UNIVERSIDAD COMPLUTENSE MADRID

Corpus annotation of causal relations and their signals in English and Spanish

TRABAJO DE FIN DE MÁSTER REALIZADO POR

DÑA. BEATRIZ ALBALADEJO PAY

BAJO LA DIRECCIÓN DE

DRA. MARÍA JULIA LAVID LÓPEZ

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Filología

Curso 2015/2016

Convocatoria de Junio

RESUMEN

Las relaciones causales ha sido un tema de gran consideración durante varios años debido al interés lingüístico de encontrar cuál es su papel en el proceso de expresar causalidad. Sin embargo, hay muchas preguntas sin responder en esta área de investigación. Por ello, este estudio pretende investigar los tipos de relaciones causales más comunes en textos escritos tanto en español como en inglés, y demostrar si hay una destacada conexión entre ambas lenguas.

En primer lugar, teniendo en cuenta el enfoque funcional las relaciones causales se pueden dividir en dos grupos: volitional (acciones volitivas) y non-volitional (acciones no volitivas). Las relaciones causales volitivas (1) ocurren cuando una situación es causada por otra situación proveniente de una acción deliberada de alguien. En el caso de las relaciones causales no volitivas (2), una situación es causada por otra situación, pero sin la acción deliberada de ningún ser humano. Las siguientes frases muestran este tipo de relaciones de causa:

(1) "Pensando que quizás el consejo era simplista, viniendo de una persona soltera, uno de ellos le preguntó: ¿Estás casado?" (Traducido de RST, n.d.)
(2) ¿Recuerdas todas las verduras que derramabas bajo la mesa? Quizás es por ello que Sparkly vivió tanto tiempo (Traducido de RST, n.d.)

En segundo lugar, el enfoque cognitivo se basa en la conjetura de Sweetser (1990), quien propuso una clasificación basada en el dominio de las frases, de modo que las relaciones causales quedaron divididas en tres grupos: content (de contenido o reales), epistemic (epistémicas o lógicas) y por último, speech-act (actos de habla). Las relaciones de contenido (3) se refieren a causas del mundo real. Aquellas que son epistémicas o lógicas (4) están basadas en la observación y la conclusión. Finalmente, las relativas a actos del habla (5) son aquellas entre el acto de habla en sí y la justificación o motivación del hablante de llevar a cabo ese acto. Los siguientes ejemplos ilustran estos tipos de causa:

(3) "John volvió porque la amaba" (Traducido de Sanders and Sweetser, 2009, p.2)
(4) "Los vecinos no están en casa porque las luces están apagadas" (Traducido de Sanders and Sweetser, 2009, p.2)

(5) Ya que eres tan listo, ¿cuándo nació George Washington" (Traducido de Sanders and Sweetser, 2009, p.2)

Asimismo, para realizar una investigación más completa, hemos añadido dos categorías más para la categorización de dichas relaciones. Por un lado, teniendo en cuenta el tipo de señal usada, las relaciones se dividen en siete grupos: conjunciones, preposiciones, sintagmas preposicionales, léxico (a su vez dividido en sintagmas nominales y verbales), oraciones con verbos en forma no personal (infinitivo, gerundio y participio), construcciones sintácticas e it-cleft (oraciones precedidas por el pronombre neutro inglés seguidas de una conjunción causal). Por otro lado, las relaciones de causa se dividirán en: primera posición y segunda posición, teniendo en cuenta si la causa precede el núcleo de la frase o no.

Nuestro estudio parte de la premisa que infiere que la elección de una expresión lingüística concreta para indicar causalidad puede deberse a la existencia de unos patrones establecidos en cada lenguaje. Por ello, el principal objetivo de este estudio es proporcionar evidencia empírica sobre dichos patrones que nos permita en un futuro automatizar la señalización de relaciones causales.

La investigación consiste en el análisis de varios textos escritos en inglés y su correspondiente traducción en español. Estos textos comprenden un total de 36 piezas de escritura extraídas de la plataforma de MULTINOT, un cualitativo, diverso y equilibrado corpus. Para llevar a cabo dicho análisis, hemos seguido tres pasos principales: primero, compilamos nuestro corpus y convertimos nuestros documentos a la codificación 8-UFT para poder procesarlos con la herramienta utilizada para la anotación. Tras ello, fue necesaria la creación de un inventario de señales con expresiones lingüísticas comúnmente conocidas y usadas para denotar relaciones causales en ambos idiomas. También fue imprescindible la creación de un esquema de anotación válido para inglés y español como base para la clasificación de las señales de acuerdo a las categorías y los enfoques explicados previamente: el enfoque cognitivo, el enfoque funcional, el orden de aparición y el tipo de señal. La anotación de cada texto se llevó a cabo gracias a la herramienta *uam corpus tool* (O'Donnell, 2016). Por último, los resultados del análisis fueron extraídos y compilados en documentos Excel para poder ser estadísticamente probados mediante la fórmula Chi-square.

Los resultados de nuestro estudio sugieren que en relación al enfoque funcional, las relaciones causales no volitivas son más comúnmente utilizadas que aquellas que son volitivas, aunque no podemos observar gran diferencia entre ellas. Por otro lado, si tenemos en cuenta el enfoque cognitivo en las relaciones causales, las relaciones de contenido son las más predominantes con diferencia; las relaciones epistémicas son también bastante recurrentes; sin embargo, aquellas referentes a actos del habla son las menos frecuentes, tanto en inglés como en español. De acuerdo al orden de aparición de las señales, podemos notar que las relaciones que ocupan la segunda posición son más recurrentes que aquellas que ocupan primera posición; sin embargo, hay una menor diferencia en su representación. Por último, teniendo en cuenta el tipo de señal utilizada para la expresión de dichas relaciones, los resultados sugieren que las conjunciones son las expresiones lingüísticas más utilizadas para denotar causa, aunque el léxico, sobre todo los sintagmas verbales, también tienen un alto grado de frecuencia.

Asimismo, tras llevar a cabo un análisis descriptivo, los resultados sugieren que los tipos de relaciones de causa en relación al enfoque funcional, el enfoque cognitivo y el orden de apariencia mantienen un comportamiento bastante similar en los textos españoles e ingleses. Igualmente, el tipo de señal utilizada para la representación de cada relación también posee cierta conexión en ambos idiomas, aunque en este caso encontramos mayores diferencias que en las categorías anteriores.

Sin embargo, ya que uno de los principales objetivos de este estudio era encontrar posibles patrones entre ambos idiomas, las relaciones de causa se han analizado teniendo en cuenta la combinación de los dos principales enfoques y las dos categorías añadidas (el enfoque funcional, el enfoque cognitivo, el orden de aparición y el tipo de señal). Los hallazgos demuestran que dichos grupos mantienen un comportamiento similar en ambos idiomas; por consecuencia, las relaciones causales más predominantes en español e inglés son las siguientes: primero, relaciones no volitivas, de contenido que ocupan segunda posición y son representadas con conjunciones; segundo, relaciones volitivas, de contenido que ocupan segunda posición y son representadas con conjunciones; y finalmente, relaciones no volitivas, de contenido que ocupan primera posición y son representadas con léxico.

Además este análisis nos muestra algunas particularidades interesantes en cada idioma. Por ejemplo, se ha descubierto que la expresión lingüística utilizada en inglés para

señalar la relación de causa más común es *because;* sin embargo, en español esta relación es expresada mediante *porque*. Como punto final, nos hemos dado cuenta de que hay cierto tipo de relaciones que están ausentes en ambos idiomas, y normalmente, este tipo de relaciones están asociadas con aquellas que se refieren al acto del habla. Por ejemplo, es imposible encontrar una relación de causa que sea acto de habla y no volitiva al mismo tiempo ni en español ni en inglés.

Por último, es importante destacar que este estudio revela interesantes hallazgos sobre las relaciones causales y el contexto en el que aparecen que permite en un futuro llegar a automatizar la señalización de dichas relaciones. Sin embargo, es necesaria una mayor investigación en esta área ya que este estudio está limitado al análisis de un pequeño número de textos, tanto en español como en inglés (veinte ensayos y dieciséis textos expositivos). Teniendo esto en consideración y percibiendo que la gran mayoría de relaciones eran de contenido debido al género de los textos analizados, los resultados tan solo pueden ser extrapolados al modo de escritura, dejando aparte la manera en la que las relaciones de causa se expresan en el habla del día a día. Finalmente, este trabajo se basa en una selección de textos ingleses y su correspondiente traducción al español; por ello, los resultados de la lengua española pueden estar influenciados por dicha traducción.

Palabras clave: relaciones causales, señales, anotación de corpus, inglés, español

---

ABSTRACT

The study of causal relations and their linguistic expression has been widely studied form different perspectives in recent years. However, few studies have attempted to combine different approaches to the meaning of these relations, and investigated the signals used to express them in a contrastive manner. This Master's thesis is an attempt at advancing knowledge in this area by investigating: a) the possibility to characterize causal relations into different types, using features which combine functional and cognitive approaches; b) the preferred types of causal relations used in original English expository texts and their translations into Spanish; b) the preferred linguistic signals used to express those causal relations in original English texts and their translations into Spanish.

The methodology used for this investigation is based on the manual corpus annotation of a bilingual dataset consisting of a total of thirty-six expository texts (including English original texts and their corresponding translations into Spanish), extracted from the non-fiction part of the MULTINOT corpus, a high-quality, register-diversified and multifunctional bilingual English-Spanish corpus, currently compiled and multidimensionally-annotated by members of the FUNCAP Research Group within the MULTINOT project (see Lavid et al., 2015).[1]

The study was carried out in four major steps: first, an annotation scheme for causal relations in English and Spanish was designed consisting of three interrelated systems and their associated features; second, an inventory of signals for causal relations in English and Spanish, and a categorisation into different types was compiled; third, the annotation scheme was implemented in the UAM Corpus Tool and the bilingual sample was annotated by the author of this study; fourth, the annotated data was statistically analysed to check possible differences between the original English texts and their Spanish translations in the selection of causal relation types and their signals.

The statistical analysis of the annotated data suggests that the preferred types of causal relations in the English original texts are 'non-volitional' and 'content' type, that the preferred order of occurrence is second-position, and the most frequent signal used

[1] I am thankful to Dr. Julia Lavid for giving me permission to use a bilingual sample of texts from the MULTINOT corpus, a research effort of the FUNCAP Research Group, which she coordinates at Universidad Complutense de Madrid.

are ‘conjunctions’, followed by the use of lexical phrases. The analysis of the Spanish translations reveals a high degree of similarity with the English original data, which suggests that the Spanish translations preserve the choices made in the original texts in most of the cases, and that these choices can be considered as indicative of expository texts in the original English texts. Future work will focus on the analysis of original Spanish texts to check whether the observed tendencies in the original English texts and their translations into Spanish are also valid for original Spanish texts, and on the specification of patterns which may help in the automatic analysis of these relations.

Key words: causal relations, signals, corpus annotation, English, Spanish

---

Table of Contents

1. **Introduction……9**
2. **Brief review of approaches to causal relations……...10**
    - **2.1 The concept of causality……11**
    - **2.2 Causal coherence relations……...17**
    - **2.3 Signalling causal coherence relations……..18**
        - **2.3.1 Causal signals in English……...23**
        - **2.3.2 Causal signals in Spanish……..25**
3. **Aims and research questions……27**
4. **Corpus data……27**
5. **Methodology……...28**
    - **5.1 Annotation scheme for causal relations……...29**
    - **5.2 Signals of causal relations…….30**
    - **5.3 Annotation of the bilingual sample……..31**
    - **5.4 Data Analysis description…….32**
    - **5.5 Statistical Analysis of the annotated data……...36**
6. **Results and discussion……...37**
    - **6.1 Comparison of Clausal Types in English and Spanish……..38**
    - **6.2 Comparison of Signals in English and Spanish……..48**
    - **6.3 Descriptive comparison between English and Spanish……..56**
7. **Summary and concluding remarks……..66**
8. **Acknowledgements……69**
9. **References……...71**

List of tables……73

List of figures……...75

Appendix……..77

1. Introduction

Causal relations have been widely studied in linguistics in order to find out their role in the process of human's expression of causality. Work by Maat and Sander's (2001) and Sanders (2005) has focused on the assumption that linguistic categorization reflects human cognition; which means that human's choice of one word over another one to express a certain meaning is considered an act of linguistic categorization. When writers want to express a concrete type of causal relationship, they have a systematic tendency to use one lexical item rather than another, even if both of them are quite similar. These kinds of choices give us an insight on language users' categorizations of causality.

Moreover, written texts are produced under several circumstances which have been shown to have a direct link with the syntactic forms of sentences within discourse. Therefore, according to Maat and Sanders's (2001) view that a central characteristic of discourse is that it shows coherence, "writers have to accommodate their utterances to specific conventions regarded in their language grammar" (Lavid, 1994, p.149).

We can add that recent research has been shown to favour the classification of causal relations due to the fact that the notion of causality is crucial to humans' understanding of the world because it provides us with the necessary knowledge to understand physical processes, actions and how humans organize real-world experiences (Pit, 2003).

While work on causal relations abounds in the literature, there are few studies that specifically address the signalling of coherence relations in general and causal relations in particular. Indeed, there is interesting work on the signalling of coherence relations in English (Taboada and Das 2013) and in Spanish (Da Cunha 2012; Duque 2014), but, to our knowledge, there are no studies which compare the signalling of

causal relations in English and Spanish in a contrastive manner. In this study, therefore, my primary aim is the study of these various signalling mechanisms of causal relations in English and Spanish texts.

The study is structured as follows. Section 2 provides a brief overview of existing approaches to the study of causal relations, focusing on the two main influential ones (the cognitive and the functional-rhetorical). The overview is not exhaustive but focused on the two main strands which have been used for the proposal presented in this study. Section 3 outlines the aims of the study and describes the research questions guiding the investigation; Section 4 describes the bilingual sample used. Section 5 presents the methodology used, consisting of a number of steps, and section 6 reports on the annotation results and discusses the main findings. Finally, section 7 summarises the work reported and provides some concluding remarks and pointers for the future.

2. Brief review of approaches to causal relations

The study of causality and causal relations has attracted a vast amount of research from different perspectives and approaches. This is not surprising, given the centrality of the concept of causality in human cognition, and also its relevance for establishing coherence in discourse.

In this section, I will present the two main approaches to the study of causal relations which have guided the work in this dissertation, namely, the cognitive approach and the functional-rhetorical approach.

The selection of the cognitive approach is based on the assumption that the distinction between using one discourse connectives in a content domain or others is essential to define the meaning of most causal connectives. Concretely, we shall focus on Sweetser

(1990) approach since this is the only author that introduced a three-level approach to differentiate the meaning and uses of connectives. Moreover, the cognitive approach is also essential in the analysis of causal relations because Rhetorical Structure Theory includes the hierarchy of constituents in order to define the coherence relations, which was missed by others authors' approaches.

