



Discours

Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics

20 | 2017
Varia

Causality and Subjectivity in Spanish Connectives: Exploring the Use of Automatic Subjectivity Analyses in Various Text Types

Andrea Santana, Dorien Nieuwenhuijsen, Wilbert Spooren and Ted Sanders



Electronic version

URL: <http://journals.openedition.org/discours/9307>
DOI: 10.4000/discours.9307
ISSN: 1963-1723

Publisher:

Laboratoire LATTICE, Presses universitaires de Caen

Electronic reference

Andrea Santana, Dorien Nieuwenhuijsen, Wilbert Spooren and Ted Sanders, « Causality and Subjectivity in Spanish Connectives: Exploring the Use of Automatic Subjectivity Analyses in Various Text Types », *Discours* [Online], 20 | 2017, Online since 22 September 2017, connection on 19 April 2019. URL : <http://journals.openedition.org/discours/9307> ; DOI : 10.4000/discours.9307

Licence CC BY-NC-ND



Revue de linguistique, psycholinguistique et informatique

<http://discours.revues.org/>

Causality and Subjectivity in Spanish Connectives: Exploring the Use of Automatic Subjectivity Analyses in Various Text Types

Andrea Santana

Utrecht Institute of Linguistics (UiL OTS)

Dorien Nieuwenhuijsen

Utrecht Institute of Linguistics (UiL OTS)

Wilbert Spooren

Radboud Universiteit

Ted Sanders

Utrecht Institute of Linguistics (UiL OTS)

.....
Andrea Santana, Dorien Nieuwenhuijsen, Wilbert Spooren, Ted Sanders, «Causality and Subjectivity in Spanish Connectives: Exploring the Use of Automatic Subjectivity Analyses in Various Text Types», *Discours* [En ligne], 20 | 2017, mis en ligne le 22 septembre 2017.

.....
URL : <http://discours.revues.org/9307>

.....
Titre du numéro : *Varia*

Coordination : Shirley Carter-Thomas & Laure Sarda

revues.org
CENTRE POUR L'ÉDITION ÉLECTRONIQUE OUVERTE
CENTRE FOR OPEN ELECTRONIC PUBLISHING

 discours

 Presses
universitaires
de Caen

Causality and Subjectivity in Spanish Connectives: Exploring the Use of Automatic Subjectivity Analyses in Various Text Types

Andrea Santana

Utrecht Institute of Linguistics (UiL OTS)

Dorien Nieuwenhuijsen

Utrecht Institute of Linguistics (UiL OTS)

Wilbert Spooren

Radboud Universiteit

Ted Sanders

Utrecht Institute of Linguistics (UiL OTS)

.....

Causality and subjectivity are relevant cognitive principles in the categorization of coherence relations and connectives. Studies in various languages have shown how both notions can explain the meaning and use of different connectives. However, the Spanish language has been understudied from this perspective. Also, most of the existing research on connectives has used manual analyses. This paper explores the use of automatic analyses of subjectivity in causal connectives. The goal is to determine the degree to which Spanish causal connectives encode subjectivity across different text types, by carrying out automatic analyses. A corpus was constructed to identify causal connectives in journalistic texts (news and editorials) and academic texts (essays, research articles and textbooks in education and psychology). A Spanish lexicon of subjectivity was used to automatically identify the frequency of subjective words in the texts and the segments linked by the most frequent causal connectives. Our assumption is that supposedly subjective connectives will occur in more subjective environments, that is, a context containing relatively many subjective words. The results show a statistically significant relationship between the use of connectives and text type, and also between the text type and subjectivity. From a methodological point of view, the use of automatic analyses appears not to be without difficulties. However, it allowed us to explore various text types, to analyze the degrees of subjectivity in each of them and to identify tendencies related to the use of connectives in Spanish. More interestingly, the combination of automatic and manual analyses can result in a promising methodology for the study of discourse coherence and connectives.

Keywords: causality, subjectivity, connectives, Spanish, automatic analyses, lexicon, text type

1. Introduction

- 1 Constructing a coherent representation by relating different discourse segments is crucial in communication (see Graesser et al., 2003; Zwaan & Rapp, 2006; and many others). Coherence relations play a fundamental role in this representation since they establish the link between two or more discourse segments. These relations

are cognitive entities that allow us to infer the meaning of two or more segments (Sanders et al., 1992 and 1993). Thus, on reading Example [1], we understand that the coherence relation between the first and the second segment (S₁ and S₂) corresponds to a *cause-consequence* relation and that the meaning of this passage can only be established by inferring this relation between the segments.

- [1] (S₁) The sun came up. As a result, (S₂) the temperature went up.
(Pander Maat & Degand, 2001: 212)

- 2 A general assumption is that coherence relations express different meanings, such as addition, contrast or causality. This study focuses on causal coherence relations, for which an implicational relation can be deduced between the connected discursive segments ($P \rightarrow Q$) (Sanders et al., 1992 and 1993). For example, [1] is a causal coherence relation because the fact described in S₁ constitutes the antecedent (P) of the fact presented in S₂ which is the consequent (Q). The two discursive segments are strongly connected, one event implying the other one.
- 3 Over the last two decades, the notion of subjectivity has been used to distinguish between various types of causal coherence relations. It has been regarded as a cognitive principle that allows us to explain the system and use of causal relations and their linguistic expressions (Sanders & Spooren, 2015). Subjectivity refers to the speaker's involvement, that is the degree to which the speaker is involved in the construal of the relation (Degand & Pander Maat, 2003; Pander Maat & Degand, 2001; Pander Maat & Sanders, 2000 and 2001). Thus, a relation will be subjective when a *subject of consciousness* (SoC) is involved in the construction of the relation. This SoC may be represented by a reasoning entity that states a conclusion or deduction, by a speaker who performs a speech act or by a thinking subject who is third person actor. In contrast, a relation is objective when the SoC is absent, and the discursive segments are related because some facts or states of affairs cause another event in the outside world. This subjectivity approach is compatible with Sweetser's trichotomy (Sweetser, 1990). *Speech act relations* are considered subjective since segments are connected by an (often implicit) SoC: the speaker who performs the speech act. *Epistemic relations* are subjective because there is a SoC who is reasoning, deducing or concluding something on the basis of observation. Finally, *content relations* are objective relations because there is no speaker involvement and the connected segments describe states of affairs that occur in the outside world (Pander Maat & Degand, 2001; Pander Maat & Sanders, 2000; Sanders et al., 2012).
- 4 Connectives and other linguistic markers seem to vary systematically in the way they encode causality and subjectivity. For example, *As a result* in [2] below is a cue phrase that prototypically signals an objective causal relation; it indicates that S₂ is the consequence of the fact described in S₁, that two states of affairs that occur in the outside world are connected, and that there is no SoC. The same marker could not be used in [3] since the idea of S₂ is not a real consequence of the fact

described in Sr. The connective *So* in Example [3] fits in with the subjective causal relation; it is often used to indicate conclusions, in this case, it marks a conclusion based on the observation that the lights are out.

- [2] (S1) There had been an avalanche at Roger's pass. As a result (S2) the road was blocked.

(Pander Maat & Sanders, 2001: 252)

- [3] (S1) The lights in the neighbors' living room are out. So (S2) they are not at home.

(Pander Maat & Sanders, 2001: 249)

5 Corpus research has been conducted in various languages to investigate to what extent such linguistic markers indeed specialize in subtypes of causal relations, as they do not seem to be interchangeable in terms of use and meaning. These systematic variations of linguistic markers are not of a black and white type (Sanders & Spooren, 2013 and 2015; Stukker & Sanders, 2012). However, there is evidence in different languages showing that some connectives preferentially express objective meanings, while others specialize in expressing subjective meanings. Most of these studies have focused on Dutch (Pander Maat & Degand, 2001; Pander Maat & Sanders, 2001; Pit, 2006; Verhagen, 2005; Sanders & Spooren, 2015; Stukker & Sanders, 2009); there are also studies on German (Günthner, 1993; Keller, 1995; Pit, 2007; Stukker & Sanders, 2012; Wegener, 2000), and other typologically less related languages have also been explored, such as Mandarin-Chinese (Li et al., 2013) and French (Degand & Pander Maat, 2003; Zufferey, 2012). However, other Romance languages, such as Spanish, are understudied from this perspective. There are some studies that have analyzed the notion of perspective or subjectivity of Spanish connectives. They have focused on some particular connectives such as *porque* ("because"), *ya que* ("since"), *pues* ("for"), *como* ("as/since") and on specific contexts of use (Blackwell, 2016; Goethals, 2002 and 2010), or they have been based on the translation of connectives from Dutch to Spanish (*omdat* and *want* corresponding to *porque* ("because"), *puesto que* ("given that") and *ya que* ("since") in Spanish) (Pit et al., 1996). Therefore, it is still not known whether Spanish connectives encode subjectivity as other languages do and if so, it remains unclear in which contexts they are used.

6 There is an extensive body of literature on Spanish linguistic markers of coherence relations (e.g., Briz, 1998; Martín Zorraquino & Montolío, 1998; Martín Zorraquino & Portolés, 1999; Montolío, 2001; Pons, 1998 and 2000; Portolés, 1998). In fact, the interest in these markers is not a recent topic in studies of the Spanish language. Casado Velarde (1991) refers to some studies of the past centuries (Bello, 1847; Garcés, 1791; Fernández Ramírez, 1951; Gili Gaya, 1961; Moliner, 1966) that already showed interest in the function of these elements in discourse organization, and provided descriptions and definitions of some conjunctions. More recently, several studies have aimed at constructing dictionaries of Spanish linguistic markers (Aschenberg & Loureda, 2011; Briz et al., 2008; Fuentes, 2009; Martín Zorraquino, 2003;

Santos Río, 2003; Vázquez Veiga, 2002), which have contributed to the field of linguistics, translation and the teaching of Spanish as a foreign language. Other studies, with the purpose of exploring the signaling of coherence relations in Spanish, have provided information about causal and concessive relations in Spanish (Duque, 2014 and 2016; Taboada & Gómez-González, 2010). There are also resources available that allow us to analyze Spanish coherence relations, specifically the “RST Spanish Treebank”¹, which is the first corpus to be annotated with rhetorical relations in Spanish (Cunha et al., 2011). Nonetheless, most of these contributions are mainly descriptive, are based on specific theoretical approaches and do not focus on the possible systematic differences in subjectivity.

7 Therefore, the present study aims to explore just that: whether Spanish causal connectives encode systematic differences in subjectivity. However, this is not the only goal pursued in this paper. In addition, this paper provides insights into automatic analyses and explores a method that could improve the analyses of coherence relations and their linguistic markers, especially in contexts that have received little attention. More specifically, this paper analyzes to what extent automatic subjectivity analyses allow us to gain insight into the meaning and use of Spanish causal connectives.

8 It is widely acknowledged that corpus-based studies are useful in linguistic analyses since they provide us with large data sets, and they capture the variability of the language in different contexts of use. However, the vast majority of studies of discourse relations and connectives are based on small corpora or random sets extracted from large corpora. The reason is that most of the analyses are hand-based (Bestgen et al., 2006). Moreover, it has been argued that the exploitation of corpora by using automatic tools (such as word lists, concordances, keywords, among others) is beneficial (it allows us to search corpora rapidly and reliably [McEnery & Hardie, 2012]), but such tools are hard to use for the study of coherence relations and their linguistic markers because of the complexity of the analyses involved.

