*: 57-82. Berlin, Mouton de Gruyter.
- Henk Pander Maat and Ted Sanders (2001). Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics* 12(3): 247-274.
- María Victoria Pavón Lucero (1999). Clases de Partículas. In V. Demonte and I. Bosque, *Gramática descriptiva de la lengua española*: 565-656. Madrid, Espasa Calpe.
- Mirna Pit (2003). *How to express yourself with causal connective: Subjectivity and causal connectives in Dutch, German and French*. New York, Rodopi.
- Mirna Pit, Jacqueline Hulst and Paander Mat (1996). Subjectiviteit en de Spaanse connectieven porque, ya que en puesto que. *Gramma/TTT* 5(3): 221-240.
- Livia Polanyi, Chris Culy, Martin Van Den Berg, Gian Lorenzo Thione and David Ahn (2004). A rule-based approach to discourse parsing. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL 2004*, pages 108-117. Cambridge, Massachusetts, USA
- Salvador Pons (1998). Conexión y conectores. Estudio de su relación en el registro informal de la lengua. *Cuadernos de Filología, Anexo XXVII*. Valencia, Universitat de València.
- José Portóles (1993). La distinción entre los conectores y otros marcadores del discurso en español. *Verba: Anuario galego de filoloxia* 20: 141-170.
- José Portóles (1998). *Marcadores del discurso*. Barcelona, Ariel.
- REAL ACADEMIA ESPAÑOLA: *Diccionario de la lengua española*, 23.^a ed., [on line version 23.3]. <<https://dle.rae.es>> [October, 2019].
- Gisela Redeker (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics* 14(3): 367-381.
- Gisela Redeker (1991). Linguistic markers of discourse structure. *Linguistics* 29(6): 1139-1172.
- Gisela Redeker and Helmut Gruber (2014). Introduction The pragmatics of discourse coherence. In H. Gruber and G. Redeker (eds.), *The pragmatics of discourse coherence: Theories and applications*: 1-22. Amsterdam, John Benjamins.
- William Salmon (1997). *Causality and explanations*. New York, Oxford University Press.
- 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.
- Ted Sanders (2005). Coherence, causality and cognitive complexity in discourse. In M. Aurnague & M. Bras (Eds.), *Proceedings of the First International Symposium on the Exploration and Modelling of Meaning*, pages 31-46. Université de Toulouse-le-Mirail, France.
- Ted Sanders, Wilbert Spooren and Leo Noordman. (1992). Toward a taxonomy of coherence relations. *Discourse Processes* 15(1): 1-35.
- 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*. (20).
- Andrea Santana, Wilbert Spooren, Dorien Nieuwenhuijsen and Ted Sanders (2018). Subjectivity in Spanish discourse: Explicit and implicit causal relations in different text types. *Dialogue & Discourse* 9(1): 163-191.
- Deborah Schiffrin (1987). Discourse markers. *Studies in Interactional Sociolinguistics*. Cambridge: Cambridge University Press.
- 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 and Ted Sanders (2008). The acquisition order of coherence relations: On cognitive complexity in discourse. *Journal of Pragmatics* 40(12): 2003–2026.
- Manfred Stede (2014). Resolving connective ambiguity: A prerequisite for discourse parsing. In H. Gruber and G. Redeker (eds.), *The Pragmatics of Discourse Coherence*: 121–141. Amsterdam, John Benjamins.
- Ninke Stukker (2005). *Causality marking across levels of language structure. A cognitive semantic analysis of causal connectives in Dutch*. PhD Thesis. Utrecht University.
- 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 (2009). Categories of subjectivity in Dutch causal connectives: a usage-based analysis. In T. Sanders and E. Sweetser (eds.), *Causal categories in discourse and cognition*: 119–171. Berlin, De Gruyter Mouton.
- Eve Sweetser (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge University Press.
- 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 Jan Renkema (ed.), *Discourse of Course*: 127–140. Amsterdam, John Benjamins.
- Maite Taboada and Debopam Das (2013). Annotation upon Annotation: Adding Signalling Information to a Corpus of Discourse Relations. *Dialogue & Discourse* 4(2): 249–281.
- Maite Taboada and María Gómez-González (2012). Discourse markers and coherence relations: Comparison across markers, languages and modalities. *Linguistics and the Human Sciences* 6(1–3): 17–41.
- Nynke Van der Vliet and Gisela Redeker (2014). Explicit and implicit coherence relations in Dutch texts. In H. Gruber and G. Redeker (eds.), *The Pragmatics of Discourse Coherence*: 23–52. Amsterdam, John Benjamins.
- Yipu Wei (2018). *Causal connectives and perspective markers in Chinese: The encoding and processing of subjectivity in discourse*. PhD Thesis. Utrecht University.
- Karen Zagana (2002). *The syntax of Spanish*. Cambridge University Press.
- Martín Zorraquino and Antonia Portolés (1999). Los Marcadores del Discurso. In V. Demonte and I. Bosque, *Gramática descriptiva la lengua española*: 4051–4213. Madrid, Espasa Calpe.
- Sandrine Zufferey (2010). *Lexical pragmatics and theory of mind: the acquisition of connectives*. Amsterdam, John Benjamins.