2.1 The cognitive approach

Human beings tend to categorize everything around them, and they usually do it unconsciously. In people's everyday language use there are some linguistic categories which give us quite interesting insights into the working of human minds. There are several types of linguistic categorization; therefore, a basic knowledge of language in discourse is crucial to any functional account of language. After all, every language user communicates through discourse. A central characteristic of discourse is that it shows coherence. One principal type of coherence is that of causality (Maat and Sanders, 2001). On the contrary, Moeschler (2003) claims that "the issue of causality is not a specific linguistic problem. Although language is a very efficient way to express causal relations, causality is not an inherent property of language. Evolution theory allows us to suppose the emergence of causal reasoning must have been a crucial step of phylogenesis and the construction of causal ties between concepts yielded a fantastic qualitative change in human cognition" (p.277). Similarly to Maat and Sanders, Pit (2003) defends that the notion of causality is essential to humans' understanding of the world. This cognitive basis provides us with the grounding necessary to understand human actions and physical processes and to organize every real-world experience. Without the concept of causality, the real world and humans' actions and emotions would become a complete chaos.

The main idea of the concept of causality is that people can connect discourse segments by a causal relationship of some kind. Although in general terms coherence is considered to be a cognitive phenomenon, and to some extent independent of the exact linguistic realization in the discourse itself, both linguists and psycholinguists tend to assume that connectives have the function of signalling relationships between discourse segments, by means of instructing interlocutors to construct a coherence relation between two clauses (Maat and Sanders, 2001). Moreover, according to Sanders (2005), when dealing with this type of categorization, there are two fundamental discourse principles at work: causality and subjectivity. This author claims that "causality and subjectivity not only account for the system and use of causal relations and their linguistic expressions – connectives (*because, so*) and lexical signals (*as a result, in conclusion*), but also for cognitive complexity of discourse connections in language acquisition and discourse processing" (p.31). The reasoning is that the different linguistic categorizations reflect basic categories in human cognition. When humans make use of language and begin to understand discourse, they connect discourse segments by inferring coherence relations, such as Cause-Consequence and Argument-Claim on the basis of a very limited set of cognitive principles among which we find Causality and Subjectivity.

Sanders (1992, as cited in Pit, 2005) defines, a coherence relation as "an aspect of meaning of two or more discourse segments which cannot be described in terms of the meaning of the segments in isolation. In other words, it is because of this coherence relation that the meaning of two discourse segments is more than the sum of its parts" (p.5). Concretely, in casual relations, the additional aspect of meaning is represented in two different and successive clauses by the causal link between the causing event and the caused event. Therefore, it seems rational that human language provides us with a

wide range of linguistic means to express causality, for instance, causative constructions such as “he made me write the book” or prepositional phrases like “due to the rain, we have to stay home”. Furthermore, coherence relations can be made both implicitly and explicitly by linguistic markers such as connectives (Pit, 2003, p.5).

Two of the most salient markers of causal constructions used by human beings are causal connectives and causative auxiliaries (Sanders and Sweetser, 2009, introduction). Many authors and linguist have studied connectives from a wide variety of perspectives. Some attempts have been made to analyse their effect on the procession of coherence relations in the field of language processing, leading to the conclusion that connectives do speed up the humans’ processing of the text (Pitt, 2003). Moreover, connectives have been studied in relation to the classifications of coherence relations. Lagerwerf (1998) states that “a connective often gives expression to a coherence relation” (p.14). Then, it can be said that linguistic markers of coherence relations help us to understand how relations are cognitively organized in humans’ mind (Pitt, 2003).

Concretely, Sander and Sweetser (2009) assume that when speakers want to express a certain type of causal relationship, they often systematically tend to use one lexical item rather than another, even if the other one is very similar. These kinds of choices may provide us with an insight of language users’ categorizations of causality. This idea comes from the assumption that linguistic categorization reflects human cognition; and choosing one word or one connective over another one to express meaning can be considered an act of linguistic categorization (Sanders, 2005).

According to Zufferey and Cartoni (2012), discourse connectives have been broadly studied in linguistics for the past decades from many theoretical perspectives and different descriptive points of view. They (Zufferey and Cartoni, 2012) defend that there are relevant criteria governing the choice of one causal connective or another. In

fact, in causal relation the two most important factors identified in the literature are the notion of domains of use (1) and the information status of the cause segment (2).

(1) The notion of domains of use: discourse connectives have a common property which gives us the opportunity to use them to relate various kinds of propositional content; and the distinction between being used in a content domain or others has proved to be essential to define the meaning of most of causal connectives (Zufferey and Cartoni, 2012). However, there are some discrepancies among authors regarding connectives' domains of use. For instance, Sweetser (1990) proposes a distinction between content or real-world uses, epistemic uses and speech act uses of connective. Other authors such as van Dijk (1979) and Halliday and Hasan (1976) believe that in many cases, connectives can either be used in a content domain or in the other two domains. Therefore, they only distinguish between semantic and pragmatic uses (van Dijk, 1979) or external and internal uses (Halliday and Hassan, 1976). In sum, the important point drawn from these views is that the separation between the domains of use plays an important role in causal connectives relations.

(2) The information status of the cause segment: this criteria leads us to make a distinction in the status of causal connectives. Zufferey and Bruno (2012) state that "this status can either be 'new', if the speaker considers that the hearer is not aware of the cause, or 'given', if the speaker considers that it is part of the common ground" (p.235). It seems that this criterion is directly related to humans' choice of a given connective over another one, which leads us to relate the notion of information status to another well-known criterion: their prototypical position in a sentence. Some authors such as Altenberg (1984) and Meier (2002) studied the relation between these two criteria and concluded that the information status of the cause and the position of the connective in the sentences usually match.

Although both factors are equally valid to classify causal connectives, most of the authors have focused on the notion of domain of use in order to set a classification of causal relations. The main distinction is based on a two-level approach supported by authors such as van Dijk (1976), Halliday and Hassan (1976) and Sanders et al. (1993) who defend that causal relations can be divided in two groups. The first one comprises semantic coherence relations and connectives associated with content-domain meanings while the second one compasses pragmatic relations that express epistemic meanings.

A similar approach is defended by Mendizábal de la Cruz (1977) who states that in the Spanish language we can find two types of causal relations: 'causa lógica' (logic cause) and 'causa real' (real cause). The first one is related to pragmatic meanings and it is used when the relation of cause makes reference to the reason why the speaker states the main clause, as in example (1). Instead, the second one agrees with the semantic relations defended by the English linguists and it is used when the causal relation makes reference to the cause of something previously stated, as in example (2):

(1) "No hay nadie en casa porque están las persianas bajadas" (There is no one at home because the blinds are closed) (Mendizábal de la Cruz, 1997, p.78).

(2) "Suspendió los exámenes porque no estudió lo suficiente" (He failed the exams because he did not studied enough) (Mendizábal de la Cruz, 1997, p.78).

Nevertheless, Sweetser (1990) introduced a three-level approach to establish differences in the meaning and use of connectives. She differentiates between three domains: the content, the epistemic and the speech act. In order to exemplify these categories, Sanders and Sweetser (2009) propose some sentences in which the use of causal conjunctives like *since* or *because* is illustrated (1):

(1) (taken from Sanders & Sweetser, 2009, p.2)

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

b. The neighbours are not at home because the lights are out.
(i.e. the observation that the lights are out causes the conclusion that the neighbours are away).

c. Since you're so smart, when was George Washington born?
(i.e. the question is presumed to be motivated or enabled by the adressee's claim to superior intelligence).

According to Lagerwerf (1998), in any interpretation in the content domain "the relation refers to real-world causality". In every sentence interpreted in the epistemic domain, "the relation is one of observation and conclusion, rather than a real-world causality". Finally, in all the examples that refer to the speech act domain, "the relation is between a speech act and the speaker's justification or motivation for performing this speech act" (Lagerwerf, 1998, p.16). Moreover, this author concludes that the content domain is a mental and internal ideal of the physical world; the epistemic content is linked to reasoning processes and contains conclusions, premises and internal relations between them; and the speech act domain comprises qualifications, justifications or motivation of speech acts.

The distinction between these three domains can sometimes be difficult due to the degree of similarity between them. Indeed, it is not always so obvious to interpret one connective as epistemic rather than semantic. The main discriminating factor between these two domains is apparently the way in which the real-world connection is established. When the connection between the contents of both propositions is likely to occur, we are dealing with a content interpretation; when that connection between the contents of the propositions is impossible to find, it is an epistemic interpretation (Lagerwerf, 1998).

In short, the domain of use of causal connectives can be very useful when it comes to categorize distinct causal relations. However, it should not be forgotten that although connectives like *because* can be used across the three domains, most of them tend to specialize in one domain. As Sanders (2005) demonstrates, this not only happens in the English language but in most languages. For instance, the English connective *since*, the French connective *puisque* and the German connective *denn* can only be used to express epistemic relations.

2.2 The functional approach: Rhetorical Structure Theory

The Rhetorical Structure Theory (RST), however, proposes a different categorization of causal relations. RST was created in order to offer an explanation of the coherence of texts by describing texts rather than focusing on the processes of creating or reading and understanding them (RST, n.d.). Duque (2014) argues that RST makes a difference between hypotactic relations, also called nuclear relations, and paratactic relations, which it calls multinuclear relations. Besides, what most differentiates RST from other theories is that RST includes the hierarchy of the constituents in order to define the coherence relations. The 'nucleus' is the constituent with the higher hierarchical position, while the 'satellite' is the one with low hierarchical position (Duque, 2014).

Then, according to RST, each text has a wide range of possibilities of structure; that is, it is made up of various sorts of 'building blocks' which occur in the text. These blocks are categorized according to their nuclearity and the relation they express. Causal relations are divided into two groups in Rhetorical Structure Theory: volitional cause and non-volitional cause (RST, n.d.).

- Volitional cause occurs when a situation (nucleus) is caused by another one (satellite) by someone's deliberate action, i.e: "Thinking that perhaps the counsel was simplistic, coming from an unmarried person, one of them asked, Are you married?" (Extracted from RST, n.d.).

  In this first case, S (satellite) could have caused the agent of the volitional action in N (nucleus) to perform that action; without the presentation of S, R (reader) might not regard the action as motivated or know the particular motivation; N is more central to W's purposes in putting forth the N-S combination than S is.

- Non-volitional cause occurs when a situation (nucleus) is caused by another one (satellite), but not by anyone's deliberate action, i.e: "Remember all those vegetables you splipped under the table? Maybe that's why Sparkly lived so long" (extracted from RST, n.d.).

  In this second case, "S, by means other than motivating a volitional action, caused N; without the presentation of S, R might not know the particular cause of the situation; a presentation of N is more central than S to W's purposes in putting forth the N-S combination" (RST, n.d.).

2.3 Signalling causal coherence relations

Mann and Thompson (1988, as cited in Duque, 2014, p.29) argue that signals for any of the relations are ambiguous, so their 'definitions do not rely on morphological or syntactic signals', but 'on functional and semantic judgments'. In addition, Duque (2014) states that even in the cases where there are no signals that allow us to define a relation, we can say that all rhetorical relations are signalled in some way or another. Therefore, this author categorises those signals into distinct groups, which are the following ones: conjunctions and discourse markers (1), anaphors and indirect

coherence relations (anaphors and verbs, anaphors and prepositional phrases, anaphors and nominal phrases) (2), lexicon and coherence relations (3), non-finite verbs (4), and genre structure (5).

(1) Conjunctions and discourse markers: as Duque (2014) states, these are the most obvious signals for causal relations. "Discourse markers belong to different grammatical classes, but the main role of all of them is to guide inferences" (Duque, 2014, p.29). Moreover, Taboada and Das (2013) defend that although the taxonomy of discourse markers has been reduced to single-word conjunctions, we can also find some multi-word expressions that function as discourse markers such as 'in the event that'. Similarly, Fraser (2009) defines discourse markers as a functional class of linguistic elements belonging to different syntactic classes, such as conjunctions, adverbs and prepositional phrases that connect two or more discourse segments and signal a relation between them.

However, we will follow the discourse marker concept defended by Portolés (1998), who insists that discourse markers are restricted in two different ways: they are invariable units morphologically speaking, and syntactically they are isolated units that do not have any function in the clause they introduce. Therefore, prepositional phrases like 'for this reason' cannot be considered a discourse marker since it plays an adverbial role in the clause it introduces and it admits morphological changes such as 'for these reasons'. We shall consider this assumption in order to focus on close criteria for the categorization of causal relations.

(2) Anaphors and indirect coherence relations: "In indirect coherence relations, through an anaphora, the construction of the relation is moved inside a single segment, so that the meaning of the relation is not a surplus caused by the connection between segments"

(Duque, 2014, p.29). These indirect coherence relations can be divided into smaller sub-groups, which are the following:

a. Anaphors and verbs: In this type of coherence relations, the relation is lexicalized in a verb which has an anaphora as one of its arguments; and this anaphora can address a sole entity previously introduced in the other segment of the relation, or it can address the whole segment. I.e. 'Ted didn't stop joking. *This/ He* caused hilarity among his friends' (Duque, 2014, p.30).
b. Anaphors and prepositional phrases: Other types of indirect coherence relations are some prepositional phrases placed at the initial position of a sentence, where the anaphora refers to a previous sentence, such as 'for this reason'. These phrases are considered an open class because any noun or pronoun can work as an anaphora (Duque, 2014).
c. Anaphors and nominal phrases: In this type of indirect relations, the anaphora is part of a nominal phrase, such as 'this result' or 'the main cause'. Hence, this group is restricted since not all nouns characterize for they own the prototypical constituents of the causal relations.

(3) Lexicon and coherence relations: According to Danlos (2001), some coherence relations are built through psychological verbs. Many psychological verbs are usually complemented with causes. Then, "when these causes are introduced in a different sentence from the one with the psychological verb, a causal coherence relation is created between the two sentences" (Duque, 2014, p.30). The relation signals are usually single lexical items considered as cognitive frames that organize information, but as Wellner et al. (2006, as cited in Duque, 2014) state, some coherence relations are held through grouping two cognitive frames provided by two lexical items.

(4) Non-finite verbs: non-finite verbs are frequently used in intra-sentential relations. According to Taboada and Das (2013), tense is one of the most prominent features helping indicate relation, as is the case with some instances of non-finite verbs in circumstance relations. Moreover, Duque (2014) adds that gerunds are commonly used to indicate causal relations.

(5) Genre structure: Genre structure is another kind of signal, mainly inter-sentential. As we are speaking about a signal of coherence relations, it is related to global coherence rather than the local coherence found in the previous cases (Duque, 2014). Accordingly, Taboada and Das (2013) state that "genre helps guide the interpretation of relations when the style of the genre is well known to the reader" (p.258).