9 The annotation and segmentation of coherence relations and their linguistic markers indeed represent a challenge. The proliferation of studies focusing on the difficulties related to annotation and segmentation procedures (Bayerl & Paul, 2011; Cartoni et al., 2013; Hoek et al., 2017; Spooren & Degand, 2010, among others) is not a coincidence. In fact, different theoretical approaches to the analysis of coherence relations such as the “Penn Discourse Treebank” (PDTB) (Prasad et al., 2008), the “Rhetorical Structure Theory Treebank” (Carlson et al., 2003) and “Segmented Discourse Representation Theory” (Asher & Lascarides, 2003) have provided extensive resources that explain the different proposed sets of coherence relations and describe the processes of annotation and segmentation in detail (see the manuals developed by the PDTB Research Group [2007] and Carlson & Marcu [2001]). Unquestionably, these processes involve a series of complex tasks, among which:

1. Available in: http://www.iling.unam.mx/rst/index_es.html.

defining criteria for the analysis of coherence relations and connectives; establishing analytical criteria for the units that are connected; discussing ambiguous examples; making decisions about several interpretations. It does not come as a surprise, therefore, that the field of text linguistics and discourse studies is dominated by manual analyses. This type of analysis is time-consuming, and the results obtained may depend on individual analysts, which makes generalizations and replications difficult (Bestgen et al., 2006).

- 10 Lately, researchers have argued in favor of automatic analyses as they have been developed in language technology and corpus-based data analysis (Bestgen et al., 2006; Bouma et al., 2001; Oostdijk et al., 2013; Pander Maat et al., 2014; Spooren & Degand, 2010; Spooren et al., 2010, among others). Although such automatic analyses cannot entirely replace manual ones, it has been suggested that they could be a complementary type of analysis in the study of coherence relations and connectives, too (Bestgen et al., 2006; Spooren & Degand, 2010). Recently, Levshina and Degand (2017) have shown that it is possible to disambiguate between objective and subjective uses of the connective *because* by presenting an integrative method that incorporates the use of a parallel corpus and semi-automatic analyses.
- 11 A final issue we want to study here is the role of the text type². The question is whether a systematic relationship exists between the text type and the type of coherence relation. We believe that the text type is an influential factor on the type of coherence relations occurring in texts. In fact, some studies support this assumption. For instance, Sanders et al. (1993), by validating the primitives proposed in their taxonomy, demonstrated that language users could recognize objective and subjective relations³ when appropriate communicative contexts were provided, i.e. when coherence relations were presented in descriptive and argumentative texts. Focusing on the interaction between the context and the primitive source of coherence, Sanders (1997) demonstrated that the interpretation of ambiguous cases of coherence relations was strongly influenced by the context. In an experiment, ambiguous relations that were in-between an objective and a subjective reading (*chameleon* cases) were interpreted by text analysts as objective relations when presented in descriptive texts and as subjective relations when appearing in argumentative texts. The same study also revealed that objective relations were predominant in informative texts, while subjective relations were predominant in expressive and persuasive texts.
- 12 Zufferey (2012) demonstrated through corpus analyses that the distribution of French connectives varies according to different modalities of texts. In written texts, the connective *car* (“because”) is more often used for epistemic relations, whereas *parce que* (“because”) prefers content relations. In spoken discourse, the

2. The term “text type” adopted in this paper refers mainly to the classical distinction between informative and persuasive/argumentative texts.

3. These are named “semantic” and “pragmatic” relations in the original papers (see Sanders et al., 1992 and 1993).

connective *car* is absent and *parce que* is used significantly more often in speech act and epistemic relations. Zufferey found that the semantic-pragmatic profile of the connective *puisque* (“since”), which was clearly subjective, was stable across text types. Stukker and Sanders (2010) also revealed interesting results related to the role that the context plays in the distribution of connectives.

13 All the findings summarized above seem to indicate a relationship between text type on the one hand, and the type of connective or relation expressed, on the other hand. The question is, however, what exactly this relationship is. Sanders and Spooren (2015) analyzed the Dutch connectives *omdat* (“because”) and *want* (“since/for”) in written texts, conversations and chat interactions. The study revealed significant interactions between medium and different indicators of subjectivity such as *propositional attitude*, *SoC*, and *linguistic realization of the SoC*, but no straightforward relationship between medium and *type of relation*: the two connectives showed a clearly different pattern irrespective of the media used in the research. In other words, the connectives show a robust semantic-pragmatic profile in terms of subjectivity, irrespective of the text type in which they occur. These results lead to the conclusion that there is a relationship between text type and type of connective, but that context does not determine the type of relation (see also Stukker & Sanders, 2012). In this paper, we want to further investigate this relationship, and put this conclusion to the test.

14 We chose text types in two types of discourse for the analysis of Spanish causal connectives: journalistic discourse and academic discourse. The former focuses on the real world, paying great attention to referentiality, factuality, accountability and reliability (Waugh, 1995); it is a public discourse, oriented to a broad audience (Van Dijk, 1988; Waugh, 1995). By contrast, academic discourse is a specialized discourse, oriented to a specific audience; it refers to the different ways of thinking and using language which exist in academia (Bhatia, 2002 and 2004; Hyland, 2009; Swales, 1990; Silver, 2006).

15 Among journalistic texts, news and editorials were selected. The news is inherently an informative text type; its purpose is to convey current, unknown, and unprecedented information and it is supposed to be about topics which are of interest to the public at large (Leñero & Marín, 1986; Waugh, 1995). Editorials are inherently an argumentative text type since a difference of opinion is resolved through arguments (Van Eemeren et al., 1997); it reveals the author’s ideological convictions, political positions and individual viewpoints (Broersma, 2010; Leñero & Marín, 1986; Van Dijk, 1993; White, 2006).

16 Among academic texts, essays, textbooks, and research articles were selected. The communicative purpose of the essay is to persuade the reader of the correctness of a central statement (Hyland, 1990); it presents and develops a proposition, by arranging several arguments to defend or explain a point of view (Hale et al., 1996; Henry & Roseberry, 1999; Hyland, 2009). The textbook is an introductory text in undergraduate fields which arranges the accepted knowledge of a discipline into

a coherent whole (Myers, 1992; Swales, 1995); it is written by experts and it is oriented to students that need to learn and acquire the fundamental concepts of their discipline (Alred & Thelen, 1993). Finally, the research article is a central text in the communication and dissemination of research results in a disciplinary field, being the principal vehicle for new knowledge (Hyland, 1996). Through the research article, writers provide their disciplines with knowledge and they demonstrate the novelty, credibility and relevance of their work (Hyland, 2001 and 2011).

17 Given the characteristics described so far, we expect to find different degrees of subjectivity across the selected texts, that is, that they contain more or fewer subjective words according to their characteristics. Our assumption is that if a connective has a subjective profile, it will occur more often in a subjective environment, that is, a context containing relatively many subjective words. In the case of journalistic texts, many subjective words are expected in editorials because of its argumentative nature, while fewer subjective words are expected in news due to its informative essence. Consequently, subjective connectives are expected to co-occur more with subjective words in editorials than those that occur in news.

18 In the case of academic texts, we predict a continuum in terms of subjectivity. On the one hand, we expect to find the essay with many subjective words because of its persuasive nature. Therefore, subjective connectives in essays will co-occur more with subjective words. On the other hand, it is expected that there will be fewer subjective words in the textbook because it aims to teach specialized disciplinary knowledge rather than to convince its audience with arguments. It can be considered as an informative text type rather than an argumentative text type. Therefore, connectives in textbooks will co-occur less with subjective words. Finally, in the middle of this continuum, we expect to find the research article, with fewer subjective words than the essay, but with more subjective words than the textbook. This claim is based on the assumption that the research article has a hybrid nature depending on its different sections. For example, the literature review section can be descriptive when presenting the theoretical background, but it can also be persuasive because it intends to justify the value of the current research and to show its newsworthiness in comparison with previous studies (Kwan, 2006). Therefore, objective as well subjective words could be expected. Likewise, other sections such as the introduction, in which the research is justified, and the discussion, in which the research is presented as a significant contribution to the field, may be more persuasive and contain more subjective words than other sections such as method or results, in which procedures and findings are described in a concise and organized way. Hence, the research article is predicted to be hybrid in terms of subjective words, and consequently, this nature would also be evident in its connectives, which are expected to co-occur both with subjective words and non-subjective words.

19 Now that the central issues have been introduced, it is time to state our hypotheses. Below, each hypothesis is presented with a brief explanation.

- Regarding causality, text type and identification of connectives:

(H1) *The distribution and the frequency of connectives vary in different text types.*

This hypothesis is based on the assumption that the text type constitutes a fundamental variable that constrains the use and distribution of coherence relations and their causal connectives.

- Concerning the degrees of subjectivity and the influence of the text types:

(H2) *The number of subjective words depends on text types:*

(a) *Journalistic discourse: subjective words are more frequent in editorials than in news.*

(b) *Academic discourse: subjective words are more frequent in essays than in research articles, and more frequent in research articles than in textbooks.*

This hypothesis is based on the idea that the argumentative nature of texts will be reflected in the number of subjective words, i.e. if the text is argumentative, it will contain more subjective words. Thus, editorials and essays will contain many subjective words since these are inherently argumentative texts; news and textbooks will contain fewer subjective words because these are mainly informative texts, and research articles will contain fewer subjective words than argumentative texts but more than informative texts because they have a hybrid nature.

- About the subjectivity of the most frequent causal connectives in journalistic and academic texts:

(H3) *Subjective connectives (connectives that occur frequently in subjective text types) co-occur more often with subjective words than non-subjective connectives.*

This hypothesis is based on our assumption that if a connective signals subjectivity, it will be used in a subjective environment, i.e. a context with many subjective words.

20 So far, we have presented the motivations of our study and the central concepts involved in it. The central goal of this study has also been established: to determine the degree to which Spanish causal connectives encode subjectivity across different text types, by carrying out automatic analyses. The following sections explain the method that was used to perform the proposed analyses.

2. Method

21 Different methodological steps were carried out: 2.1) Construction of a Spanish corpus; 2.2) Counting of causal connectives; 2.3) Manual analysis and selection of frequent causal connectives; 2.4) Automatic counting of subjective words per text; 2.5) Automatic counting of subjective words per segment; and 2.6) Data analysis and processing. In this section, the description of these methodological steps is provided in detail.

2.1. Construction of a Spanish corpus

22 Different sources were selected depending on the text type. In the case of journalistic texts, the source was the newspaper *El País*⁴ and each sample was selected randomly by using a sequence generator⁵. Approximately 90,000 words per text type were selected, resulting in 344 texts, evenly distributed between news and editorials. The detailed composition of the journalistic sub-corpora is presented in Table 1.

23 In the case of academic texts, it was necessary to determine the domain in which the texts would be selected. The exploration of scientific databases⁶ revealed that social sciences is the area of knowledge with the largest number of Spanish scientific publications, specifically in education and psychology. Consequently, texts were selected in these two specific domains. It should be mentioned that it was necessary to select different sections from the texts (initial, middle and final) since their length varied considerably.

24 Concerning the sources, different criteria were considered. The criteria for textbooks were the availability in digital format and the original language. Therefore, texts written originally by Spanish authors were selected from the bibliographies of different academic programs of education and psychology at Spanish universities (Universidad Complutense de Madrid and Universidad Autónoma de Madrid). The criteria for research articles were the availability in digital format and the scientific quality of the publications. Thus, after reviewing different indicators of quality⁷, the most highly-ranked journals in education and psychology were selected, *Revista de Educación* and *Anales de Psicología*, respectively. Finally, the collection of essays involved the same criteria as for research articles and a third criterion was added: the explicitness of a communicative purpose (Swales, 1990). In the journal *Revista de Educación*, it was possible to collect essays because a special section is devoted to research articles and another one to essays. However, this was not the case in the journal *Anales de Psicología*, in which all the contributions are labelled as *research articles*. Therefore, it was necessary to use the third criterion, the explicitness of the communicative purpose of the essay, i.e. to persuade the reader of the correctness of a central statement (Hyland, 1990). Consequently, other journals were explored and *The UB Journal of Psychology* was selected since it met all the criteria: it made distinctions among its publications, and some of the

4. Available in: <http://elpais.com/>.

5. Available in: <http://www.random.org/>.

6. "Dialnet", "Sistema de Información Científica Redalyc – Red de Revistas Científicas de América Latina y el Caribe, España y Portugal", and "Latindex: Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal".

7. "Clasificación Integrada de Revistas Científicas" (CIRC), "Revistes Científiques de Ciències Socials i Humanitats" (CARHUSplus), "Revistas Españolas de Ciencias Sociales y Humanidades" (RESH), "Scimago Journal & Country Rank" (SJR) and "Journal Citation Reports" (JCR).

| | News | Editorial | Total |
|-------|--------|-----------|---------|
| Texts | 172 | 172 | 344 |
| Words | 88,014 | 87,452 | 175,466 |

Table 1 – Composition of the journalistic sub-corpora in terms of texts and words per text type

| | | Essay | Research article | Textbook | Total |
|----------------|-------|--------|------------------|----------|---------|
| Education | Texts | 20 | 20 | 20 | 60 |
| | Words | 45,409 | 45,668 | 45,510 | 136,587 |
| Psychology | Texts | 20 | 20 | 20 | 60 |
| | Words | 44,853 | 46,829 | 45,090 | 136,772 |
| Total of texts | | 40 | 40 | 40 | 120 |
| Total of words | | 90,262 | 92,497 | 90,600 | 273,359 |

Table 2 – Composition of the academic sub-corpora in terms of texts and words per text type in each domain

requirements for one type of publication were the explicit indication of a thesis or an academic or theoretical orientation, which fits with the communicative purpose of the essay. The detailed composition of the academic sub-corpora is presented in Table 2.

- 25 As shown in Tables 1 and 2, the total number of texts differs in the two sub-corpora, but the main criterion of selection was the number of words per each text (news, editorials, essays, research articles and textbooks), which is quite similar (around 90,000).

2.2. Counting of causal connectives

- 26 In order to identify the causal connectives, an extensive and varied list of connectives was built after a bibliographical review of different proposals of Spanish causal connectives (Domínguez García, 2007; Martí, 2008; Martínez, 1997; Montolio, 2001; Portolés, 1998). The selection consisted of 57 Spanish causal connectives, in which cue phrases were also assumed to be connectives (see Appendix).

- 27 The selected causal connectives were used as keywords. The identification of them in the corpus was carried out automatically using the Antconc⁸ toolkit

8. Available in: <http://www.laurenceanthony.net/software.html>.

and a particular R-script. The results are presented in Section 3.1 for journalistic texts and 3.2 for academic texts.

2.3. Manual analysis and selection of frequent causal connectives

28 Once the causal connectives had been identified automatically in the corpus, it was possible to identify the most frequent ones. In order to do this, it was necessary to check whether every case corresponded to a causal coherence relation marked by the specific linguistic markers of our list. This procedure is what we called *manual analysis* since the recognition of coherence relations had to be made by the analyst and could not be done automatically. This analysis implies exploring every case, discussing the problematic examples with peers, and deciding in each case whether it should be included in the analysis or not.

29 The manually selected relations were established between two adjacent clauses, in which there was a clearly identifiable conjugated verb. In Example [4], S₁ and S₂ are adjacent clauses with conjugated verbs (*tiende* [“tends to”] and *resta* [“diminishes”], respectively) and are linked by the connective *porque* (“because”).

[4] (S₁) La situación tiende a empeorar políticamente, porque (S₂) la desaceleración económica en Europa resta capacidad de crecimiento a Estados Unidos y Japón.

‘(S₁) The situation tends to become worse politically because (S₂) the economic slowing down in Europe diminishes the growth capacity of the United States and Japan.’

30 The manual analysis led to the exclusion of 634 examples, 234 in journalistic texts and 400 in academic texts. The main reason for this was the fact that coherence relations could not always be established between two adjacent clauses, which was the basic criterion for the selection. In some cases, S₁ corresponded to a clause, but S₂ corresponded to a phrase or vice versa. A similar situation consisted of connectives appearing between brackets. These cases were also excluded from the analyses. In Example [5], S₂ is a phrase instead of a main clause and in [6], S₂ is not a main clause and the connective appears between brackets:

[5] La conferencia anual del Círculo de Economía ha marcado una pauta al propugnar el retorno al diálogo, una tercera vía que podría desembocar en una reforma constitucional y, por tanto, una consulta legal.

‘The annual conference of the Circle of Economy has set a standard by advocating the return to dialogue, a third way that could lead to a constitutional reform and, therefore, a legal query.’

[6] El AMPK también actúa cuando detecta una bajada de glucosa (el nutriente básico de las células), y activa un sistema de ahorro que reduce las divisiones celulares (y, por tanto, el envejecimiento).

‘The AMPK also acts when it detects a drop in glucose (the basic nutrient of cells) and it activates a saving system which reduces cell divisions (and, therefore, aging).’

31 The same decision was made for those cases in which the linguistic marker functioned as a modal, comparative, temporal adverb or as an interjection instead of establishing an explicit causal coherence relation. In [7] the connective *así* (“thus”) has the role of modal adverb, and in [8] *pues* (“so”) functions as an interjection that could emphasize the meaning about what is communicated in the first segment, but neither of them establishes an explicit causal coherence relation.

[7] Sin duda Rusia tiene un papel fundamental que jugar en la resolución de cada vez más enrevesado tablero de Oriente Próximo y así debe ser reconocido.

‘Undoubtedly Russia has a fundamental role to play in the resolution of the increasingly confusing scene in the Middle East, and it must be recognized like that.’

[8] Una modificación de la política económica puede obtenerse por dos vías políticas. O el gobernante acaba rectificando por sí solo (opción bastante ardua y azarosa) o cambia su socio de coalición (pues parece que, en todo caso, deberá haber coalición).

‘A change in economic policy can be obtained in two political ways. Either the leader himself makes the change (a rather arduous and risky option), or he changes his coalition partner (well, it seems that, in any case, there must be a coalition).’

32 Another issue was the juxtaposition of linguistic markers of a different nature. In some cases, a causal connective appeared next to another additive, negative or conditional linguistic marker, which caused ambiguity. For example, the relation in [9] could be considered as an additive relation since the connective *y* (“and”) indicates that the two segments are connected, but at the same time, it could be considered as a causal relation because the second segment could be a consequence of the first one.

[9] Ojalá pudiera decirse que la deuda aumenta porque ha aumentado la inversión pública, sobre todo la más productiva, y por tanto hay que financiarla.

‘Hopefully, it could be said that the debt increases because public investment has increased, especially the most productive one, and therefore it has to be funded.’

33 These cases were not included as they can have several interpretations. The purpose was to identify purely causal linguistic markers. Likewise, those cases in which a list of causes was presented were excluded from the analysis although they were linked by a causal connective. The reason for this was that the relation between one segment and another was found to be additive instead of causal. Example [10] illustrates this type of case:

[10] Esta estrategia de moderación universal propicia alguna mejora competitiva en la economía dominante (en este caso Alemania), bien porque sus exportaciones aumenten, bien porque absorba los flujos de capitales procedentes de países cuya deuda sufre una presión mayor.

‘This universal restraint strategy promotes some competitive improvement in the dominant economy (in this case Germany) either because its exports increase, or because it absorbs the capital flows from countries whose debt is under greater pressure.’

34 In order to identify the most frequent causal connectives in the corpus, a threshold of frequency for the selection of the causal relations was adopted. In the case of journalistic texts, we considered those connectives as frequent that occurred five times or more in either news or editorials, whereas for academic texts the threshold was put at five per domain (education, psychology). This decision was made to avoid the influence of the sampling distribution in the subsequent chi-square analyses (Field et al., 2012). The total number of the most frequent causal connectives in the corpus and the most relevant cases of the manual analysis are described in Section 3.

2.4. Automatic counting of subjective words per text

35 Once the most frequent causal connectives had been identified, the analysis focused on subjectivity. The purpose was to explore degrees of subjectivity in the global context, i.e. to identify whether texts contain more or fewer subjective words. The identification of subjective words in each text was carried out by using a Spanish lexicon of subjectivity (Molina-González et al., 2013).

36 This lexicon was originally built for *sentiment analysis* (SA), which focuses on the automatic identification of personal states in natural languages (such as opinions, emotions, beliefs, etc.). Thus, texts are analyzed as objective or subjective texts. Particularly, they are analyzed according to their polarity, i.e. texts are labeled as either positive, negative or neutral (Pérez-Rosas et al., 2012). Several SA-studies have constructed lexicons of subjectivity and polarity (Díaz Rangel et al., 2014; Jiménez Zafra et al., 2014; Pérez-Rosas et al., 2012; Sidorov et al., 2013; Villena Román et al., 2014). One of them is the lexicon used in our study (Molina-González et al., 2013).

37 This resource was chosen for two reasons. The first one is the quality of this lexicon; it constitutes an enriched version of the improved “Spanish Opinion Lexicon” (iSOL) created by Molina-González et al. (2013)⁹. It is the result of an empirical analysis of data on subjectivity in the Spanish language, which serves as our initial point of reference. This lexicon is often used in the Spanish SA research community (Jiménez Zafra et al., 2014) and has already been validated in other research areas. The second reason is the accessibility of this lexicon; it is a freely available resource that provides us with a set of 8,171 subjective words in different word classes: verbs, nouns, adverbs, and adjectives. Our aim was to establish whether this instrument is valid for the identification of degrees of subjectivity across texts from different types of discourse, journalistic as well as academic.

38 The analysis consisted in identifying the subjective words of the Spanish lexicon of subjectivity in every text of the corpus. This identification was carried out automatically using a particular R-script. Additional analyses were applied to determine the statistical differences on the basis of chance (chi-square). The results are presented in Section 3.1 for journalistic texts and 3.2 for academic texts.

9. Available in: <http://sinai.ujaen.es/?p=1188>.

2.5. Automatic counting of subjective words per segment

39 The preceding methodological step (2.4) allowed us to explore the global environment of the most frequent causal connectives identified in the corpus. The following step was to analyze subjectivity in the local context, i.e. to identify the subjective words of every segment linked by the most frequent causal connectives (S₁ and S₂). The purpose was to identify whether a specific linguistic marker co-occurs more or less often with subjective words. We assume that if a linguistic marker has a subjective profile, it will occur in a context containing relatively many subjective words. The identification of subjective words in every segment was carried out by using the same lexicon and the same automatic analyses employed in the previous methodological step (see 2.4). This enabled us to establish whether the use of the lexicon is a valid instrument for the identification of degrees of subjectivity not only in a global context such as texts (see 2.4) but also in a local context such as linked segments. The results are presented in Section 3.3.

2.6. Data analysis and processing

40 When the corpus was collected, five methodological steps were carried out. The first step consisted of the automatic identification of Spanish connectives in both sub-corpora, journalistic and academic. Statistical analyses were carried out using a log likelihood test for the journalistic sub-corpora and a log-linear analysis for the academic sub-corpora. The second step was the manual analysis of these connectives, which led to the identification of the most frequent causal connectives in the corpus and the determination and exclusion of inappropriate cases. The third step resulted in an overview of the number of subjective words for each of the sub-corpora. These were analyzed using a χ^2 test in the journalistic sub-corpora and a log-linear analysis in the academic sub-corpora to verify significant differences. Finally, the fifth step was the automatic overview of the number of subjective words for each segment connected by the most frequent causal connectives. These were also analyzed using χ^2 tests for each of the sub-corpora.

3. Results

41 The first two sections describe the distribution of connectives, the identification of the most frequent causal connectives and of subjective words in journalistic and academic texts. The last section presents the identification of subjective words in S₁ and S₂ of the most frequent connectives identified in the corpus.