Taboada and Das (2013, p.258) propose a similar categorization of signals to the one suggested by Duque (2014), but adding some new groups that were previously ignored such as the syntactic level (6), graphical features (7) and numerical elements (8).

(6) The syntactic level: There are several constructions that help us identify a relation at the syntactic level such as word order, subject-verb inversion or the use of interrogatives.

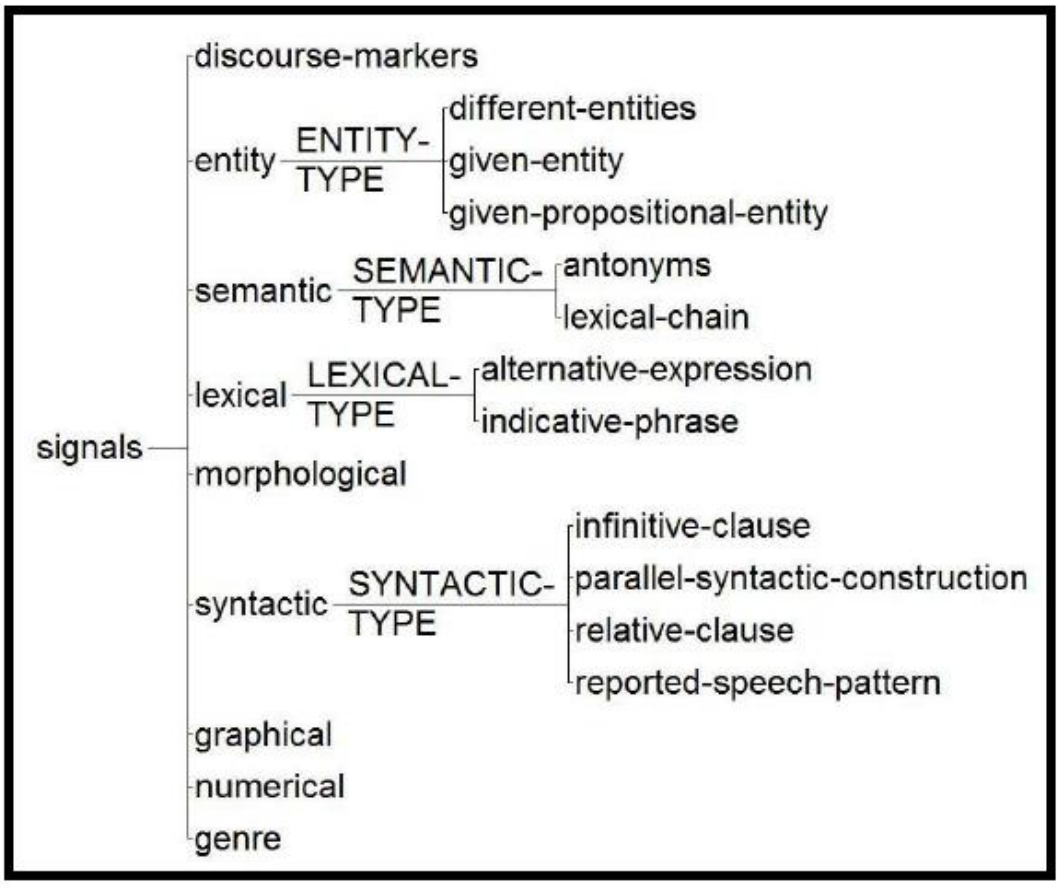
(7) Graphical features: graphical features such as punctuation, lists, headings or layout are also potential indicators of coherence relations.

(8) Numerical elements: Numerical elements can also be considered signals of relations, but they are mainly used in List relations.

These new groups are not considered by Duque (2014) because those types of primary signals are related to prosody and intonation or paralinguistic signalling. Duque defends that:

> "Coherence relations, like many other aspects of meaning, can be conveyed not only by a single signal, but also by clusters. In this sense, other linguistic phenomena, such as tense, aspect, verbal mood, sentence mood, punctuation, and order of segments implied in the coherence relation, participate in signalling some coherence relations; but they are less clear signals of specific coherence relations" (Duque, 2014, p.32).

Taboada and Das (2013) designed a scheme of what they considered the top-level classification of signals (picture 2.3).



Picture 2.3 Classification of signals according to Taboada and Das (2013).

2.3.1 Causal signals in English

In English we find a wide variety of signals to express causal relations. Nevertheless, some of the most used signals were extracted from different authors (Cenoposiciones, 2010; Duque, 2014; Matt and Sanders, 2001; Sanders, 2005; Sanders and Sweetser, 2009), and were grouped according to the previous section (2.3).

Conjunctions: *because, so, since, as, while, therefore*

- I.e. "While you are in the kitchen, bring me another bear" (Cenoposiciones, 2010, p.14). In this case the coherence relation is a blend between a causal and temporal relation.
- I.e. "John came back because he loved her" (Sanders and Sweetser, 2009, p.2).
- I.e. "Since you're so smart, when was George Washington born?" (Sanders and Sweetser, 2009, p.2).

Prepositions: *for, from, out of, through, with(out)*

- I.e: "She cried for joy" (Cenoposiciones, 2010, p.12).
- I.e: "he died from cancer" (Cenoposiciones, 2010, p.12).

Prepositional phrases: *because of, out of, on account of, from, for fear of, now that, by virtue of, in light of, in view of the fact that, owing to, due to, as long as, inasmuch as, as a result, for this reason, on the grounds that.*

- I.e. "As long as you're here, we might as well talk about your last game" (Cenoposiciones, 2010, p.12).
- I.e. "The sun was shining. As a result the temperature rose" (Sanders, 2005, p.32).

Lexicon: *psychological verbs, cause, make, lead, result in, give rise to, bring about, become.*

- I.e. “Marta failed the exam. Miguel got angry” (Duque, 2014, p.31).
- I.e. “High temperatures will make the metal melt” (Cenoposiciones, 2010, p.15).

Non-finite clauses: *past participle, gerund, infinitive*

- I.e. “Miguel was punished for arriving home late” (Duque, 2014, p.31).
- I.e. “Assured of your support, I will make it” (Cenoposiciones, 2010, p.13).

Nominal phrases: *this result, the main cause.*

It-cleft: *it is because*

Syntactic constructions: *that is why, in that, what with, come to think of it, now that I come to think of it.*

- I.e. “The neighbours suddenly left for Paris last Friday. That’s why they are not at home” (Matt and Sanders, 2001, p.248).
- I.e. “The evidence is invalid in that it was obtained through illegal means” (Cenoposiciones, 2010, p.14).
- I.e. “What with one thing and another, I couldn’t sleep last night” (Cenoposiciones, 2010, p.14).

Although all these kinds of signal may express causality in one way or another, it is clear that they cannot always be used interchangeably. As Matt and Sanders (2001) show in the following examples:

> *“*The neighbors suddenly left for Paris last Friday. *As a result/ That’s why/ So* they are not at home” (p.248).
>
> “The lights in the neighbors’ living room are out. *#As a result/ ?That’s why/ So* they are not at home” (p.249).

2.3.2 Causal signals in Spanish

As well as English, in the Spanish language we find several ways of expressing causal relations. Therefore, following Piñero (2001), Bellés (2006) and Mendizábal (1997), some of the most common signals were compiled and categorised in groups.

Conjunctions: *porque, ya que, puesto que, dado que, como*

- I.e. "¿A qué hora dijo Manuel que llegaría? Porque no vamos a estar esperándolo eternamente" (Piñero, 2001, p.155).
- I.e. "Ya que no quieres estudiar, ponte a trabajar" (Mendizábal, 1997, p.80).
- I.e. "si no era hijo era sobrino y en todo caso era yerno, porque se casó con una de sus hijas con una hermana o hermanastra, esto era normal" (Bellés, 2006, p.185).

Prepositions: *por*

- I.e. "No pudo estudiar medicina por la nota de selectividad" (Mendizábal, 1997, p.81).

Prepositional phrases: *a causa de, por culpa de, gracias a, en vista de que, a la vista de que, con motivo de, por lo tanto.*

- I.e. "En vista de que obtuvimos malos resultados en las elecciones disolvimos el partido" (Mendizábal, 1997, p.81).
- I.e. "El salario mínimo ha quedado reducido a una cantidad irrisoria. Por culpa de esto, los empresarios y la administración han acordado elevarlo" (Piñero, 2001, p.157).
- I.e. "de demostrar teóricamente, que b, que b es igual a la correlación entre x e y, por eso os decía, hay una redundancia entre la correlación y la regresión" (Bellés, 2006, p.185).

Lexicon: *causar, llevar a*

Non-finite clauses: *infinitivo, gerundio, participio.*

- I.e. “Por venir tarde, te has quedado sin nada bueno” (Mendizábal, 1997, p.80).
- I.e. “Rechazadas estas propuestas pasemos a otras” (Mendizábal, 1997, p.80).
- I.e. “Comiendo tan poco lo único que lograste fue enfermar” (Mendizábal, 1997, p.80).

Nominal phrases: *la causa principal, este resultado*

Syntactic constructions: *es que, de (lo)+ adjetivo/sustantivo + que*

- I.e. “¿Vienes al cine con nosotros? Es que no puedo, he quedado con mi amiga” (Mendizábal, 1997, p.80).
- I.e. “De lo bueno que es, lo llaman tonto” (Mendizábal, 1997, p.81).

3. Aims and research questions

Both English and Spanish speakers denote actions and physical processes, and organize real world events by using different types of causal relations. However, the choice of certain linguistic expression to indicate causality may be because of some fixed pattern established by each language.

The main aims of this study are the following:

(1) to provide a characterisation of causal relations combining a cognitive and a functional-rhetorical perspective in order to unveil the meanings they convey in English and Spanish texts.

(2) to examine which types of causal relation is preferred in English and Spanish expository texts;

(3) to investigate which are the language-specific signals used to express causal relations in original English texts and their translations into Spanish.

(4) to establish possible correspondence patterns between different types of causal relations and their linguistic signals which can be used in computational contexts such as Machine Translation or Automated Discourse Parsing and Generation.

In view of these aims, the research questions that are investigated in the study are the following:

RQ1: Is it possible to combine a cognitive and a rhetorical perspective in the study of causal relations in English and Spanish? If so, which are the features involved?

RQ2: Is there any preferred type of causal relation in the expository English and the Spanish texts selected for this study? Are there any statistically-significant differences between both languages in the selection of causal relations in expository texts?

RQ3: Which types of signals are preferred for expressing causal relations in each language? Are there statistically-significant differences between English and Spanish in the selection of signals?

4. Corpus data

For this study, a sample corpus of English original texts and their translations into Spanish was extracted from the MULTINOT corpus, a high-quality, register-diversified

and multifunctional bilingual English-Spanish corpus, currently compiled and multidimensionally-annotated within the MULTINOT project (see Lavid et al., 2015).[2]

The bilingual sample consisted of a total of thirty six written essays and expository texts, half of them original English texts and the other half translations into Spanish of those English texts. Table 1 summarizes the features of the bilingual sample collected for this study.

Domain	Sub-register	English	Spanish
Non-fiction	20 Essays	18.570 words	21.046 words
	16 Expository texts	40.668 words	44.160 words

Table 1 Bilingual sample used in this study

The bilingual sample consists on English original texts and their correspondent translation into Spanish instead of original Spanish texts since the focus of our study is on possible translations patterns of English causal relations and their signals into Spanish.

5. Methodology

The study was carried out in four major steps:

1. The first step focused on the creation of an annotation scheme for causal relations in English and Spanish consisting of three interrelated systems and their associated features, as described in 5.1.
2. The second step consisted of the creation of an inventory of signals for causal relations in English and Spanish, and a categorisation into different types, as presented in 5.2.

[2] I am thankful to Dr. Julia Lavid for giving me permission to use a bilingual sample of texts from the MULTINOT corpus, a research effort of the FUNCAP Research Group, which she coordinates at Universidad Complutense de Madrid.

3. The third step was the implementation of the annotation scheme in the UAM Corpus Tool, followed by the annotation of the bilingual sample extracted from the parallel corpus MULTINOT, as described in 5.3.
4. The fourth step focused on the statistical analysis of the annotated data, as described in 5.4.

5.1 Annotation scheme for causal relations

The creation of an annotation scheme is a fundamental step when having to annotate texts, since it reflects the features which the analyst considers to be relevant for his/her corpus study. Given the fact that the current work is a pilot study, I created a preliminary annotation scheme combining features of causal relations which had been proposed in cognitive and functional-rhetorical approaches. [3]

The scheme consists of four simultaneous systems, each with their own features, as shown in Figure 1 below:

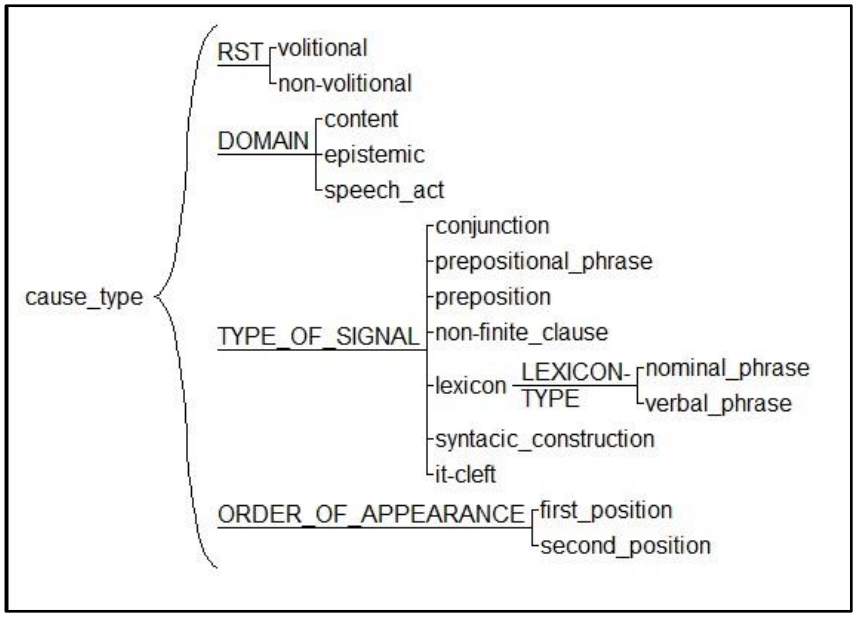


Figure 1: Preliminary annotation scheme for causal relations in English and Spanish

[3] The current annotation scheme is preliminary and has to be validated in future work through an agreement study with two independent annotators.

The systems reflect the paradigmatic options available to English and Spanish when constructing a causal relation.

The RST system captures the division proposed in RST of causal relations into volitional cause and non-volitional cause.

The DOMAIN system captures the classification of causes into: content, epistemic and speech-act.

The TYPE_OF_SIGNAL system captures the potential linguistic realisations of causal relations in English and Spanish.

The ORDER_OF_APPEARANCE system captures the linear order in which a causal relation can occur, i.e. in first position, that is, before the nucleus, or in second position, after the nucleus.

5.2 Signals of causal relations

In order to study the signals which express causal relations in English and Spanish, I consulted different classifications proposed in the literature for English and Spanish,[4] and compiled a contrastive inventory which is graphically displayed in Table 2:

Type of signal		**English**	**Spanish**
Conjunctions		*because, so, since, as, while, therefore*	*porque, ya que, puesto que,*
Prepositions		*for, from, out of, through, with(out)*	*por, de*
Prepositional phrases		*because of, out of, on account of, from, for fear of*	*a causa de, por culpa de, gracias a,*
Lexicon	Verbal phrase	*cause, make, lead*	*causar, llevar a*
	Nominal phrase	*this result, the main cause.*	*la causa principal, este resultado*
Non-finite clause		*past participle, gerund, infinitive*	*infinitivo, gerundio, participio.*
Syntactic constructions		*that is why, in that, what with,*	*es que, de (lo)+ adjetivo/sustantivo +*

[4] For English I consulted Cenoposiciones, 2010; Duque, 2014; Matt and Sanders, 2001; Sanders, 2005; Sanders and Sweetser, 2009. For Spanish I consulted Bellés 2006; Mendizábal 1997; and Piñero 2001.

		que
It-cleft	*It is because, it is why*	-

Table 2: Inventory of English and Spanish signals of causal relations

5.3. Annotation of the bilingual sample

The English and the Spanish texts of the bilingual sample were manually annotated using the UAM Corpus Tool (O'Donnell, 2016). The annotation considered all the options of the annotation schema, as graphically shown in Figure 2 (English text) and in Figure 3 (Spanish text).

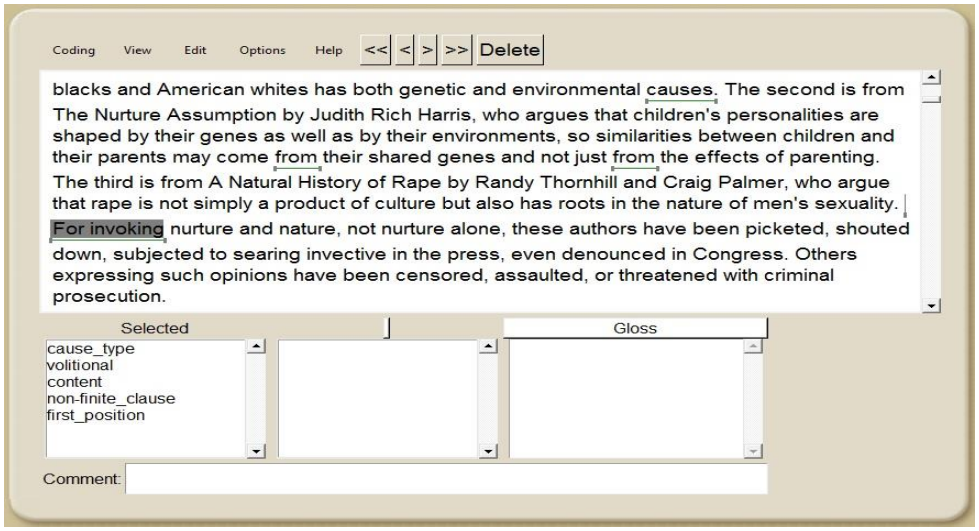


Figure 2: Annotating an English text

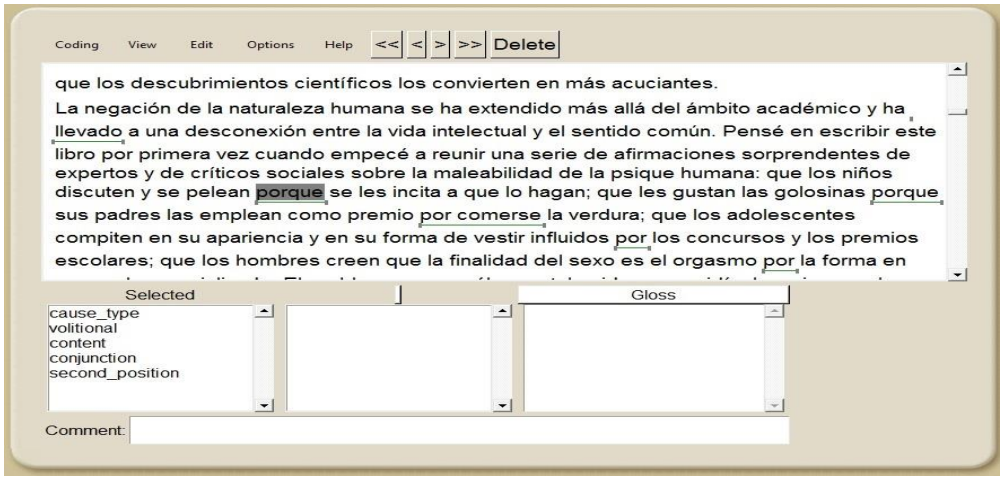


Figure 3: Annotating a Spanish text

5.4. Data analysis description

The data was extracted from the annotation of the thirty six texts. All annotation features were tabulated in Excel documents (added in the appendix). In order to answer the research questions, the analysis was divided into three main sections.

One part of the analysis was carried out in answer to RQ nº 1, which was aimed at combining a cognitive and a rhetorical perspective in the study of causal relations in English and Spanish. For this reason, each essay and expository text was manually annotated according to the sub-categories proposed by the cognitive and the functional approach.

The second part of the analysis was carried out in answer to RQ nº 2, which aims at finding any preferred type of causal relation in the expository English and the Spanish texts. Therefore, each text was analysed descriptively in order to figure out which type of causal relations were more prominent to occur in each language. For this purpose, we have used some statistical measures such as frequency, mean and standard deviation. Furthermore, we shall indicate that we have statistically tested the results by using the Qui-score formula, to have a fairly reliable comparison between both languages.

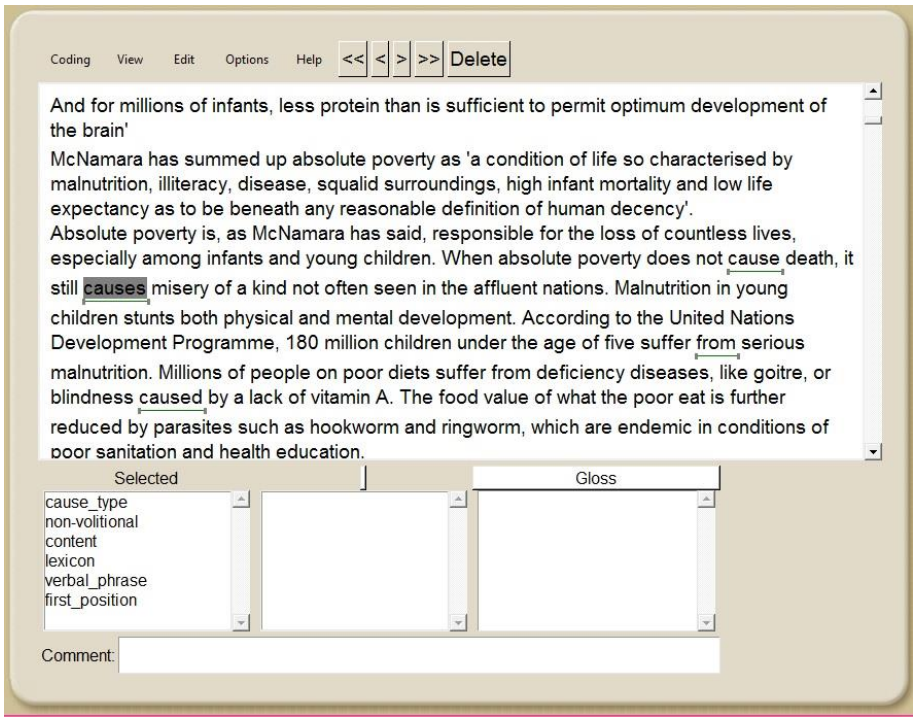
In answer to RQ nº3 which aims at searching for the types of signals preferred for expressing causal relations, we analysed the different texts in depth so as to denote which signals are used in certain causal relations. For this purpose, we also used some statistical measures such as frequency, mean and standard deviation. Moreover, the results were also statistically tested by using the Qui-score formula.

Finally, once the causal relations were annotated and analysed, we compiled a table which contains all the signals appearing in our corpus and indicates their features

so as to determine any potential pattern in their use. Besides, the data obtained in our analysis was compared between both languages in order to determine their similarities and differences that can be useful for future research.

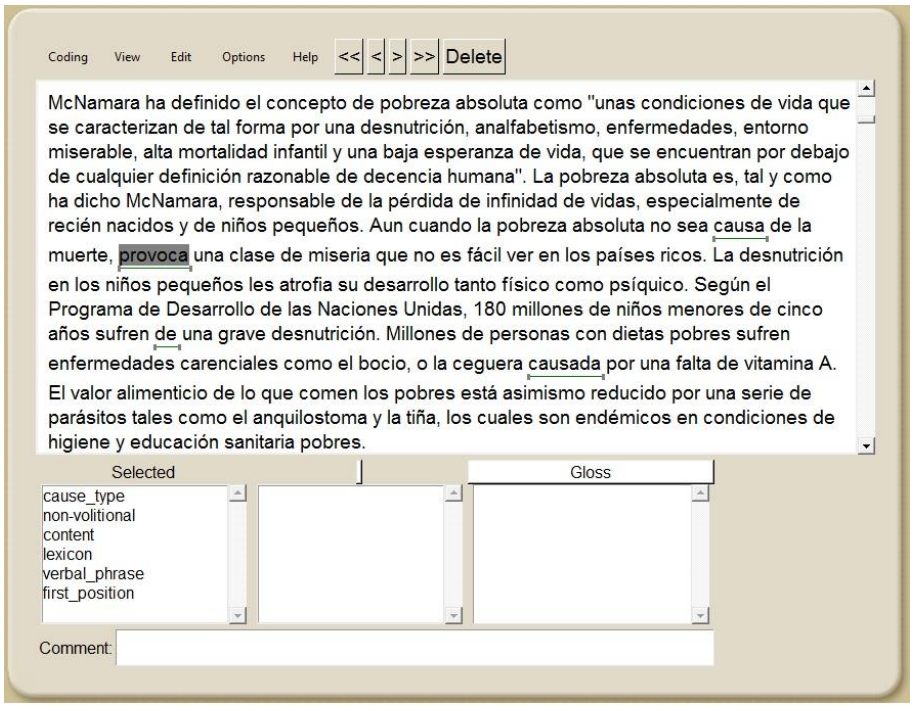
This analysis aimed at measuring the frequency of causal relations that can be observed in English and Spanish. With this observation, we could see whether some signals are more prone to occur in one category or another and we could find possible patterns for their realization.

See below an example of an English and Spanish essay and the annotations that have been made in order to carry out the analysis.



1. **Causes**
    a. Non-volitional
    b. Content
    c. Lexicon: verbal-phrase
    d. First position
2. **From**
    a. volitional
    b. Content
    c. Preposition
    d. Second position

Figure 4: Sample of an English essay annotation



3. **Provoca**
    a. Non-volitional
    b. Content
    c. Lexicon: verbal-phrase
    d. First position
4. **De**
    a. Non-volitional
    b. Content
    c. Preposition
    d. Second

Figure 5: Sample of a Spanish essay annotation

Before continuing with a clear illustration of results, we shall start by summarizing on a table the terminology which has been used in order to name the type of causal relations and its corresponding abbreviation. Linguistics expressions will be categorised according to four categories: the functional approach (RST), the cognitive approach (DO), the type of signal (SG) and the order of appearance (OR). In addition, within those categories, we could find a subdivision in order to focus on certain aspects for a more detailed analysis. Exemplification can be seen in the table below.

Causal type of relation	Sub-category	Description	Abbreviations
The functional approach (RST)	Volitional (V)	An action caused volitionally	RST_V
	Non-Volitional (N-v)	An action caused non-volitionally	RST_N-V

Table 3: Description of causal relations according to the functional approach

Causal type of relation	Sub-category	Description	Abbreviations
The cognitive approach (DO)	Content (C)	Real-world causality	DO_C
	Epistemic (E)	The relation is one of observation and conclusion	DO_E
	Speech-act (S-A)	The relation is between a speech act and the speaker's justification or motivation for performing this speech act	DO_S-A

Table 4: Description of causal relations according to the cognitive approach

Causal type of relation	Sub-category	Description	Abbreviations
Order of appearance (OR)	First Position (1)	A relation where the cause appears before the nucleus	OR_1
	Second Position (2)	A relation where the cause appears after the nucleus	OR_2

Table 5: Description of causal relations according to the order of appearance

Causal type of relation	Sub-category	Description	Abbreviations
Type of signal (SG)	Conjunctions (CJ)	Since, because	SG_CJ
	Prepositions (P)	For, from	SG_P
	Prepositional phrases (PP)	Because of, for fear of	SG_PP
	Lexicon (L)	To cause, the cause	SG_L
	Non-finite clause (NF)	Infinitive, gerund, past-participle	SG_NF
	Syntactic constructions (SC)	That is why	SG_SC
	It-cleft (IT)	It is because	SG_IT

Table 6: Description of causal relations according to the type of signal.

5.5 Statistical Analysis of the annotated data

Once the annotated data were created and saved with the *UAM Corpus Tool*, they were standardised by using an *Excel* document and statistically analysed with the Qui-square test. The procedure was as follows: first, the frequency distribution of the annotated data was divided into two types:

1. Data resulting from the annotation of the type of causal relation in the English and the Spanish texts. This is displayed on table 7 below.

Type of Causal relation	**English dataset**		**Spanish dataset**	
	#	%	#	%
RST_V	109	41,44	103	40,71
RST_N-V	154	58,56	150	59,29
DO_C	226	85,93	215	84,98
DO_E	32	12,17	34	13,44
DO_S-A	5	1,90	4	1,58
OR_1	120	45,63	114	45,06
OR_2	143	54,37	139	54,94

Table 7: Frequency distribution of type of causal relation in bilingual dataset

2. Data resulting from the annotation of the type of signal used to express the different types of causal relations in the bilingual dataset. This is displayed on table 8:

Type of signal	**English dataset**		**Spanish dataset**	
	#	%	#	%
Conjunctions	102	38,78	91	35,97
Prepositions	39	14,83	46	18,18
Prepositional Phrases	11	4,18	13	5,14
Lexicon	53	20,15	55	21,74
Non-finite clause	20	7,60	15	5,93
Syntactic construction	35	13,31	33	13,04
It-cleft	3	1,14	0	0

Table 8: Frequency distribution of signal types in the bilingual dataset

Second, the raw data was analysed with the Chi-Square test to find out whether the frequencies were statistically-significant or just due to chance. The Chi-Square results were also divided into two types. See tables below.