3.1. Subjectivity in journalistic texts

42 The total word count (see 2.2) in the journalistic sub-corpora was 175,466, and the absolute number of causal connectives was 472. Then, after carrying out the manual analysis (see 2.3), the total number of causal connectives decreased to 238. Table 3 illustrates the distribution of causal connectives per text type.

| | News | Editorial | Total |
|-------------|--------|-----------|---------|
| Words | 88,014 | 87,452 | 175,466 |
| Connectives | 97 | 141 | 238 |
| % | 0.11 | 0.16 | 0.14 |

Table 3 – Causal connectives in the journalistic sub-corpora

43 Table 3 shows that the percentage of linguistic markers is 0.14% of the total word count in the journalistic sub-corpora. Regarding the distribution of linguistic markers, the highest amount is concentrated in editorials (0.16%). The statistical analyses demonstrated that there was a significant relationship between the use of connectives and the text type (LL2 (1): 8.47, $p = .000$); following the odds ratio, the use of connectives was 0.68 times more likely to occur in a persuasive text such as editorials than in an informative text such as news.

44 The manual analysis (see 2.3) also shed light on the most frequent linguistic markers. Eleven causal connectives were identified as the most frequent connectives in the journalistic sub-corpora on the basis of a minimal frequency of five. Table 4 gives the absolute numbers of these connectives and their percentages per text type, showing that they occur more frequently in editorials (58.7%) than in news texts (41.3%).

45 The automatic identification of subjective words in the journalistic sub-corpora using the Spanish lexicon of subjectivity (Molina-González et al., 2013) showed that in the editorials 5,219 words or 5.93% of the words were subjective, while the news texts contain 3,360 or 3.84% subjective words. The relationship between the text type and subjectivity was significant (χ^2 (1): 411.11, $p = .000$); following the odds ratio, subjective words were 1.58 times more likely to occur in a persuasive text such as editorials than in an informative text such as news. The interaction between these two variables is represented in Figure 1.

46 Figure 1 shows that the observed frequency of subjective words in editorials was higher than was expected on the basis of chance (indicated in violet), and the observed frequency of non-subjective words was lower (in purple) than was expected on the basis of chance. In the case of news, the opposite holds.

3.2. Subjectivity in academic texts

47 The total word counting (see 2.2) in the academic sub-corpora was 273,359, and the absolute number of causal connectives was 1,185. After carrying out the manual analysis (see 2.3), the total number of causal connectives decreased to 785 (cf. Section 2.3). Table 5 illustrates the distribution of causal connectives per text type in both domains. The abbreviation ESS corresponds to essays, RA to research articles, TB to textbooks, Ed to education and Ps to psychology.

| Connectives | News | | Editorial | | Total | % |
|----------------------------------|-------|------|-----------|------|-------|------|
| | Total | % | Total | % | | |
| <i>Porque</i> (“because”) | 26 | 30.6 | 53 | 43.8 | 79 | 38.3 |
| <i>Ya que</i> (“since”) | 18 | 21.2 | 4 | 3.3 | 22 | 10.7 |
| <i>Por eso</i> (“that’s why”) | 7 | 8.2 | 13 | 10.7 | 20 | 9.7 |
| <i>Pues</i> (“so”) | 6 | 7.1 | 11 | 9.1 | 17 | 8.3 |
| <i>Por lo que</i> (“so that”) | 13 | 15.3 | 4 | 3.3 | 17 | 8.3 |
| <i>Puesto que</i> (“given that”) | 4 | 4.7 | 10 | 8.3 | 14 | 6.8 |
| <i>Dado que</i> (“given that”) | 1 | 1.2 | 8 | 6.6 | 9 | 4.4 |
| <i>Por tanto</i> (“therefore”) | 1 | 1.2 | 8 | 6.6 | 9 | 4.4 |
| <i>Así</i> (“thus”) | 3 | 3.5 | 4 | 3.3 | 7 | 3.4 |
| <i>De modo que</i> (“so that”) | 2 | 2.4 | 5 | 4.1 | 7 | 3.4 |
| <i>Por ello</i> (“thus”) | 4 | 4.7 | 1 | 0.8 | 5 | 2.4 |
| Total | 85 | 41.3 | 121 | 58.7 | 206 | 100 |

Table 4 – Absolute numbers and percentages of the most frequent causal connectives in the journalistic sub-corpora¹⁰

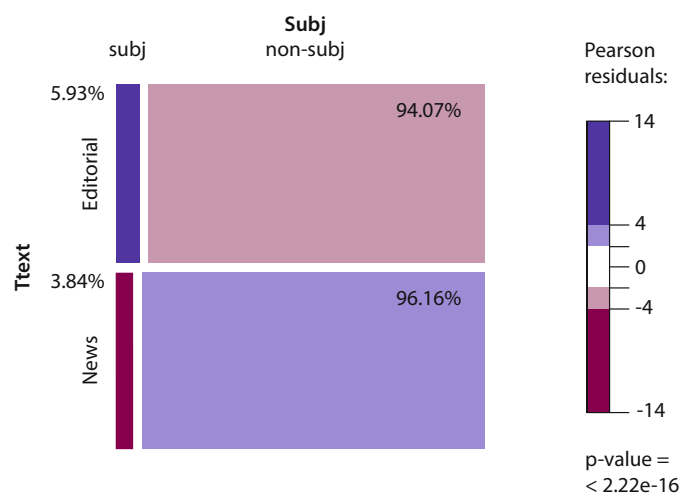


Figure 1 – Mosaic plot of relative frequencies of subjective and non-subjective words in journalistic texts

10. The percentages are column percentages. For example, 30.6% of the causal connectives identified in news texts were *porque*.

| | ESS Ed | ESS Ps | RA Ed | RA Ps | TB Ed | TB Ps | Total |
|-------------|--------|--------|--------|--------|--------|--------|---------|
| Words | 45,409 | 44,853 | 45,668 | 46,829 | 45,510 | 45,090 | 273,359 |
| Connectives | 164 | 130 | 147 | 89 | 121 | 134 | 785 |
| % | 0.36 | 0.29 | 0.32 | 0.19 | 0.27 | 0.30 | 0.29 |

Table 5 – Causal connectives in the academic sub-corpora

48 Table 5 shows that the percentage of linguistic markers is 0.29% of the total word count in the academic sub-corpora. The highest percentage in education is found in essays (0.36%), then in research articles (0.32%), and the lowest in textbooks (0.27%). In psychology, the highest percentage is found in textbooks (0.30%), then in essays (0.29%), and the lowest in research articles (0.19%). The statistical analysis showed a significant three-way interaction between text type, domain and the use of connectives (χ^2 (2): 12.29, $p = .01$). This three-way interaction was further explored by analyzing the relationship between the text type and the use of connectives in the two domains separately. Results demonstrated that there was a statistically significant relationship between the text type and the use of connectives in education (χ^2 (2): 6.61, $p = .037$) and in psychology (χ^2 (2): 12.85, $p = .002$). The standardized residuals in the education domain showed that there were fewer connectives than was expected by chance in textbooks, whereas in psychology there were fewer connectives than was expected by chance in research articles. The manual analysis (see 2.3) identified eight specific causal connectives as the most frequent connectives in the academic sub-corpora on the basis of a minimal frequency of five per domain, psychology and education. Table 6 shows the absolute numbers and percentages of these linguistic markers per text type in the two domains. The same pattern identified in the previous automatic analysis is observed: the highest number of linguistic markers is found in essays (38.7%), then in textbooks (32.3%) and the lowest in research articles (29%).

49 The automatic analysis of subjective words in the academic sub-corpora using the Spanish lexicon of subjectivity (Molina-González et al., 2013) revealed that the essay contains 4.1% subjective words in education and 4.3% in psychology; the textbook contains 4.1% in education and 4.5% in psychology; and the research article contains 2.9% in education and 3.8% in psychology. The statistical analysis showed a significant three-way interaction between the text type, subjectivity and domain (χ^2 (2): 25.59, $p = .000$). This three-way interaction was further explored by analyzing the relationship between the text type and subjectivity in the two domains separately. Results demonstrated that there was a statistically significant relationship between the text type and subjectivity in education (χ^2 (2): 120.90, $p = .000$) and in psychology (χ^2 (2): 27.25, $p = .000$). The standardized residuals in education showed that there were more subjective words than was expected by chance in essays and textbooks, whereas there were fewer subjective words than was

| Connectives | Essays | | | | Research Articles | | | | Textbooks | | | | Total | % |
|--------------------------------------|--------|----|-------|------|-------------------|----|-------|------|-----------|----|-------|------|-------|------|
| | Ed | Ps | Total | % | Ed | Ps | Total | % | Ed | Ps | Total | % | | |
| <i>Ya que</i> ("since") | 29 | 17 | 46 | 22.2 | 14 | 9 | 23 | 14.8 | 19 | 18 | 37 | 21.4 | 106 | 19.8 |
| <i>Porque</i> ("because") | 22 | 25 | 47 | 22.7 | 10 | 5 | 15 | 9.7 | 20 | 16 | 36 | 20.8 | 98 | 18.3 |
| <i>Así</i> ("thus") | 15 | 13 | 28 | 13.5 | 18 | 16 | 34 | 21.9 | 12 | 5 | 17 | 9.8 | 79 | 14.8 |
| <i>Pues</i> ("so") | 26 | 11 | 37 | 17.9 | 10 | 5 | 15 | 9.7 | 8 | 18 | 26 | 15 | 78 | 14.6 |
| <i>Por tanto</i> ("therefore") | 7 | 5 | 12 | 5.8 | 17 | 7 | 24 | 15.5 | 8 | 8 | 16 | 9.2 | 52 | 9.7 |
| <i>Por lo que</i> ("so that") | 6 | 9 | 15 | 7.2 | 14 | 8 | 22 | 14.2 | 8 | 7 | 15 | 8.7 | 52 | 9.7 |
| <i>Por lo tanto</i> ("therefore") | 7 | 6 | 13 | 6.3 | 5 | 5 | 10 | 6.5 | 8 | 8 | 16 | 9.2 | 39 | 7.3 |
| <i>Puesto que</i> ("given that") | 7 | 2 | 9 | 4.3 | 6 | 6 | 12 | 7.7 | 1 | 9 | 10 | 5.8 | 31 | 5.8 |
| Total | 119 | 88 | 207 | 38.7 | 94 | 61 | 155 | 29 | 84 | 89 | 173 | 32.3 | 535 | 100 |

Table 6 – Absolute numbers and percentages of the most frequent causal connectives in the academic sub-corpora¹¹

11. The percentages are column percentages. For example, 22.2% of the causal connectives identified in essays were *ya que*.

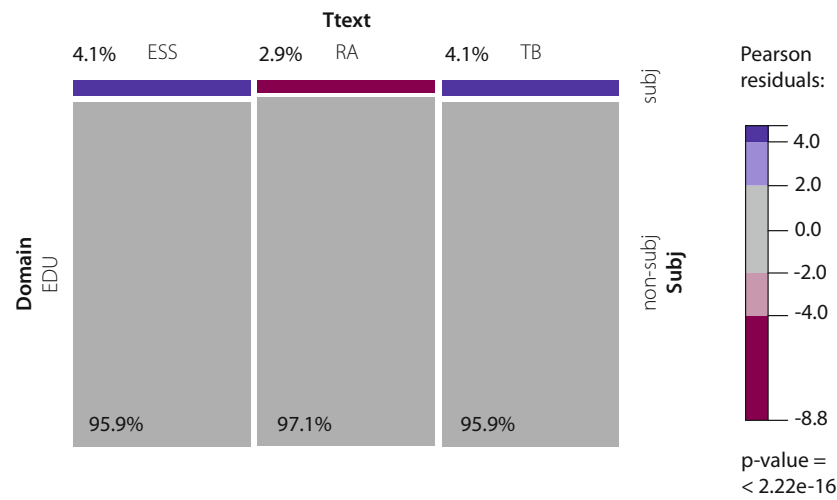


Figure 2 – Mosaic plot of relative frequencies of subjective and non-subjective words in academic texts of education

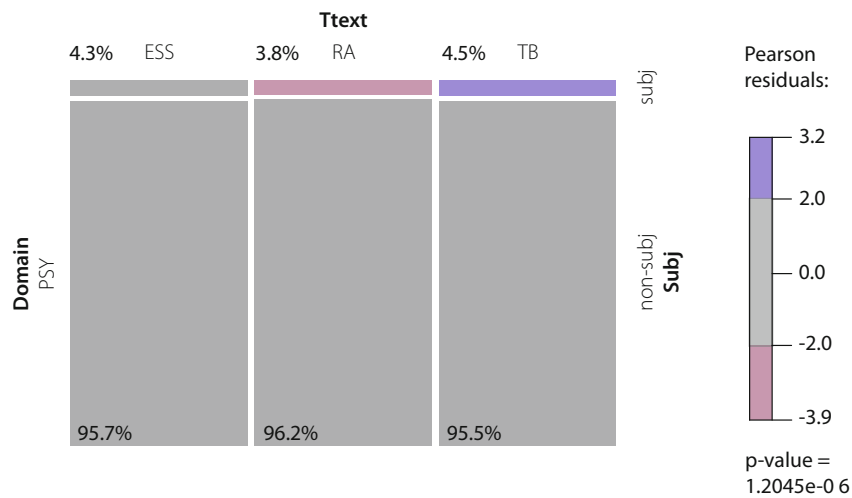


Figure 3 – Mosaic plot of relative frequencies of subjective and non-subjective words in academic texts of psychology

expected by chance in research articles. Likewise, in psychology, the standardized residuals in research articles and textbooks go in the same direction, but are less extreme. In the case of essays, the results did not differ from what was expected on the basis of chance. This interaction between text type and subjectivity in education and psychology is represented in Figures 2 and 3, respectively.

3.3. Subjectivity in segments

50 The goal of this part of the study was to see whether the most frequent causal connectives co-occur more or less with subjective words and whether these connectives are likely to differ in the degree of subjectivity in their local environment (S₁ and S₂). The percentages of subjective words in S₁ and S₂ of journalistic texts are illustrated in Figures 4 and 5.

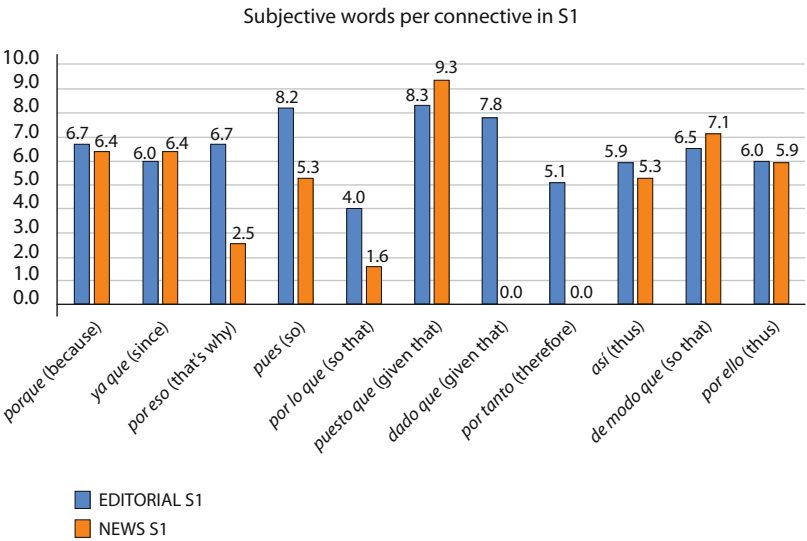


Figure 4 – Percentage of subjective words in S₁ of journalistic texts

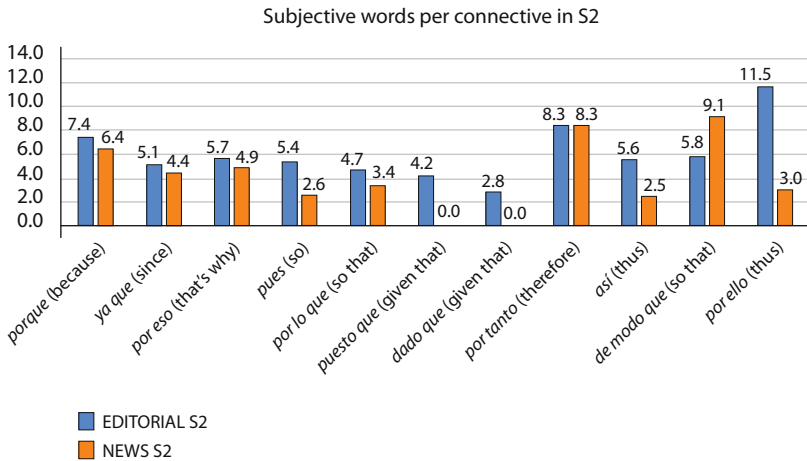


Figure 5 – Percentage of subjective words in S₂ of journalistic texts

Figure 4 shows that most of the percentages of subjective words in S1 are higher in editorials (6.5% average) than in news (4.5% average). In Figure 5, we observe the same pattern, although with lower percentages (6.0% average in editorials and 4.1% average in news). The statistical analyses carried out for S1 (Figure 4) and S2 (Figure 5) demonstrated that there was no significant relationship between connectives and subjectivity (all p 's > .10). This implies that, overall, none of the connectives have a clear preference for subjective or objective contexts. Still, the analysis revealed some tendencies of Spanish connectives that are worth mentioning. For instance, we can observe that in S1 (Figure 4) the connective *por lo que* ("so that") co-occurs with few subjective words, while the connectives *puesto que* ("given that"), *pues* ("so"), *de modo que* ("so that") co-occur with more subjective words. Likewise, in S2 (Figure 5) we can observe a similar pattern with the connectives *dado que* ("given that") and *por lo que* ("so that"), which co-occur with few subjective words, while the connectives *por tanto* ("therefore"), *por ello* ("thus"), *de modo que* ("so that") and *porque* ("because") co-occur with more subjective words. These observed tendencies may be useful in further studies on semantic-pragmatic profiles of Spanish connectives.

The percentages of subjective words in S1 and S2 in academic texts are shown in Figures 6 and 7, respectively. The statistical analyses carried out for S1 (Figure 6) and S2 (Figure 7) demonstrated that there was no significant relationship between connectives and subjectivity (all p 's > .10). By observing the percentages of subjective words in S1 (Figure 6), we can see that all the connectives co-occur with a similar number of subjective words. In S2 (Figure 7) the connectives *por tanto* ("therefore") and *puesto que* ("given that") co-occur with more subjective words, while no connective co-occurs with few subjective words.

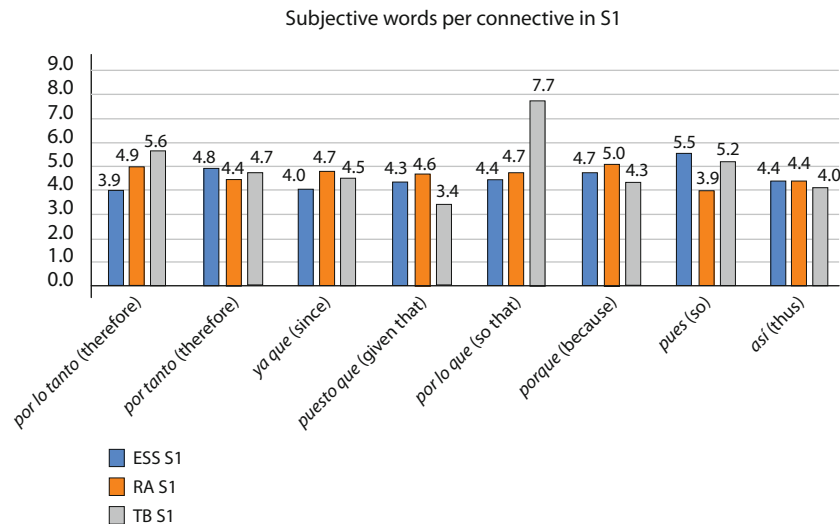


Figure 6 – Percentage of subjective words in S1 of academic texts

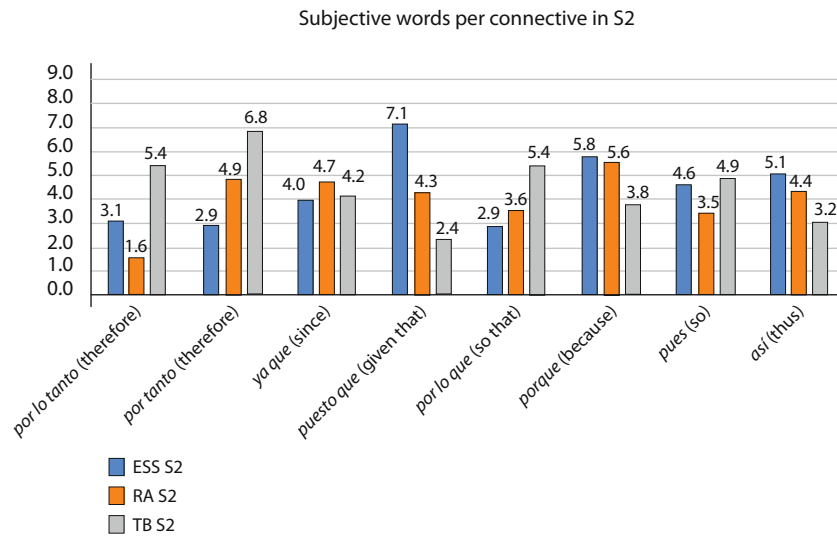


Figure 7 – Percentage of subjective words in S2 of academic texts

4. Discussion and conclusions

53 The purpose of the current study was to explore whether we could determine the degree to which Spanish causal connectives encode subjectivity across different text types, by carrying out automatic analyses. We explored a type of automatic analysis that allowed us to gain insight into coherence relations and their linguistic markers in terms of subjectivity. We expected that causal connectives would more often signal subjectivity if they occurred in a subjective environment, that is, in a context containing relatively many subjective words. For this reason, the first methodological step was to construct a corpus to identify the most frequent causal connectives in different text types. Then, the automatic identification of subjective words in the different text types using a Spanish lexicon of subjectivity (Molina-González et al., 2013) was carried out to explore the general contexts of linguistic markers. Finally, this same procedure was carried out in the segments connected by the most frequent causal connectives with the purpose of exploring the local contexts of linguistic markers.

54 Three hypotheses were formulated:

- (H1) *The distribution and the frequency of connectives vary in different text types.*
- (H2) *The number of subjective words depends on text types:*
 - (a) *Journalistic discourse: subjective words are more frequent in editorials than in news.*
 - (b) *Academic discourse: subjective words are more frequent in essays than in research articles, and more frequent in research articles than in textbooks.*

- (H3) *Subjective connectives (connectives that occur frequently in subjective text types) co-occur more often with subjective words than non-subjective connectives.*

55 Regarding H1, we can conclude that it is borne out. The results showed that there was a significant relationship between the use of connectives and the text type in the journalistic sub-corpora and that there was a significant three-way interaction between the use of connectives, text type and domain in the academic sub-corpora. We also identified the most frequent causal connectives in both types of discourse. Moreover, we identified that some connectives are more often used in journalistic texts, specifically *por eso* (“that’s why”), *dado que* (“given that”), *de modo que* (“so that”) and *por ello* (“thus”); other connectives are more often used in academic texts, particularly *por lo tanto* (“therefore”); and we also identified some connectives that are used in both, such as *porque* (“because”), *ya que* (“since”), *pues* (“so”), *por lo que* (“so that”), *puesto que* (“given that”), *por tanto* (“therefore”) and *así* (“thus”). Hence, these results allow us to suggest that the text type plays a relevant role, and that it is a variable that constrains the use and distribution of linguistic markers. Consequently, the text type should be considered in future studies focusing on the phenomenon of coherence relations and linguistic markers.