1. Data resulting from the Chi-Square test of the type of causal relation in the English and the Spanish texts. This is displayed on table 9 and table 10 below.

Type of Signal	**English dataset**		**Spanish dataset**	
	Observed	Expected	Observed	Expected
RST_V	41,44	20,54	40,71	20,54
RST_N-V	58,56	29,46	59,29	29,46
DO_C	85,93	42,73	84,98	42,73
DO_E	12,17	6,40	13,44	6,40
DO_S-A	1,9	0,87	1,58	0,87
OR_1	45,63	22,63	45,06	22,63
OR_2	54,37	27,33	54,94	27,33

Table 9: Contingency table of type of causal relation in bilingual dataset

Type of Signal	**English Chi-Square results**	**Spanish Chi-Square results**
RST_V	21,27	19,81
RST_N-V	28,74	30,20
RST Chi-Square total	100,02	
DO_C	43,68	41,78
DO_E	5,20	7,74
DO_S-A	1,22	0,58
DO Chi-Square total	100,20	
OR_1	23,25	22,11
OR_2	26,76	27,90
Chi-Square total	100,01	

Table 10: Chi-Square results of type of causal relation in bilingual dataset

2. Data resulting from the Chi-Square test of the type of signal used to express the different types of causal relations in the bilingual dataset. This is displayed on table 11 and table 12:

Type of Causal relation	English dataset		Spanish dataset	
	Observed	Expected	Observed	Expected
Conjunctions	38,78	18,69	35,97	18,69
Prepositions	14,83	8,25	18,18	8,25
Prepositional Phrases	4,18	2,33	5,14	2,33
Lexicon	20,15	10,47	21,74	10,47
Non-finite clause	7,6	3,38	5,93	3,38
Syntactic construction	13,31	6,59	13,04	6,59
It-cleft	1,14	0,28	0	0,29

Table 11: Contingency table of type of signal in bilingual dataset

Type of Causal relation	English Chi-Square results	Spanish Chi-Square results
Conjunctions	21,61	15,98
Prepositions	5,24	11,94
Prepositional Phrases	1,47	3,39
Lexicon	8,94	12,12
Non-finite clause	5,26	1,92
Syntactic construction	6,86	6,32
It-cleft	2,57	0,29
Qui-Square total	103,90	

Table 12: Chi-Square results of type of causal relation in bilingual dataset

6. Results

In this section the results of the two types of analyses outlined in section 5.3 above are presented: first, the preferred types of clausal relations in English and Spanish (section 6.1), and secondly, the results of the comparison of the preferred signals of those clausal types (section 6.2.).

6.1. Comparison of Clausal Types in English and Spanish

When comparing the RST causal relations preferred by the English original texts and their Spanish translations, we can observe that English tends to express causality more frequently with non-volitional relations (58,56%) rather than volitional ones (41,44%),

as graphically shown in Figure 6. This preference has a great influence on Spanish translations, as graphically shown in Figure 7, where non-volitional relations are more frequent than volitional ones.

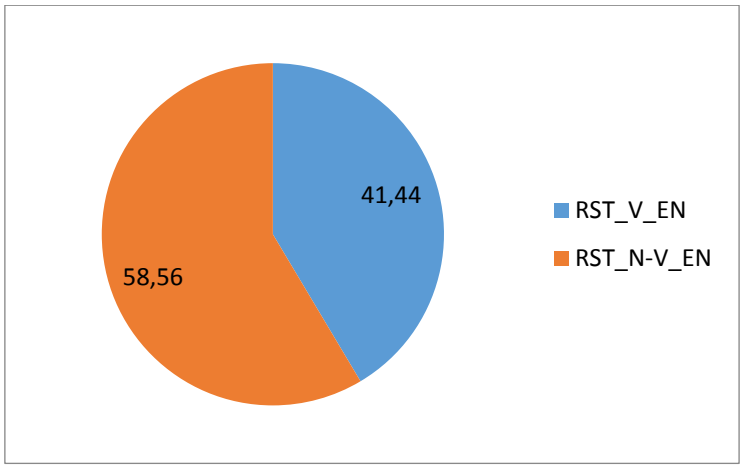


Figure 6: Distribution of RST causal relations in English data set

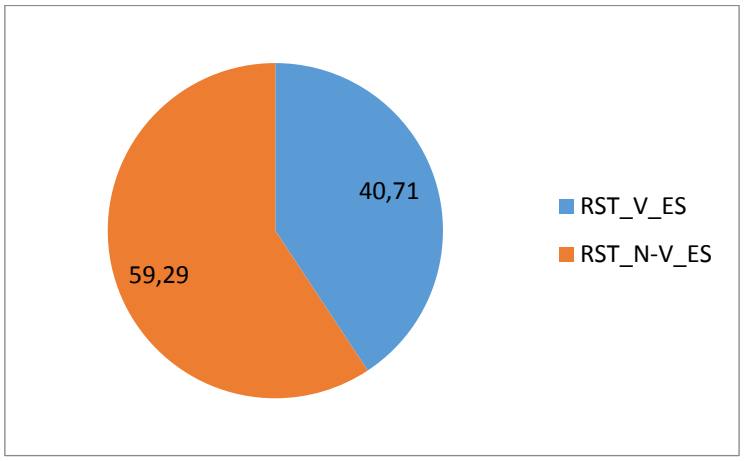


Figure 7: Distribution of RST causal relations in Spanish data set

With respect to the other features of causal relations, it can be observed that in the English dataset DO_C relations are the most frequent (85,93%), followed at a distance by DO_E (12,17%), and by an insignificant amount of , DO_S-A (1,90%)

Similar results can be found in the Spanish translations: DO_C relations are the most frequently used (84,98%), followed at a distance by DO_E relations (13,44%),.

Finally, DO_S-A are again the least preferred type of relation. However, it is interesting to highlight that the frequency of DO_C and DO_S-A relations is higher in the English language, while DO_E are more frequent in the Spanish translations than in the English originals. This means that Spanish translators prefer to express some content or speech-act causal relations by reasoning the purpose of the cause; therefore, epistemic relations are more recurrent in the Spanish texts than in English original ones. Figures 8 and 9 illustrate these results.

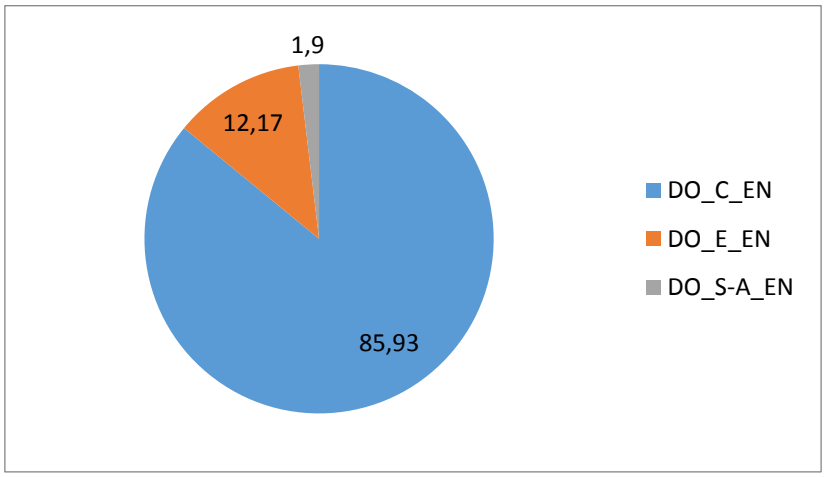


Figure 8: Distribution of causal relations by Domain type in English

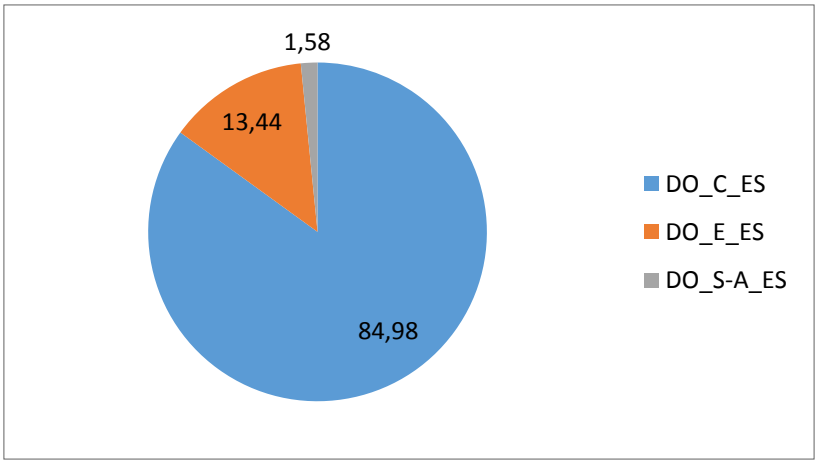


Figure 9: Distribution of causal relations by Domain type in Spanish

As far as the order of appearance is concerned, we can ascertain that OR_2 has the largest degree of relations (54,37%). English original texts, though, present a high tendency to express causality with OR_1 relations (45.63%). Moreover, when comparing both languages we can see almost no difference between them. Spanish translators also have preference for OR_2 relations (54,94%) rather than OR_1 (45,06), and basically, it can be said that they do not change the type of relation when translating from English into Spanish. Figure 10 and 11 summarize these results.

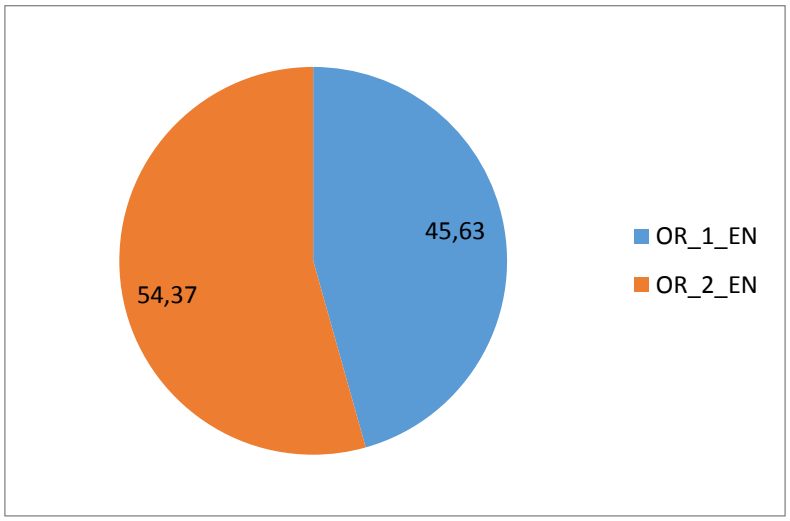


Figure 10: Order of appearance results summary in English

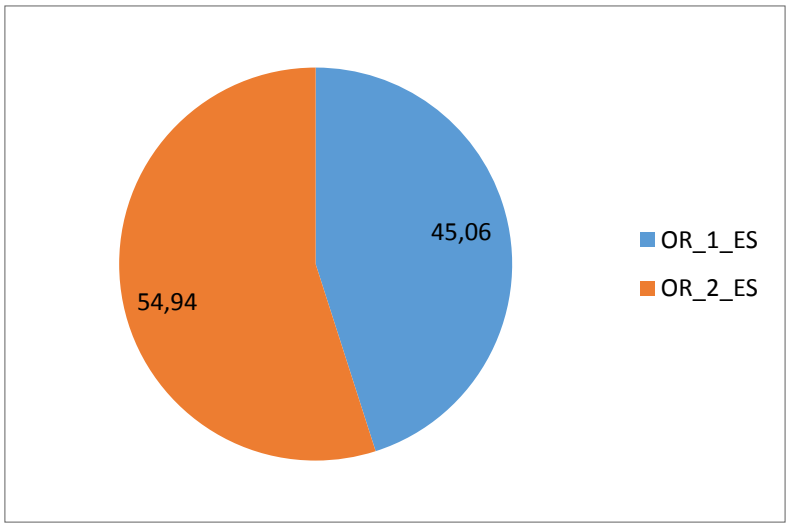


Figure 11: Order of appearance results summary in Spanish

In order to test whether the differences encountered were statistically-significant or due to chance, they were analysed with the Chi-square test.

With respect to the RST volitional or non-volitional relations, the result from the Chi-square test was 100,02. Therefore, taking into account that our degree of freedom in this case is 1 and the *p value* < 0.05, which is the appropriate significance level used in linguistics, we proceeded to check the result according to the Chi-Square distribution table[5]. We have a Chi-Square of 100,02 with *df = 1*, so according to the distribution table, we would need our Chi-value to be equal to or greater than 3,841; which it is. Following with the data from the cognitive approach, the result from this test was 100,20. Thus, taking into account that the degree of freedom in this case is 2 and we maintain the *p value* < 0.05, we proceeded to check the result. We have a Chi-Square of 100,20 with *df = 2*, so according to the distribution table, we would need our Chi-value to be equal to or greater than 7,815; which it also is. Finally, taking into account the order of appearance, the result from this test was 100,01. In this case, taking into account that the degree of freedom is 1 as well as the functional approach, we would also need a Chi-value to be equal to or greater than 3,841; which it is.

This means that the difference found between English and Spanish both in terms of RST relations, the Domain type of relation and the order of appearance is statistically significant at $p < 0.05$. Therefore, we can conclude, with quite a high degree of certainty, that the differences are not due to chance, but due to a true reflection of variation in the use of causal relations in the two languages. Table 13 and figure 12 summarize these results.