56 With respect to H2 (a), we conclude that it is also borne out. The results revealed a relationship between the text type and subjectivity. The frequency of subjective words was higher than was expected on the basis of chance in editorials, whereas it was lower than was expected on the basis of chance in news. These results are in accordance with the characteristics of these text types. Therefore, the argumentative and informative nature of these texts is reflected in the use of more or fewer subjective words, respectively.

57 In the case of academic texts, we can conclude that H2 (b) is partially borne out since essays contain more subjective words than expected, but this was only observed in the domain of education. A possible explanation for these results could be that academic disciplines have different conventional ways of communication (Hyland, 2008) and that these differences can be reflected in some of the characteristics of the texts used in these communities. For example, in this study, we identified that essays in education contain more subjective words than essays in psychology, which could indicate that the use of subjective words is more acceptable in this domain than in psychology. Another possible explanation could be related to the source of the essays in psychology. The journal that was selected for these essays does not name the texts explicitly “essays”, as the journal of education does. It includes the essays in another category called “theoretical reports”. Therefore, this may have influenced the writing of authors.

58 Regarding the textbooks, the results lead us to reject H2 (b) since the frequency of subjective words was higher than expected in both domains, education and psychology. In order to find possible explanations for these results, we reviewed the local context of the five most frequent subjective words identified in textbooks of education and psychology (*mayor* [“greater”], *bien* [“well”], *trastornos* [“disorders”],

importante [“important”] and *mejor* [“better”]). Ten examples of each of these subjective words in the two domains (100 examples) were selected randomly. This review showed that indeed most of the examples reveal the author’s evaluation and that the identified subjective words were used in a subjective context. Therefore, this refutes our previous expectations about textbooks, which led us to reconsider our assumptions about the degrees of subjectivity in textbooks. Examples [11], [12] and [13] illustrate some of these subjective cases.

- [11] Las necesidades más importantes en este apartado son las referidas a aquellos aprendizajes que requieren mayor abstracción, que se apoyan en la representación y simbolización.

‘The most important requirements in this section are related to those types of learning that require greater abstraction, that rely on representation and symbolization.’

- [12] Por tanto, nadie mejor que la propia familia para, con una serie de principios y pautas fáciles de llevar a la práctica a la hora de comunicarse con sus hijos e hijas, convertirse en un elemento vital para el buen desarrollo del lenguaje.

‘Therefore, no one is better than your own family to become, with a set of principles and guidelines easy to put into practice when communicating with your children, a vital element for good language development.’

- [13] Otro aspecto relevante que se indica en el capítulo de atención son los procesos controlados versus procesos automáticos. Como bien señala Neisser (1976) nuestra capacidad de procesamiento no está limitada ni por las características estructurales ni por las funcionales del sistema de procesamiento, sino que, sobre todo, depende de que desarrollemos las habilidades específicas necesarias para ejecutar una determinada tarea.

‘Another important aspect indicated in the care chapter is controlled processes versus automatic processes. As Neisser (1976) points out well, our processing capacity is not limited by the structural characteristics nor the functional characteristics of the processing system but, above all, depends on if we develop the specific skills that are necessary to perform a certain task.’

59 However, it should be mentioned that in particular cases (19 out of 100), the identified subjective words were in fact used to express other meanings than the supposed subjective usage. For example, the Spanish word *mayor* (“greater”) in [14] refers to “the majority”, therefore it is not used subjectively as a way of evaluating the superiority or relevance of something. In [15], the use of the Spanish word *bien* (“well”) functions as part of the concessive connective *si bien* (“although”), and therefore, is not used subjectively as a way of evaluating the appropriate quality of something. Another similar case is the Spanish word *trastorno* (“disorder”) in [16] since it can be considered as a technical medical term rather than a word used in a negative subjective context. Therefore, not all subjective words identified with the

lexicon necessarily express the “true” subjectivity we were looking for. This finding shows the need for a better identification of subjectively used words by using more sophisticated automatic analyses than simply literal word counts.

- [14] También resultaría necesario investigar subgrupos minoritarios. Por ejemplo, el colectivo de mujeres, ya que la mayor parte de los estudios han sido realizados con hombres (Echeburúa, 1993), [...].

‘It would also be necessary to look at minority subgroups. For example, the collective of women, since the major part of the studies has been carried out with men (Echeburúa, 1993), [...].’

- [15] El test de Stroop ha demostrado ser discriminativo, si bien su uso se restringe a la sintomatología impulsiva de los niños o niñas hiperactivos.

‘The Stroop test has proved to be discriminative, although its use is restricted to impulsive symptoms of hyperactive boys and girls.’

- [16] Se han incluido en la GPC comorbilidades psiquiátricas y no psiquiátricas que pueden requerir otro tipo de atención: epilepsia, trastornos del espectro autista, trastornos del estado de ánimo, trastorno bipolar y trastorno por abuso de sustancias. No incluye las intervenciones específicas para los trastornos comórbidos psiquiátricos y no psiquiátricos del TDAH.

‘In the CPG, psychiatric and non-psychiatric comorbidities have been included that may require another type of care: epilepsy, autism spectrum disorders, mood disorders, bipolar disorder, and substance abuse disorders. It does not include specific interventions for psychiatric and non-psychiatric comorbid disorders of ADHD.’

60 Regarding the research articles, the results also lead us to reject H₂ (b), since these texts contain fewer subjective words than expected, which was observed in both domains. This implies a reconsideration of our assumptions regards research articles, which requires further research. A possible explanation for these unexpected results could be the fact that the research article is a text by which researchers accrue credibility with their peers and consolidate their position in their disciplines. Therefore, the authors may avoid the use of subjective words because these could diminish the credibility of what they are communicating. It seems to be that the authors of research articles present the information and the arguments in the most acceptable and convincing way; they aim to demonstrate certainty and maximum plausibility through their research claims (Hyland, 2001). In this sense, the persuasive nature of a research article could be reflected through other rhetorical strategies, not necessarily using subjective words. One such resource seems to be the inclusion of references to the work of other authors, that helps to establish a framework for the acceptance of arguments by showing how the work is based on previous studies and displays the author as an expert in the discipline (Hyland, 2008); another is self-mention, that allows writers to claim authority by expressing their convictions, emphasizing their contribution to the field and seeking recognition for their work (Hyland, 2001; Kuo, 1999).

- 61 Finally, regarding H₃, we conclude that it should be rejected since the statistical analyses did not allow us to conclude whether some connectives co-occur more or less with subjective words. We were able to identify some tendencies that could be put to the test in further studies of possible differences of subjectivity in Spanish causal connectives. We found that the following connectives co-occurred frequently with subjective words: *puesto que* (“given that”), *pues* (“so”) and *de modo que* (“so that”) in S₁ of journalistic texts; *por tanto* (“therefore”), *por ello* (“thus”), *de modo que* (“so that”), *porque* (“because”) in S₂ of journalistic texts; and *por tanto* (“therefore”) and *puesto que* (“given that”) in S₂ of academic texts. These results suggest that these connectives seem to be subjective. On the other hand, some connectives with low percentages of subjective words were identified. These cases are the connectives *por lo que* (“so that”) in S₁ of journalistic texts and *dado que* (“given that”), *por lo que* (“so that”) and *puesto que* (“given that”) in S₂ of journalistic texts. These results can be taken to indicate that such connectives mainly express objective relations. Furthermore, it is interesting that *por tanto* (“therefore”) showed high percentages of subjective words in news, editorials and textbooks, whereas *puesto que* (“given that”) had high percentages of subjective words in news, editorials, and essays. This last result provides support for the study of Pit et al. (1996), which states that *puesto que* (“therefore”) is a connective that signals subjectivity and especially with a speaker in an evaluative role. In the same vein, it is remarkable that the connective *ya que* (“since”) did not obtain high percentages of subjective words in any of the text types, despite the fact that previous studies have considered it as a subjective connective (Goethals, 2002 and 2010; Pit et al., 1996). Do such results suggest that it is possible to identify a systematic variation of Spanish connectives in terms of subjectivity? We take them as preliminary indications that require further study of these specific connectives.
- 62 As for methodology, the use of the lexicon of subjectivity is proposed as a first step in the exploration and evaluation of the environments of connectives that have received little attention. This lexicon allowed us to operationalize subjectivity; it proved to be a valid instrument for the identification of degrees of subjectivity across different text types because it confirmed that some texts were indeed more or less subjective than others.
- 63 As for the use of automatic analyses, the results obtained allow us to conclude that the type of analysis that was carried out is far from sophisticated, especially if we compare it with other automatic tools that have been designed for the analysis of cohesion and coherence of texts (Coh-Metrix: Graesser et al., 2011) or for the analysis of several features that aim to gain insight into the complexity of texts (T-Scan: Pander Maat et al., 2014), to mention just a few. In this study, we performed basic automatic analyses which facilitated the identification of connectives and degrees of subjectivity across large data sets in different text types. This type of analysis showed some limitations. Additional manual analyses were required to check whether the connectives were actually functioning as connectives, and whether they linked two clauses. In addition, we noticed that some words registered

as subjective words in the corpus, did not in fact have a subjective meaning. Therefore, an evaluation by a human analyst was decisive to elucidate their specific use and meaning. This situation shows that this type of analysis is not sensitive enough to local contexts, and it demands the search for the development of more sophisticated tools, that reduce the need for manual revisions. In this sense, further research that integrates collective work between computational and discourse linguists is required.

64 Another limitation was the small proportion of some frequent causal connectives, specifically in the case of journalistic texts. Although eleven connectives with a total number of 206 occurrences were identified as the most frequent causal connectives, at least five of them occurred fewer than ten times. The situation was different in the case of academic texts, even when a similar number of words per text type was considered (90,000 words). Eight connectives with a total number of 535 occurrences were identified as the most frequent connectives, and all of them occurred more than thirty times each. This difference was not foreseen because of the exploratory nature of the present study. However, future studies should consider the inclusion of more words per text type to prevent data sparseness, especially as automatic analyses allow us to explore even more data.

65 In spite of these limitations, we still want to argue in favor of the innovations that come with partially automatic analyses. Manual analyses involve certain disadvantages for the study of coherence relations and connectives (Spooren & Degand, 2010). The analysts have to deal with complex activities that can be laborious and time-consuming. Consequently, manual analyses are restricted to small data sets and the reliability of this type of analysis is often questioned because the results obtained depend exclusively on the analysts' interpretations. On the contrary, automatic analyses can offer certain advantages for the study of coherence relations and connectives. Although supervision is always required, this type of analysis can improve the quality of analyses and give greater reliability to the studies of discourse coherence because it is based on computational systems and is not totally dependent on the interpretation of the analyst. Furthermore, the use of automatic analyses allows us to explore a large amount of data and facilitates the corroboration of previous analyses carried out manually, which favors the generalization and replication in different contexts of use. Therefore, we believe that manual analyses can benefit from the inclusion of automatic analyses; the combination of the two types of analyses can result in a more reliable, useful and accurate methodology for the study of coherence relations and connectives in context.