[5] See appendix for Chi-Square distribution table

	English dataset	Spanish dataset
Content/Volitional	85	85
Content/Non-Volitional	140	134
Epistemic/Volitional	18	17
Epistemic/Non-volitional	14	11
Speech-act/Volitional	5	4
Speech-act/Non-volitional	0	0

Table 13: Summarised distribution of causal relations types in English and Spanish

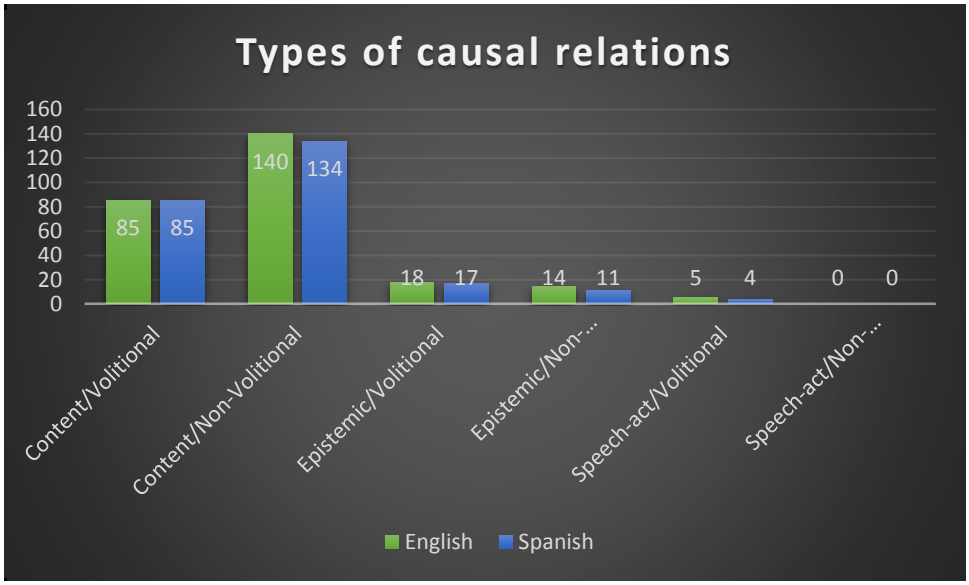


Figure 12: Types of causal relations results in English and Spanish

At first sight, we can infer that English has a higher degree of frequency in most of the types of causal relations, while Spanish translations only equalises English data in content volitional relations. However, in order to make a fair comparison between these languages, these results are standardized according to the number of causal relations signalled in each case. Therefore, as shown in table 14 and figure 13, content volitional relations are more commonly used in the Spanish translations than in the English original texts, while the rest of causal relations still have a higher degree of frequency in English texts.

	English (263)	Spanish (253)
Content/Volitional	32,32	33,60
Content/Non-Volitional	53,23	52,96
Epistemic/Volitional	6,84	6,72
Epistemic/Non-volitional	5,32	4,35
Speech-act/Volitional	1,90	1,58
Speech-act/Non-volitional	0,00	0,00

Table 14: Types of causal relations standardized results in English and Spanish

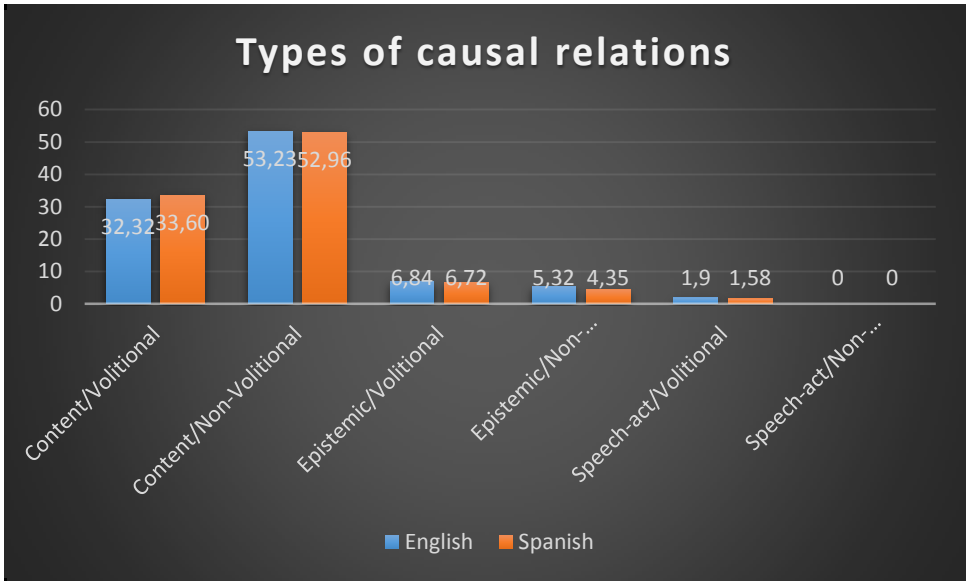


Figure 13: Types of causal relations standardized results in English and Spanish

It seems that there are not great differences between English and Spanish, but we cannot make a conclusion without a significance test. In order to determine if the differences found between both languages are statistically significant or not, we carried out a Chi-square test since it is by far the most common type of significance test used in linguistics. The result from this test was 99.76. Therefore, taking into account that our degree of freedom is 5 and the *p value* < 0.05, which is the appropriate significance level, we proceeded to check the result according to the chi-square distribution table. We have a Chi-value of 99.76 with *df = 5*, so according to the distribution table, we would need our chi-value to be equal to or greater than 11.07, which it is. It means that the difference found between English and Spanish regarding the use of causal relations is statistically significant at $p < 0.05$, and we can therefore, with quite a high degree of

certainty, conclude that the differences are not due to chance, but due to a true reflection of variation in the use of causal relations in the two languages.

To sum up, the data analysis has revealed statistically significant differences are between the English and the Spanish, dataset although they seem to have a great percentage of similarity. Furthermore, our data show that a significant number of causal relations are related to content non-volitional and volitional relations, whose results are quite close (53.23% and 32.32% in English; 52.96% and 33.60% in Spanish respectively). In addition, epistemic volitional relations are also reveling (6.84% and 6.72%). Similarly, the least common types of causal relations are those concerning speech-act relations.

In what follows, we shall comment and discuss each causal category related to English and Spanish texts in more detail.

With respect to RST volitional and non-volitional causal relations, roughly half of the relations are non-volitional and the other half volitional. It should be mentioned that in Spanish translations non-volitional relations have a higher frequency (59.29%) than English original texts (58.56%). On the contrary, volitional relations tend to be more common among English writers (41.44%) than Spanish translators (40.71%). We shall illustrate it with some examples:

"*Por desgracia, no hay ningún lugar al que retirarse, **porque** no hay ningún lugar fuera de la singularidad."* (STrans_EXPE_001)

*"It was named Pluto at least partly **because** the first two letters made a monogram from Lowell's initials."* (EO_EXPE_001)

The first example, extracted from a Spanish expository text, is a clear example of a non-volitional action since there is no one that volitionally carries out an action provoked by the cause. On the contrary, in the English expository text we can notice that someone volitionally named Pluto because of some reasons.

Taking into account the cognitive approach, we can infer that content relations are by far the more prone to occur in both languages (85.93% in English and 84.98% in Spanish). Moreover, epistemic relations are the second most common to be found (12.1% in English and 13.44% in Spanish), while there is almost no manifestation of speech-act relations since they have a very low degree of frequency (1.90% and 1.58%). Finally, it can be noticed that content and speech-act relations are more frequent in English original texts; but, when we deal with Spanish translations, epistemic relations tend to be more used in this language than in English texts. The examples below let us visualize these features:

"If these are the facts, we cannot avoid concluding that ***by not giving*** *more than we do, people in rich countries are allowing those in poor countries to suffer from absolute poverty, with consequent malnutrition, ill health, and death."* (EO_ESSAY_010)

"Finalmente, podría ser que el médico fuera personalmente responsable de la muerte del niño que decide no operar, ***pues que*** *quizá sepa que si no hubiera aceptado este caso, otro médico del hospital lo habría operado."* (STrans_ESSAY_010)

*"****In face of*** *such overpowering numbers, what is the likelihood that only one ordinary star, the Sun, is accompanied by an inhabited planet?"* (EO_EXPE_002)

On one hand, the first example deals with a content relation since it describes a real-world situation. On the other hand, in the second sentence we can notice that the nucleus describes a hypothetical situation derived from someone's reasoning (*podría*

ser); therefore, it is an epistemic relation. The last example is a clear realization of a speech-act relation since the cause leads to the formulation of an interrogative clause.

To end with this detailed analysis, we shall look at the order of appearance of causal relations. First of all, it is important to remind that first position relations occurs when the cause appears before the nucleus. On the contrary, when we find the cause after the nucleus we are dealing with a second position relation. English writers and Spanish translators are more prone to use second position relations when expressing causal relations (54.37% and 54.94% respectively). However, there is not a significant difference with first position relations (45.63% and 45.06% respectively). That is, we could say that causal relation can be expressed either in first position or second position indistinctly both in English and Spanish. Some instances that reflect these results are found:

"Uno de sus envidiosos contemporáneos le apodó Beta, la segunda letra del alfabeto griego, ***porque*** *según decía Eratóstenes era en todo el segundo mejor del mundo."* (STrans_EXPE_002)

"But, ***given*** *limited interregional transfers and labor mobility, this means that the continent has far less ability to absorb disparate shocks through the operation of so-called automatic stabilizers."* (EO_ESSAY_005)

In the Spanish example we deal with a second position relation since the nucleus, which is *'le apodó'* appears before the cause signaled by *'porque'*. The second examples, instead, presents a first position relation where the cause is realized before the nucleus.

6.2. Comparison of Signals in English and Spanish

The type of signals is a relevant factor in the selection of causal relations. In relation to the type of signals through which causal relations are expressed we find larger differences.

On the one hand, the most frequent type of signal in English and Spanish is a conjunction, SG_C, with a 38.78% and a 35.97% respectively. They are followed by causal relations expressed through lexicon, SG_L, (20.15% and 21.74%) and prepositions, SG_P, (14.83% and 18.18%). On the other hand, the type of signal with the lower degree of realization is it cleft, SG_IT, with only 1.14% in English and 0% in Spanish. Besides, we can identify some types of signals which are more used in English than Spanish such as: conjunctions, non-finite clauses, syntactic constructions and it-cleft clauses. Nevertheless, prepositions, prepositional phrases and lexicon have a higher degree of frequency in the Spanish translations than in the English original texts. The figure below is an exemplification of these results.

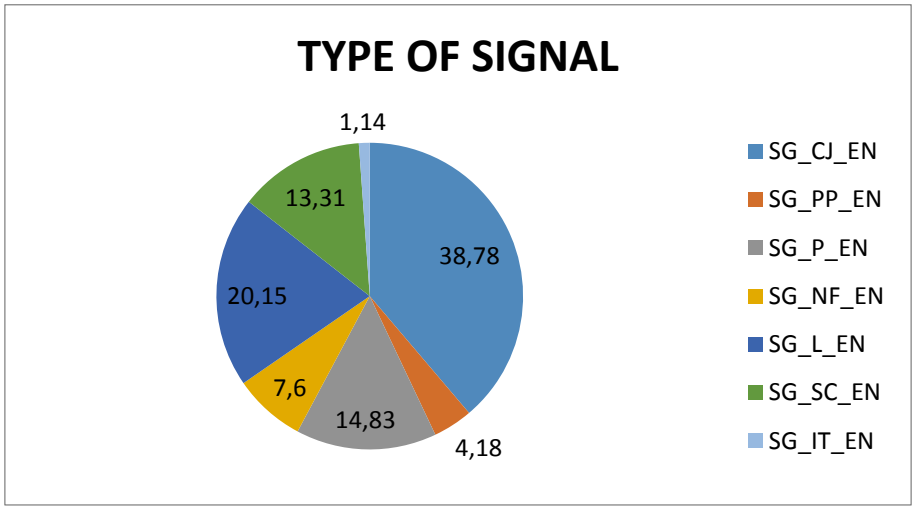


Figure 14: Type of signal results summary in English

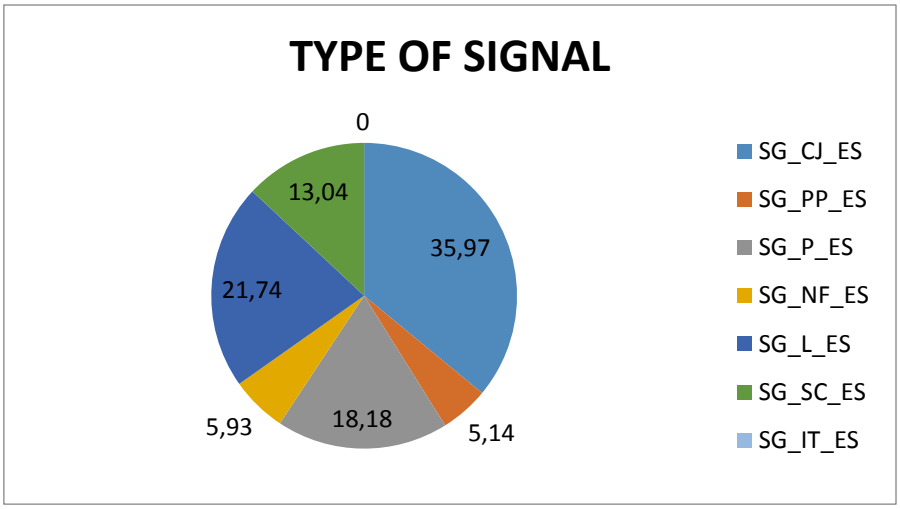


Figure 15: Type of signal results summary in Spanish

In order to carry on with the explanation of results, we shall focus on the Qui-Square test so as to determine whether the difference between English and Spanish texts in the selection of signals is statistically significant or not.

Basing on the data shown above on section 5.5, the result from this test was 103,90. Therefore, taking into account that our degree of freedom is 6 and the *p value* < 0.05, which is the appropriate significance level used in linguistics, we proceeded to check the result according to the Chi-square distribution table. We have a Chi-value of 100,01 with *df = 6*, so according to the distribution table, we would need our chi-value to be equal to or greater than 12,592; which it is. This means that the difference found between English and Spanish regarding the use of signals in causal relations is statistically significant at $p < 0.05$, and we can therefore, with quite a high degree of certainty, conclude that the differences are not due to chance, but due to a true reflection of variation in the use of causal relations in the two languages.

To continue with a more detailed analysis, we will now be looking at all types of signals used in the realization of the different types of causal relations in order to determine which ones are more frequently in each language when the features of the two approaches, the cognitive and the functional one, come into play. In this part of the analysis, the order of appearance will be disregarded because it has been shown that there is almost no different between English texts and the correspondent Spanish translations.

We can see in table 15 and table 16 that the most common relations expressed through conjunctions are content non-volitional (51 and 43 times), followed by content volitional, which also have a high degree of frequency (34 and 35 times). On the contrary, speech-act relations are rare to be found in neither language.

CONJUNCTIONS	Volitional	Non-volitional
content	34	51
epistemic	8	4
Speech-act	3	0

Table 15: Causal relations with conjunctions in English

CONJUNCTIONS	Volitional	Non-volitional
content	35	43
epistemic	10	2
Speech-act	3	0

Table 16: Causal relations with conjunctions in Spanish

As far as prepositions are concerned, the most common relation is content non-volitional (24 and 28 times), followed by content volitional (13 times in both cases). The rest of relations are not much used in either language. However, it is interesting to highlight that Spanish translators use prepositions when they want to express a speech-act volitional relation, while in English original texts there is no manifestation of any speech-act relation, regardless if it is volitional or non-volitional. See tables below.

PREPOSITIONS	Volitional	Non-volitional
content	13	24
epistemic	1	2
Speech-act	0	0

Table 17: Causal relations with prepositions in English

PREPOSITIONS	Volitional	Non-volitional
content	13	28
epistemic	3	1
Speech-act	1	0

Table 18: Causal relations with prepositions in Spanish

Regarding to prepositional phrases, content non-volitional relations continue to have the higher degree of frequency. However, as the following tables show, it is interesting to highlight that in English there no realization of epistemic volitional and speech-act non-volitional relations, while the ones that are absence in Spanish are epistemic volitional and speech-act volitional and non-volitional.

PREPOSITIONAL PHRASES	Volitional	Non-volitional
content	4	5
epistemic	0	1
Speech-act	1	0

Table 19: Causal relations with prepositional phrases in English

PREPOSITIONAL PHRASES	Volitional	Non-volitional
content	4	8
epistemic	0	1
Speech-act	0	0

Table 20: Causal relations with prepositional phrases in Spanish

Taking into account how lexicon is used to express causal relations (table 21 and table 22), we see that there is an important difference between the typical causal relation, which is content non-volitional (37 times in each case), and the rest of them. Furthermore, we can notice that there is no manifestation of speech-act relations within this category.

LEXICON	Volitional	Non-volitional
content	11	37
epistemic	4	1
Speech-act	0	0

Table 21: Causal relations with lexicon in English

LEXICON	Volitional	Non-volitional
content	9	37
epistemic	3	2
Speech-act	0	0

Table 22: Causal relations with lexicon in Spanish

We can appreciate in table 23 and table 24 that non-finite clauses are more prone to occur in content volitional relations (9 and 10 times). Nevertheless, it should be pointed out that in English content non-volitional relations are also quite used. In Spanish instead, none of the relations, with the exception of content volitional are highly manifested by translators.

NON-FINITE CLAUSES	Volitional	Non-volitional
content	9	4
epistemic	3	3
Speech-act	1	0

Table 23: Causal relations with non-finite clauses in English

NON-FINITE CLAUSES	Volitional	Non-volitional
content	10	2
epistemic	1	2
Speech-act	0	0

Table 24: Causal relations with non-finite clauses in Spanish

With regards to syntactic constructions, both content non-volitional and volitional relations are commonly used in English and Spanish. Contrarily, speech-act volitional and non-volitional relations are absent, and epistemic ones have a low degree of frequency. The following tables exemplify these results.

SYNTACTIC CONSTRUCTIONS	Volitional	Non-volitional
content	12	18
epistemic	2	3
Speech-act	0	0

Table 25: Causal relations with syntactic constructions in English

SYNTACTIC CONSTRUCTIONS	Volitional	Non-volitional
content	14	16
epistemic	0	3
Speech-act	0	0

Table 26: Causal relations with syntactic constructions in Spanish

Finally, the last type of signals, it-cleft, can only be found in English since this type of linguistic expression is not possible in Spanish. However, this case is not very prone to occur even in English. We can only find some content volitional and non-volitional expressions. As we can see in the table below, the rest of them have no realization at all.

IT-CLEFT	Volitional	Non-volitional
content	2	1
epistemic	0	0
Speech-act	0	0

Table 27: Causal relations with it-clefts in English

Generally, we could say that the most typical types of causal relations are non-volitional content expressed through conjunctions, followed by non-volitional content relations that take place with lexicon features and volitional content ones which also are manifested with conjunctions. Besides, prepositions are also commonly used by English writers and Spanish translators when expressing non-volitional relations, as well as syntactic constructions. On the contrary speech-act relations are the least frequent ones, precisely, they only occur with volitional relations expressed through non-finite clauses and prepositional phrases in English, prepositions in Spanish and conjunctions in both languages.

In conclusion, the data analysis has revealed that the differences found in English original texts when translating into Spanish are statistically significant although it seems that these languages have a high degree of similarity. Besides our data show that a significant number of causal relations are expressed through conjunctions no matter if we are dealing with an English original text or a Spanish translation.

To follow with our analysis, we shall comment and discuss in depth which linguistic expressions are more recurrent to express the most common type of signals in each language.

As far as the type of signal is concerned, conjunctions are the linguistics expressions most used to express causal relations, with a 38.78% degree of frequency in English and 35.97% in Spanish. Similarly, causal relations apropos of lexicon and prepositions are also prone to occur in both languages (20.15% and 14.83% in English, and 21.74% and 18.18% in Spanish). In addition, it is important to point out that regarding lexicon, most of realizations tend to be expressed with verbal phrases (18.25% and 17%) rather than nominal phrases (1.90% and 4.74%). It-clefts and prepositional phrases are, however, the least prominent types of causal relations (1.14% and 4.18% in English, and 0.00% and 5.93%). Concretely, it can be seen that in Spanish causal relations may not be expressed through an it-cleft clause since this type of linguistics expression is not possible in this language.

Moreover, we should specify which linguistic expressions are more prone to be used with the types of signals which have a higher frequency of realization. In English texts we notice that conjunctions are the most common signals and the typical linguistic expression in this case in *because.* Although we can find some variations with expressions such as *for* or *since*, *because* is by far the one preferred by English writers. Spanish translations also have a significant number of causal relations expressed through conjunctions; in this case, the linguistic expression *porque* is the most salient. Nevertheless, we can find some other expressions such as *pues*, *puesto* que and *ya que* that also are quite recurrent.

With regards to lexicon signals, we mentioned above that verbal phrases are more prominent than nominal phrases in both English and Spanish. English writers have

a clear tendency to use two main linguistic expressions: *to cause* and *to lead*, although the expression *to make* is quite frequent as well. Instead, their correspondent translations in Spanish have a wider variety of representation. In Spanish there several verbal phrases that are preferred by translators in order to express the causal relations appearing in English original texts. These expressions are the following one: *causar, llevar, hacer, conducir* and *provocar.*

Finally, as far as prepositions are concerned, there is a great tendency to use prepositions such as *for* or *from* in English original texts. When it comes the translation into Spanish, we can find signals such as *por* or *de* as the main choice; which means that there is a high level of similarity between the original texts and the translations. However, it is interesting to point out that both English writers and Spanish translators have a rich and wide variety of preposition choices. In order to illustrate the data we show the following examples:

"He believed this ***because*** *he felt, for mystical reasons, that the earth was the center of the universe and that circular motion was the most perfect."* (EO_EXPE_007)

"Subrayo este punto ***pues*** *sé que estoy en peligro de ser mal interpretado por aquellas personas, demasiado numerosas, que no pueden distinguir una declaración que denote convencimiento de una defensa de lo que debería ser."* (STrans_EXPE_003)

"According to one study, 14 million children under five die every year ***from*** *the combined effects of malnutrition and infection."* (EO_ESSAY_010)

"Aun cuando la pobreza absoluta no sea ***causa*** *de la muerte,* ***provoca*** *una clase de miseria que no es fácil ver en los países ricos."* (STrans_ESSAY_010)

"Pero, ***debido a*** *los límites a las trasferencias interregionales y la movilidad laboral, esto significa que hay mucho menos capacidad de absorber crisis dispares mediante los llamados estabilizadores automáticos."* (STrans_ESSAY_005)

*"****It is because*** *America has embraced a policy agenda in recent decades that* ***has caused*** *its economy to become wildly unequal, leaving the most vulnerable segments of society further and further behind."* (EO_ESSAY_003)

The two first sentences are clear examples of causal relations realized with conjunctions, both in English and Spanish. They are followed by examples with prepositions and lexicon. In addition, the example extracted from STrans_ESSAY_010 illustrates the use of nominal phrases and verbal phrases within lexicon signals. Finally, the two last example deals with the least frequent types of signals, prepositional phrases and it-cleft clauses. Interestingly, in the example extracted from EO_ESSAY_003, the relation expressed through an it-cleft clause is strengthened with the use of a lexicon (*has caused*).

6.3 Descriptive comparison between English and Spanish

A descriptive analysis whose results are shown in the following table was also carried out. In this analysis, we account for all possible causal relations taking into account all the different features included in our analysis.

'N' corresponds to the total number of texts, which in this case were eighteen for English original texts and eighteen for Spanish translations; 'sum' is the total number of times that the type of causal relation appears within all the corpus; 'mean' is the sum of the causal relation divided into the number of texts (eighteen in each case); the 'standard deviation' concerns how the causal relation deviates from the mean; the 'variance' is the

squared standard deviation; and the ‘coefficient of variance’ corresponds to the division of the standard deviation and the mean in order to calculate the most common types of causal relations.

On account of this, we can see the number of time that causal relations occur in total in each language, the mean of all values regarding the causal relations and the number of texts analysed and how all the causal relations deviate from the mean. See table 28 and table 29.

Causal relations will be listed according to their frequency, once the coefficient of variance has been calculated: We can find some frequent causal relations such as N-V_C_2_CJ (non-volitional, content, second position relations expressed through conjunctions), V_C_2_CJ (volitional, content, second position relations expressed through conjunctions), and N-V_C_1_L (non-volitional, content, first position relations expressed through lexicon) in English; and N-V_C_2_CJ (non-volitional, content, second position relations expressed through conjunctions), V_C_2_CJ (volitional, content, second position relations expressed through conjunctions), N-V_C_1_L (non-volitional, content, first position relations expressed through lexicon) and N-V_C_2_P (non-volitional, content, second position relations expressed through prepositions) in Spanish. Contrarily, some less frequent causal relations are: V_E_1_SC (volitional, content, first position relations expressed through syntactic constructions, V_S-A_1_PP (volitional, speech-act, first position relations expressed through prepositional phrases), V_E_2_P, V_E_2_NF, V_E_2_SC (volitional, epistemic, second position relations expressed through prepositions, non-finite clauses and syntactic constructions), N-V_C_1_PP, N-V_C_1_IT (non-volitional, content, first position relations expressed through prepositional phrases and it-clefts), N-V_E_1_P, N-V_E_1_PP and N-V_E_1_L (non-volitional, epistemic, first position relations expressed through

prepositions, prepositional phrases and lexicon) in English; and V_C_1_PP (volitional, content, first position relations expressed through prepositions), V_E_1_NF (volitional, epistemic, first position relations expressed through non-finite clauses), V_S-A_1_P (volitional, speech-act, first position relations expressed through prepositions), V_E_2_P (volitional, epistemic, second position relations expressed through prepositions), N-V_E_1_CJ, N-V_E_1_PP (non-volitional, epistemic, first position relations expressed through conjunctions and prepositional phrases), N-V_E_2_CJ, N-V_E_2_P and N-V_E_2_NF (non-volitional, epistemic, second position relations expressed through conjunctions, prepositions and non-finite clauses) in Spanish. Moreover, it is important to highlight that there some types of causal relations that are not possible either in English or Spanish when all the features of the different approaches and categories come into play.

Type of Causal relation	N	Sum	Mean	Standard Deviation	Variance	Coefficient of Variance
V_C_1_CJ	18	3	0,167	0,397	0,157	2,380
V_C_1_P	18	9	0,500	0,687	0,472	1,374
V_C_1_PP	18	0	0,000	0,000	0,000	0
V_C_1_L	18	11	0,611	0,760	0,577	1,243
V_C_1_NF	18	7	0,389	0,606	0,367	1,558
V_C_1_SC	18	7	0,389	0,606	0,367	1,558
V_C_1_IT	18	2	0,111	0,324	0,105	2,915
V_E_1_CJ	18	4	0,222	0,458	0,210	2,062
V_E_1_P	18	0	0,000	0,000	0,000	0
V_E_1_PP	18	0	0,000	0,000	0,000	0
V_E_1_L	18	4	0,222	0,458	0,210	2,062
V_E_1_NF	18	2	0,111	0,324	0,105	2,915
V_E_1_SC	18	1	0,056	0,229	0,052	4,123
V_E_1_IT	18	0	0,000	0,000	0,000	0
V_S-A_1_CJ	18	3	0,167	0,397	0,157	2,380
V_S-A_1_P	18	0	0,000	0,000	0,000	0
V_S-A_1_PP	18	1	0,056	0,229	0,052	4,123
V_S-A_1_L	18	0	0,000	0,000	0,000	0
V_S-A_1_NF	18	1	0,056	0,229	0,052	4,123
V_S-A_1_SC	18	0	0,000	0,000	0,000	0
V_S-A_1_IT	18	0	0,000	0,000	0,000	0
V_C_2_CJ	18	31	1,722	1,275	1,627	0,741
V_C_2_P	18	4	0,222	0,458	0,210	2,062
V_C_2_PP	18	4	0,222	0,458	0,210	2,062
V_C_2_L	18	0	0,000	0,000	0,000	0
V_C_2_NF	18	2	0,111	0,324	0,105	2,915
V_C_2_SC	18	5	0,278	0,512	0,262	1,844
V_C_2_IT	18	0	0,000	0,000	0,000	0

V_E_2_CJ	18	4	0,222	0,458	0,210	2,062
V_E_2_P	18	1	0,056	0,229	0,052	4,123
V_E_2_PP	18	0	0,000	0,000	0,000	0
V_E_2_L	18	0	0,000	0,000	0,000	0
V_E_2_NF	18	1	0,056	0,229	0,052	4,123
V_E_2_SC	18	1	0,056	0,229	0,052	4,123
V_E_2_IT	18	0	0,000	0,000	0,000	0
V_S-A_2_CJ	18	0	0,000	0,000	0,000	0
V_S-A_2_P	18	0	0,000	0,000	0,000	0
V_S-A_2_PP	18	0	0,000	0,000	0,000	0
V_S-A_2_L	18	0	0,000	0,000	0,000	0
V_S-A_2_NF	18	0	0,000	0,000	0,000	0
V_S-A_2_SC	18	0	0,000	0,000	0,000	0
V_S-A_2_IT	18	0	0,000	0,000	0,000	0
N-V_C_1_CJ	18	11	0,611	0,760	0,577	1,243
N-V_C_1_P	18	8	0,444	0,648	0,420	1,458
N-V_C_1_PP	18	1	0,056	0,229	0,052	4,123
N-V_C_1_L	18	28	1,556	1,212	1,469	0,779
N-V_C_1_NF	18	3	0,167	0,397	0,157	2,380
N-V_C_1_SC	18	4	0,222	0,458	0,210	2,062
N-V_C_1_IT	18	1	0,056	0,229	0,052	4,123
N-V_E_1_CJ	18	2	0,111	0,324	0,105	2,915
N-V_E_1_P	18	1	0,056	0,229	0,052	4,123
N-V_E_1_PP	18	1	0,056	0,229	0,052	4,123
N-V_E_1_L	18	1	0,056	0,229	0,052	4,123
N-V_E_1_NF	18	2	0,111	0,324	0,105	2,915
N-V_E_1_SC	18	2	0,111	0,324	0,105	2,915
N-V_E_1_IT	18	0	0,000	0,000	0,000	0
N-V_S-A_1_CJ	18	0	0,000	0,000	0,000	0
N-V_S-A_1_P	18	0	0,000	0,000	0,000	0
N-V_S-A_1_PP	18	0	0,000	0,000	0,000	0
N-V_S-A_1_L	18	0	0,000	0,000	0,000	0
N-V_S-A_1_NF	18	0	0,000	0,000	0,000	0
N-V_S-A_1_SC	18	0	0,000	0,000	0,000	0
N-V_S-A_1_IT	18	0	0,000	0,000	0,000	0
N-V_C_2_CJ	18	40	2,222	1,449	2,099	0,652
N-V_C_2_P	18	16	0,889	0,916	0,840	1,031
N-V_C_2_PP	18	4	0,222	0,458	0,210	2,062
N-V_C_2_L	18	9	0,500	0,687	0,472	1,374
N-V_C_2_NF	18	1	0,056	0,229	0,052	4,123
N-V_C_2_SC	18	14	0,778	0,857	0,735	1,102
N-V_C_2_IT	18	0	0,000	0,000	0,000	0
N-V_E_2_CJ	18	2	0,111	0,324	0,105	2,915
N-V_E_2_P	18	1	0,056	0,229	0,052	4,123
N-V_E_2_PP	18	0	0,000	0,000	0,000	0
N-V_E_2_L	18	0	0,000	0,000	0,000	0
N-V_E_2_NF	18	1	0,056	0,229	0,052	4,123
N-V_E_2_SC	18	1	0,056	0,229	0,052	4,123
N-V_E_2_IT	18	0	0,000	0,000	0,000	0
N-V_S-A_2_CJ	18	0	0,000	0,000	0,000	0
N-V_S-A_2_P	18	0	0,000	0,000	0,000	0
N-V_S-A_2_PP	18	0	0,000	0,000	0,000	0
N-V_S-A_2_L	18	0	0,000	0,000	0,000	0
N-V_S-A_2_NF	18	0	0,000	0,000	0,000	0
N-V_S-A_2_SC	18	0	0,000	0,000	0,000	0
N-V_S-A_2_IT	18	0	0,000	0,000	0,000	0