References

- ALRED, G.J. & THELEN, E.A. 1993. Are Textbooks Contributions to Scholarship? *College Composition and Communication* 44 (4): 466-477.
- ASCHENBERG, H. & LOUREDA, Ó. 2011. *Marcadores del discurso: de la descripción a la definición*. Madrid – Frankfurt: Iberoamericana – Vervuert.
- ASHER, N. & LASCARIDES, A. 2003. *Logics of Conversation*. Studies in Natural Language Processing. Cambridge – New York: Cambridge University Press.
- BAYERL, P.S. & PAUL, K.I. 2011. What Determines Inter-coder Agreement in Manual Annotations? A Meta-analytic Investigation. *Computational Linguistics* 37 (4): 699-725.
- BELLO, A. 1847. *Gramática de la lengua castellana destinada al uso de los americanos*. R. Trujillo (ed.). Santa Cruz de Tenerife: Instituto universitario de lingüística Andrés Bello, Aula de cultura de Tenerife.
- BESTGEN, Y., DEGAND, L. & SPOOREN, W. 2006. Toward Automatic Determination of the Semantics of Connectives in Large Newspaper Corpora. *Discourse Processes* 41 (2): 175-193.
- BHATIA, V. 2002. A Generic View of Academic Discourse. In J. FLOWERDEW (ed.), *Academic Discourse*. Harlow – New York: Longman: 21-39.
- BHATIA, V. 2004. *Worlds of Written Discourse: A Genre-Based View*. London: Continuum.
- BLACKWELL, S.E. 2016. *Porque* in Spanish Oral Narratives: Semantic *Porque*, (Meta) Pragmatic *Porque* or Both? In A. CAPONE & J.L. MEY (eds.), *Interdisciplinary Studies in Pragmatics, Culture and Society*. Berlin – Heidelberg – New York: Springer: 615-651.
- BOUMA, G., VAN NOORD, G. & MALOUF, R. 2001. Alpino: Wide-Coverage Computational Analysis of Dutch. In W. DAELEMANS, K. SIMA'AN, J. VEENSTRA & J. ZAVREL (eds.), *Computational Linguistics in the Netherlands 2000*. Language and Computers 37. Amsterdam – New York: Rodopi: 45-59.
- BRIZ, A. 1998. *El español coloquial en la conversación: esbozo de pragmagramática*. Barcelona: Ariel.
- BRIZ, A., PONS, S. & PORTOLÉS, J. (eds.) 2008. *Diccionario de partículas discursivas del español*. Available online: <http://www.dpde.es>.
- BROERSMA, M. 2010. Journalism as Performative Discourse. The Importance of Form and Style in Journalism. In V. RUPAR (ed.), *Journalism and Meaning-Making: Reading the Newspaper*. Cresskill: Hampton Press: 15-35.
- CARLSON, L. & MARCU, D. 2001. *Discourse Tagging Reference Manual*. Available online: <http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>.
- CARLSON, L., MARCU, D. & OKUROWSKI, M.E. 2003. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In J. VAN KUPPEVELT & R.W. SMITH (eds.), *Current and New Directions in Discourse and Dialogue*. Dordrecht – Boston: Kluwer Academic Publishers: 85-112.
- CARTONI, B., ZUFFEREY, S. & MEYER, T. 2013. Annotating the Meaning of Discourse Connectives by Looking at Their Translation: The Translation-Spotting Technique. *Dialogue and Discourse* 4 (2): 65-86.
- CASADO VELARDE, M. 1991. Los operadores discursivos *es decir, esto es, o sea y a saber* en español actual: valores de lengua y funciones textuales. *Lingüística española actual* 13 (1): 87-116.

- CUHNA, I. DA, TORRES-MORENO, J.-M. & SIERRA, G. 2011. On the Development of the RST Spanish Treebank. In *Proceedings of the 5th Linguistic Annotation Workshop – LAW 5 (Portland, Oregon, 23-24 June 2011)*. Stroudsburg: Association for Computational Linguistics: 1-10. Available online: <https://pdfs.semanticscholar.org/0aa4/c504c324f4f-28f9a075ae611dic844824a22.pdf>.
- DEGAND, L. & PANDER MAAT, H. 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. Lexicon, Grammar, Discourse*. Utrecht: LOT: 175-199. Available online: <https://dspace.library.uu.nl/bitstream/handle/1874/295390/bookpart.pdf?sequence=2&isAllowed=y>.
- DÍAZ RANGEL, I., SIDOROV, G. & SUÁREZ GUERRA, S. 2014. Creación y evaluación de un diccionario marcado con emociones y ponderado para el español. *Onomazein* 29 (1): 31-46.
- DOMÍNGUEZ GARCÍA, M. 2007. *Conectores discursivos en textos argumentativos breves*. Madrid: Arco Libros.
- DUQUE, E. 2014. Signaling Causal Coherence Relations. *Discourse Studies* 16 (1): 25-46.
- DUQUE, E. 2016. *Las relaciones de discurso*. Madrid: Arco Libros.
- ECHEBURÚA, E. 1993. Las conductas adictivas: ¿una ruta común desde el “crack” al juego patológico? *Psicología conductual* 1 (3): 321-337.
- FERNÁNDEZ RAMÍREZ, S. 1951. *Gramática española: los sonidos, el nombre y el pronombre*. Madrid: Revista de Occidente.
- FIELD, A., MILES, J. & FIELD, Z. 2012. *Discovering Statistics Using R*. London – Thousand Oaks: Sage Publications.
- FUENTES, C. 2009. *Diccionario de conectores y operadores del español*. Madrid: Arco Libros.
- GARCÉS, G. 1791. *Fundamento del vigor y elegancia de la lengua castellana, expuesto en el propio y vario uso de sus partículas*. Madrid: Imprenta de la viuda Ibarra.
- GILI GAYA, S. 1961. *Curso superior de sintaxis española*. Barcelona: Bibliograf.
- GOETHALS, P. 2002. *Las conjunciones causales explicativas españolas “como”, “ya que”, “pues” y “porque”: un estudio semiótico-lingüístico*. Leuven – Paris – Dudley: Peeters.
- GOETHALS, P. 2010. A Multi-layered Approach to Speech Events: The Case of Spanish Justificational Conjunctions. *Journal of Pragmatics* 42 (8): 2204-2218.
- GRAESSER, A.C., GERNSBACHER, M.A. & GOLDMAN, S.R. (eds.) 2003. *Handbook of Discourse Processes*. Mahwah: L. Erlbaum.
- GRAESSER, A.C., MCNAMARA, D.S. & KULIKOWICH, J.M. 2011. Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher* 40 (5): 223-234.
- GÜNTHER, S. 1993. “... weil – man kann es ja wissenschaftlich untersuchen”. Diskurspragmatische Aspekte der Wortstellung in WEIL-Sätzen. *Linguistische Berichte* 143: 37-59.
- HALE, G., TAYLOR, C., BRIDGEMAN, B., CARSON, J., KROLL, B. & KANTOR, R. 1996. *A Study of Writing Tasks Assigned in Academic Degree Programs*. TOEFL Research Reports 54. Princeton: Educational Testing Service.
- HENRY, A. & ROSEBERRY, R.L. 1999. Raising Awareness of the Generic Structure and Linguistic Features of Essay Introductions. *Language Awareness* 8 (3-4): 190-200.

- HOEK, J., EVERS-VERMEUL, J. & SANDERS, T. 2017. Segmenting Discourse: Incorporating Interpretation into Segmentation? *Corpus Linguistic and Linguistic Theory*. Ahead of print: <http://doi.org/10.1515/cllt-2016-0042>.
- HYLAND, K. 1990. A Genre Description of the Argumentative Essay. *REL C Journal* 21 (1): 66-78.
- HYLAND, K. 1996. Writing without Conviction? Hedging in Science Research Articles. *Applied Linguistics* 17 (4): 433-454.
- HYLAND, K. 2001. Humble Servants of the Discipline? Self-Mention in Research Articles. *English for Specific Purposes* 20 (3): 207-226.
- HYLAND, K. 2008. Genre and Academic Writing in the Disciplines. *Language Teaching* 41 (4): 543-562.
- HYLAND, K. 2009. *Academic Discourse: English in a Global Context*. London: Continuum.
- HYLAND, K. 2011. Disciplines and Discourses: Social Interactions in the Construction of Knowledge. In D. STARKE-MEYERRING, A. PARÉ, N. ARTEMEVA, M. HORNE & L. YOUSOUBOVA (eds.), *Writing in Knowledge Societies*. Perspectives on Writing. Fort Collins – Anderson: The WAC Clearinghouse – Parlor Press: 193-214.
- JIMÉNEZ ZAFRA, S.M., MARTÍNEZ CÁMARA, E., MARTÍN VALDIVIA, M.T. & UREÑA LÓPEZ, L.A. 2014. SINAI-ESMA: An Unsupervised Approach for Sentiment Analysis in Twitter. In *Proceedings of the TASS Workshop at SEPLN Conference (Girona, Spain, September 16-19, 2014)*. See online: <http://www.sepln.org/workshops/tass/2014/tass2014.php>.
- KELLER, R. 1995. The Epistemic *Weil*. In D. STEIN & S. WRIGHT (eds.), *Subjectivity and Subjectivisation: Linguistic Perspectives*. Cambridge – New York: Cambridge University Press: 16-30.
- KUO, C.-H. 1999. The Use of Personal Pronouns: Role Relationships in Scientific Journal Articles. *English for Specific Purposes* 18 (2): 121-138.
- KWAN, B.S.C. 2006. The Schematic Structure of Literature Reviews in Doctoral Theses of Applied Linguistics. *English for Specific Purposes* 25 (1): 30-55.
- LEÑERO, V. & MARÍN, C. 1986. *Manual de periodismo*. México: Grijalbo.
- LEVSHINA, N. & DEGAND, L. 2017. Just *Because*: In Search of Objective Criteria of Subjectivity Expressed by Causal Connectives. *Dialogue and Discourse* 8 (1): 132-150.
- LI, F., EVERS-VERMEUL, J. & SANDERS, T. 2013. Subjectivity and Result Marking in Mandarin. *Chinese Language and Discourse* 4 (1): 74-119.
- MARTÍ, M. 2008. *Los marcadores en español L/E: conectores discursivos y operadores pragmáticos*. Madrid: Arco Libros.
- MARTÍN ZORRAQUINO, M.A. 2003. Marcadores del discurso y diccionario: sobre el tratamiento lexicográfico de desde luego. In M.T. ECHENIQUE ELIZONDO & J.P. SÁNCHEZ MÉNDEZ (eds.), *Lexicografía y lexicología en Europa y América. Homenaje a Günter Haensch en su 80 aniversario*. Madrid – Valencia: Gredos – Biblioteca valenciana: 439-452.
- MARTÍN ZORRAQUINO, M.A. & MONTOLÍO, E. 1998. *Marcadores del discurso*. Barcelona: Ariel.
- MARTÍN ZORRAQUINO, M.A. & PORTOLÉS, J. 1999. Los marcadores del discurso. In I. BOSQUE & V. DEMONTE (eds.), *Gramática descriptiva de la lengua española*. Madrid: Espasa Calpe. Vol. 3: *Entre la oración y el discurso. Morfología*: 4051-4214.

- MARTÍNEZ, R. 1997. *Conectando texto: guía para el uso efectivo de elementos conectores en castellano*. Barcelona: Octaedro.
- MCENERY, T. & HARDIE, A. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge – New York: Cambridge University Press.
- MOLINA-GONZÁLEZ, M.D., MARTÍNEZ-CÁMARA, E., MARTÍN-VALDIVIA, M.-T. & PEREA-ORTEGA, J.M. 2013. Semantic Orientation for Polarity Classification in Spanish Reviews. *Expert Systems with Applications* 40 (18): 7250–7257.
- MOLINER, M. 1966. *Diccionario de uso del español*. Madrid: Gredos.
- MONTOLÍO, E. 2001. *Conectores de la lengua escrita*. Barcelona: Ariel.
- MYERS, G.A. 1992. Textbooks and the Sociology of Scientific Knowledge. *English for Specific Purposes* 11 (1): 3–17.
- NEISSER, U. 1976. *Cognition and Reality: Principles and Implications of Cognitive Psychology*. San Francisco: W. H. Freeman.
- OOSTDIJK, N., REYNAERT, M., HOSTE, V. & VAN DEN HEUVEL, H. 2013. *SoNaR User Documentation – version 1.0.4*. Available online: http://ticclops.uvt.nl/SoNaR_end-user_documentation_v.1.0.4.pdf.
- PANDER MAAT, H. & DEGAND, L. 2001. Scaling Causal Relations and Connectives in Terms of Speaker Involvement. *Cognitive Linguistics* 12 (3): 211–245.
- PANDER MAAT, H., KRAF, R., VAN DEN BOSCH, A., VAN GOMPEL, M., KLEIJN, S., SANDERS, T. & VAN DER SLOOT, K. 2014. T-Scan: A New Tool for Analyzing Dutch Text. *Computational Linguistics in the Netherlands Journal* 4: 53–74.
- PANDER MAAT, H. & SANDERS, T. 2000. Domains of Use or Subjectivity? The Distribution of Three Dutch Causal Connectives Explained. In E. COUPER-KUHLEN & B. KORTMANN (eds.), *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives*. Topics in English Linguistics 33. Berlin – New York: M. de Gruyter: 57–82.
- PANDER MAAT, H. & SANDERS, T. 2001. Subjectivity in Causal Connectives: An Empirical Study of Language in Use. *Cognitive Linguistics* 12 (3): 247–273.
- PDTB RESEARCH GROUP 2007. *The Penn Discourse Treebank 2.0 Annotation Manual*. Available online: <http://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>.
- PÉREZ-ROSAS, V., BANEJA, C. & MIHALCEA, R., 2012. Learning Sentiment Lexicons in Spanish. In *Proceedings of the 8th International Conference on Language Resources and Evaluation – LREC 2012*. Paris: European Language Resources Association: 3077–3081. Available online: http://www.lrec-conf.org/proceedings/lrec2012/pdf/1081_Paper.pdf.
- PIT, M. 2006. Determining Subjectivity in Text: The Case of Backward Causal Connectives in Dutch. *Discourse Processes* 41 (2): 151–174.
- PIT, M. 2007. Cross-Linguistic Analyses of Backward Causal Connectives in Dutch, German and French. *Languages in Contrast* 7 (1): 53–82.
- PIT, M., HULST, J. & PANDER MAAT, H. 1996. Subjectiviteit en de Spaanse connectieven *porque, ya que en puesto que*. *Gramma/TTT* 5 (3): 221–240.
- PONS, S. 1998. *Conexión y conectores. Estudio de su relación en el registro informal de la lengua*. Cuadernos de filología 27. Valencia: Universidad de Valencia.

- PONS, S. 2000. Los conectores. In A. BRIZ GÓMEZ (ed.), *¿Cómo se comenta un texto coloquial?* Barcelona: Ariel: 193-220.
- PORTOLÉS, J. 1998. *Marcadores del discurso*. Barcelona: Ariel.
- PRASAD, R., DINESH, N., LEE, A., MILTSAKAKI, E., ROBALDO, L., JOSHI, A. & WEBBER, B. 2008. The Penn Discourse Treebank 2.0. *Proceedings of the 6th International Conference on Language Resources and Evaluation – LREC 2008*. Paris: European Language Resources Association: 2961-2968. Available online: http://www.lrec-conf.org/proceedings/lrec2008/pdf/754_paper.pdf.
- SANDERS, J., SANDERS, T. & SWEETSER, E. 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.
- SANDERS, T. 1997. Semantic and Pragmatic Sources of Coherence: On the Categorization of Coherence Relations in Context. *Discourse Processes* 24 (1): 119-147.
- SANDERS, T. & SPOOREN, W. 2013. Exceptions to Rules: A Qualitative Analysis of Backward Causal Connectives in Dutch Naturalistic Discourse. *Text and Talk* 33 (3): 377-398. Available online: https://dspace.library.uu.nl/bitstream/handle/1874/299045/Exceptions%20to%20rules_SandersSpooren2013.pdf?sequence=1&isAllowed=y.
- SANDERS, T. & SPOOREN, W. 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, T., SPOOREN, W. & NOORDMAN, L. 1992. Toward a Taxonomy of Coherence Relations. *Discourse Processes* 15 (1): 1-35.
- SANDERS, T., SPOOREN, W. & NOORDMAN, L. 1993. Coherence Relations in a Cognitive Theory of Discourse Representation. *Cognitive Linguistics* 4 (2): 93-133.
- SANTOS RÍO, L. 2003. *Diccionario de partículas*. Salamanca: Luso-Española de Ediciones.
- SIDOROV, G., MIRANDA-JIMÉNEZ, S., VIVEROS-JIMÉNEZ, F., GELBUKH, A., CASTRO-SÁNCHEZ, N., VELÁSQUEZ, F., DÍAZ-RANGEL, I., SUÁREZ-GUERRA, S., TREVIÑO, A. & GORDON, J. 2013. Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets. In I. BATYRSHIN & M. GONZÁLEZ MENDOZA (eds.), *Advances in Artificial Intelligence: 11th Mexican International Conference on Artificial Intelligence – MICAI 2012 (San Luis Potosí, Mexico, October 27-November 4, 2012)*. Berlin – Heidelberg – New York: Springer: 1-14.
- SILVER, M. 2006. *Language Across Disciplines: Towards a Critical Reading of Contemporary Academic Discourse*. Boca Raton: BrownWalker Press.
- SPOOREN, W. & DEGAND, L. 2010. Coding Coherence Relations: Reliability and Validity. *Corpus Linguistics and Linguistic Theory* 6 (2): 241-266.
- SPOOREN, W., SANDERS, T., HUISKES, M. & DEGAND, L. 2010. Subjectivity and Causality: A Corpus Study of Spoken Language. In S. RICE and J. NEWMAN (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*. Stanford: CSLI Publications: 241-255.
- STUKKER, N. & SANDERS, T. 2009. Another('s) Perspective on Subjectivity in Causal Connectives: A Usage-Based Analysis of Volitional Causal Relations. *Discours* 4: 1-33. Available online: <https://discours.revues.org/7260>.

- STUKKER, N. & SANDERS, T. 2010. Diverging Frequency Effects in Causal Connectives across Discourse Contexts. A Usage-Based Interpretation. In *Proceedings of the 34th International LAUD Symposium Cognitive Sociolinguistics: Language Variation in Its Structural, Conceptual and Cultural Dimensions*. Landau: LAUD: 110-111.
- STUKKER, N. & SANDERS, T. 2012. Subjectivity and Prototype Structure in Causal Connectives: A Cross-Linguistic Perspective. *Journal of Pragmatics* 44 (2): 169-190.
- SWALES, J. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge – New York: Cambridge University Press.
- SWALES, J. 1995. The Role of the Textbook in EAP Writing Research. *English for Specific Purposes* 14 (1): 3-18.
- SWEETSER, E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge – New York: Cambridge University Press.
- TABOADA, M. & GÓMEZ-GONZÁLEZ, M. 2010. Discourse Markers and Coherence Relations: Comparison across Markers, Languages and Modalities. *Linguistics and the Human Sciences* 6 (1-3): 17-41.
- VAN DIJK, T.A. 1988. *News as Discourse*. Hillsdale: L. Erlbaum.
- VAN DIJK, T.A. 1993. *Elite Discourse and Racism*. Newbury Park: Sage Publications.
- VAN EEMEREN, F., GROOTENDORST, R., JACKSON, S. & JACOBS, S. 1997. Argumentation. In T.A. VAN DIJK (ed.), *Discourse as Structure and Process*. Newbury Park: Sage Publications: 208-229.
- VÁZQUEZ VEIGA, N. 2002. Diccionario de colocaciones y marcadores del español: esbozo de una entrada de un marcador discursivo. In M.T. DÍAZ HORMIGO (ed.), *IV Congreso de Lingüística General (Cádiz, del 3 al 6 de abril 2000)*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz. Vol. 4: 2459-2472.
- VERHAGEN, A. 2005. *Constructions of Intersubjectivity: Discourse, Syntax, and Cognition*. New York: Oxford University Press.
- VILLENA ROMÁN, J., GARCÍA MORERA, J., LANA SERRANO, S. & GONZÁLEZ CRISTÓBAL, J.C. 2014. TASS 2013 – A Second Step in Reputation Analysis in Spanish. *Procesamiento del Lenguaje Natural* 52: 37-44. Available online: <https://recyt.fecyt.es/index.php/PLN/article/download/29397/15634>.
- WAUGH, L.R. 1995. Reported Speech in Journalistic Discourse: The Relation of Function and Text. *Text* 15 (1): 129-173.
- WEGENER, H. 2000. *Da, denn und weil – der Kampf der Konjunktionen*. Zur Grammatikalisierung im kausalen Bereich. In R. THIEROFF, M. TAMRAT, N. FUHRHOP & O. TEUBER (eds.), *Deutsche Grammatik in Theorie und Praxis*. Berlin – New York: W. de Gruyter: 69-82.
- WHITE, P.R. 2006. Evaluative Semantics and Ideological Positioning in Journalistic Discourse. In I. LASSEN, J. STRUNCK, T. VESTERGAARD (eds.), *Mediating Ideology in Text and Image: Ten Critical Studies*. Amsterdam – Philadelphia: J. Benjamins: 37-67.
- ZUFFEREY, S. 2012. “Car, Parce que, Puisque” Revisited: Three Empirical Studies on French Causal Connectives. *Journal of Pragmatics* 44 (2): 138-153.
- ZWAAN, R.A. & RAPP, D.N. 2006. Discourse Comprehension. In M.J. TRAXLER & M.A. GERNSBACHER (eds.), *Handbook of Psycholinguistics*. Boston: Elsevier: 725-764.

Appendix

List of connectives ¹²

- *A fin de cuentas* (“after all”)
- *Al fin y a la postre* (“after all”)
- *Al fin y al cabo* (“in the end”)
- *Así* (“thus”)
- *Así es que* (“so”)
- *Así pues* (“so”)
- *Así que* (“so”)
- *Como* (“since”)
- *Como consecuencia* (“as a consequence”)
- *Conque* (“so then”)
- *Consecuencia de ello* (“as a consequence of this”)
- *Consecuentemente* (“therefore”)
- *Consiguientemente* (“consequently”)
- *Dado que* (“given that”)
- *De ahí que* (“thus”)
- *De ese modo* (“in that way”)
- *De esta forma* (“in this way”)
- *De esta manera* (“in this way”)
- *De este modo* (“in this way”)
- *De manera que* (“so that”)
- *De modo que* (“so that”)
- *De suerte que* (“so that”)
- *De tal forma que* (“in such a way that”)
- *De tal manera que* (“in such a way that”)
- *Debido a* (“because of”)
- *En consecuencia* (“consequently”)
- *Entonces* (“so”)
- *Por consecuencia* (“consequently”)
- *Por consiguiente* (“therefore”)
- *Por ello* (“thus”)
- *Por ende* (“thus”)
- *Por esa causa* (“for that reason”)
- *Por esa razón* (“for that reason”)
- *Por ese motivo* (“for that reason”)
- *Por eso* (“that’s why”)
- *Por esta causa* (“for this reason”)
- *Por esta razón* (“for this reason”)
- *Por este motivo* (“for this reason”)

12. The translation provided for each marker can be subject to changes depending on the context.

- *Por esto* (“that’s why”)
- *Por lo cual* (“whereby”)
- *Por lo que* (“so that”)
- *Por lo tanto* (“therefore”)
- *Por tal causa* (“for such a reason”)
- *Por tal motivo* (“for such a reason”)
- *Por tal razón* (“for such a reason”)
- *Por tanto* (“therefore”)
- *Por todo ello* (“for all that”)
- *Por todo eso* (“for all that”)
- *Por todo lo anterior* (“for all the above mentioned”)
- *Porque* (“because”)
- *Pues* (“so”)
- *Puesto que* (“given that”)
- *Resultado de lo anterior* (“as a result of the above mentioned”)
- *Visto que* (“given that”)
- *Y* (“and”)
- *Y es que* (“and the thing is”)
- *Ya que* (“since”)