Table 28: Results of the descriptive analysis in English

Type of Causal relation	N	Sum	Mean	Standard Deviation	Variance	Coefficient of Variance
V_C_1_CJ	18	2	0,111	0,324	0,105	2,915
V_C_1_P	18	9	0,500	0,687	0,472	1,374
V_C_1_PP	18	1	0,056	0,229	0,052	4,123
V_C_1_L	18	9	0,500	0,687	0,472	1,374
V_C_1_NF	18	8	0,444	0,648	0,420	1,458
V_C_1_SC	18	8	0,444	0,648	0,420	1,458
V_C_1_IT	18	0	0,000	0,000	0,000	0
V_E_1_CJ	18	4	0,222	0,458	0,210	2,062
V_E_1_P	18	2	0,111	0,324	0,105	2,915
V_E_1_PP	18	0	0,000	0,000	0,000	0
V_E_1_L	18	3	0,167	0,397	0,157	2,380
V_E_1_NF	18	1	0,056	0,229	0,052	4,123
V_E_1_SC	18	0	0,000	0,000	0,000	0
V_E_1_IT	18	0	0,000	0,000	0,000	0
V_S-A_1_CJ	18	3	0,167	0,397	0,157	2,380
V_S-A_1_P	18	1	0,056	0,229	0,052	4,123
V_S-A_1_PP	18	0	0,000	0,000	0,000	0
V_S-A_1_L	18	0	0,000	0,000	0,000	0
V_S-A_1_NF	18	0	0,000	0,000	0,000	0
V_S-A_1_SC	18	0	0,000	0,000	0,000	0
V_S-A_1_IT	18	0	0,000	0,000	0,000	0
V_C_2_CJ	18	33	1,833	1,316	1,731	0,718
V_C_2_P	18	4	0,222	0,458	0,210	2,062
V_C_2_PP	18	3	0,167	0,397	0,157	2,380
V_C_2_L	18	0	0,000	0,000	0,000	0
V_C_2_NF	18	2	0,111	0,324	0,105	2,915
V_C_2_SC	18	7	0,389	0,606	0,367	1,558
V_C_2_IT	18	0	0,000	0,000	0,000	0
V_E_2_CJ	18	6	0,333	0,561	0,315	1,683
V_E_2_P	18	1	0,056	0,229	0,052	4,123
V_E_2_PP	18	0	0,000	0,000	0,000	0
V_E_2_L	18	0	0,000	0,000	0,000	0
V_E_2_NF	18	0	0,000	0,000	0,000	0
V_E_2_SC	18	0	0,000	0,000	0,000	0
V_E_2_IT	18	0	0,000	0,000	0,000	0
V_S-A_2_CJ	18	0	0,000	0,000	0,000	0
V_S-A_2_P	18	0	0,000	0,000	0,000	0
V_S-A_2_PP	18	0	0,000	0,000	0,000	0
V_S-A_2_L	18	0	0,000	0,000	0,000	0
V_S-A_2_NF	18	0	0,000	0,000	0,000	0
V_S-A_2_SC	18	0	0,000	0,000	0,000	0
V_S-A_2_IT	18	0	0,000	0,000	0,000	0
N-V_C_1_CJ	18	9	0,500	0,687	0,472	1,374
N-V_C_1_P	18	9	0,500	0,687	0,472	1,374
N-V_C_1_PP	18	2	0,111	0,324	0,105	2,915
N-V_C_1_L	18	30	1,667	1,255	1,574	0,753
N-V_C_1_NF	18	2	0,111	0,324	0,105	2,915
N-V_C_1_SC	18	4	0,222	0,458	0,210	2,062
N-V_C_1_IT	18	0	0,000	0,000	0,000	0
N-V_E_1_CJ	18	1	0,056	0,229	0,052	4,123
N-V_E_1_P	18	0	0,000	0,000	0,000	0
N-V_E_1_PP	18	1	0,056	0,229	0,052	4,123
N-V_E_1_L	18	2	0,111	0,324	0,105	2,915
N-V_E_1_NF	18	2	0,111	0,324	0,105	2,915
N-V_E_1_SC	18	3	0,167	0,397	0,157	2,380
N-V_E_1_IT	18	0	0,000	0,000	0,000	0

N-V_S-A_1_CJ	18	0	0,000	0,000	0,000	0
N-V_S-A_1_P	18	0	0,000	0,000	0,000	0
N-V_S-A_1_PP	18	0	0,000	0,000	0,000	0
N-V_S-A_1_L	18	0	0,000	0,000	0,000	0
N-V_S-A_1_NF	18	0	0,000	0,000	0,000	0
N-V_S-A_1_SC	18	0	0,000	0,000	0,000	0
N-V_S-A_1_IT	18	0	0,000	0,000	0,000	0
N-V_C_2_CJ	18	35	1,944	1,355	1,836	0,697
N-V_C_2_P	18	19	1,056	0,998	0,997	0,946
N-V_C_2_PP	18	6	0,333	0,561	0,315	1,683
N-V_C_2_L	18	7	0,389	0,606	0,367	1,558
N-V_C_2_NF	18	0	0,000	0,000	0,000	0
N-V_C_2_SC	18	14	0,778	0,857	0,735	1,102
N-V_C_2_IT	18	0	0,000	0,000	0,000	0
N-V_E_2_CJ	18	1	0,056	0,229	0,052	4,123
N-V_E_2_P	18	1	0,056	0,229	0,052	4,123
N-V_E_2_PP	18	0	0,000	0,000	0,000	0
N-V_E_2_L	18	0	0,000	0,000	0,000	0
N-V_E_2_NF	18	1	0,056	0,229	0,052	4,123
N-V_E_2_SC	18	0	0,000	0,000	0,000	0
N-V_E_2_IT	18	0	0,000	0,000	0,000	0
N-V_S-A_2_CJ	18	0	0,000	0,000	0,000	0
N-V_S-A_2_P	18	0	0,000	0,000	0,000	0
N-V_S-A_2_PP	18	0	0,000	0,000	0,000	0
N-V_S-A_2_L	18	0	0,000	0,000	0,000	0
N-V_S-A_2_NF	18	0	0,000	0,000	0,000	0
N-V_S-A_2_SC	18	0	0,000	0,000	0,000	0
N-V_S-A_2_IT	18	0	0,000	0,000	0,000	0

Table 29: Results of the descriptive analysis in Spanish

When combining the four categories that are analysed in our study, we have realized that the most common causal relations are the following ones. In English original texts the relations more prone to occur are: N-V_C_2_CJ (non-volitional content second position relations expressed through conjunctions), V_C_2_CJ (volitional content second position relations expressed through conjunctions) and N-V_C_1_L (non-volitional content first position relations expressed through lexicon). See examples below.

"His conclusions and their implications form the Tasmanian's Tale, a tale of particular relevance to this rendezvous ***because*** [N-V_C_2_CJ] *Concestor o is the most recent common ancestor of till living humans."* (EO_EXPE_006)

"The Greeks observed only five planets ***because*** [V_C_2_CJ] *five are all we can see with the naked eye: Mercury, Venus, Mars, Jupiter, and Saturn."* (EO_EXPE_007)

"I will retrace the history that ***led*** [N-V_C_1_L] *people to see human nature as a dangerous idea, and I will try to unsnarl the moral and political rat's nests that have entangled the idea along the way."* (EO_EXPE_008)

In case of Spanish translations, we have also noticed that the most prominent type of relations are: N-V_C_2_CJ (non-volitional content second position relations expressed through conjunctions), V_C_2_CJ (volitional content second position relations expressed through conjunctions) and N-V_C_1_L (non-volitional content first position relations expressed through lexicon). In order to visualize these causal relations, we have chosen the following examples.

"Quienes han regresado de un intento de circunnavegar la Tierra no dicen que lo haya impedido la presencia de un continente en su camino, ***porque*** [N-V_C_2_CJ] *el mar se mantenía perfectamente abierto, sino más bien la falta de decisión y la escasez de provisiones…"* (STrans_EXPE_002)

"Nuestro anterior análisis sobre la eutanasia ilustra la naturaleza extrínseca de estas diferencias, ***puesto que*** [V_C_2_CJ] *no ofrecen una base para distinguir entre eutanasia activa y pasiva."* (STrans_ESSAY_010).

"Voy a volver sobre los pasos de la historia que ***condujo*** [N-V_C_1_L] *a las personas a pensar que la idea de la naturaleza humana es peligrosa, e intentaré desenmarañar cuantas confusiones morales y políticas han ido enredando tal idea."* (STrans_EXPE_008)

In general, we can conclude that at first sight it seems that there is great level of similarity between English original texts and Spanish translations. However, after

analysing the results in depth, we have realized that those little differences found in the expression of causal relations by each language are statistically significant, and therefore, changes in the use of those relations are not due to chance but they are based on some linguistics rules governing each language.

To end up with the explanation and discussion of our analysis, we shall mention some peculiarities that we have encountered about those causal relations when comparing English and Spanish.

It seems that English writers and Spanish translators are more prone to use the same types of causal relations (N-V_C_2_CJ, V_C_2_CJ and N-V_C_1_L). However it is quite interesting to point out the lexical variety that each language has to signal all type of causal relations. On the one hand, English has a great variety of signals in relation to conjunctions (*because, for, as, however much, if, since, after all*), prepositions (*for, without, from, through, on, with*), lexicon (*to cause, to make, to lead, to come from, the fear, the reason, the cause*), and syntactic constructions (*this is one reason that/for, the/one reason…is that, why, so + adj + that, that/this is because, such + adj/noun + that/to, that is why*). In case of Spanish, we also find wide variety of signals regarding conjunctions (*si, porque, debido a que, pues, puesto que, ya que, con tal de que, mientras, a medida que, por cuanto, dado que, como, al fin y al cabo*), lexicon (*causar, conducir, llevar, hacer, provocar, deberse a, inducir a, empujar a, proceder de, la causa, el miedo, el motivo, los causantes*), and syntactic constructions (*por qué, es por ello/eso que, es por esta razón que, éste es uno de los motivos de que, la/una razón de… es/sea que, tan/tanto + adj/noun + que, el motivo… es que, ésta es una de las razones por las que, por la razón de que, tal + noun+ que, por lo mucho que, habida cuenta de que, razón por la cual*). Besides, although relations with prepositions

are also manifested with several signals, in Spanish we have less choice in comparison with the types of signals mentioned above.

As we can see, in general Spanish language is richer that English when talking about lexical variance. Proof of this is the following example in which one sentence in English seems to repeat the same signal to express different causal relations, while in Spanish the use of distinct signals is preferred when dealing with diverse causal relations.

"The correlations between parents and children may be telling us only that the same genes that ***make*** *adults loving, authoritative, and talkative* ***make*** *their children self-confident, well-behaved, and articulate."* (EO_EXPE_008)

"Es posible que las correlaciones entre padres e hijos sólo nos indiquen que los mismos genes que ***causan*** *que los padres sean cariñosos, responsables o dialogantes* ***provocan*** *que los hijos sean personas seguras de sí mismas, educadas o que saben expresarse correctamente."* (STrans_EXPE_008)

Other feature that caught our attention is that causal relations with infinitive non-finite clauses are not found in English, while in Spanish the infinitive is a potential signal of causal relation by its own or preceded by a preposition such as *por.* Apparently, when a causal relations is expressed through an infinitive clause in Spanish, its equivalence in English is a gerund clause, which is the most recurrent signal in English non-finite clauses. This feature is exemplified as follows:

*"****Por querer*** *responder a esta pregunta con una sola respuesta hemos caído en la misma trampa que el documental Motherland."* (STrans_EXPE_006)

*"****By expecting*** *a single answer to this question we have fallen into the same trap as the Motherland television documentary."* (EO_EXPE_006)

With regards to the types of signals manifested with lexicon, more concretely with verbal phrases, we should point out that despite most of them tend to be content and non-volitional, they are more likely to appear in first position since second position relations only occur when we are dealing with passive instances. Moreover, in both English and Spanish texts we only find two signals which tend to be used in second position (*To cause* and to *come from* in English, and *causar* and *proceder de* in Spanish). See the example below.

"*Millones de personas con dietas pobres sufren enfermedades carenciales como el bocio, o la ceguera* ***causada*** *por una falta de vitamina A.*" (STrans_ESSAY_10).

In order to come up with some fixed patterns, we shall now have a close look at the most frequent types of causal relations derived from our analysis, which are: N-V_C_2_CJ, V_C_2_CJ and N-V_C_1_L. To start with the first one, we noticed that in English is far expressed through the signal *because*, although it can also appears with other signals such as *for, as* and *since*. In case of Spanish the most recurrent signal in this type of causal relation is *porque*, but signals such as *si, debido a que, pues, puesto que, ya que* and *por cuanto* are also found in this type of causal relations. Secondly, in English V_C_2_CJ are mainly used with the signal *because*, being *for* and *after all* possible options as well. The choice of Spanish signals in this type of causal relation is a little more reduced than the previous one. *Porque* is still the most frequent signals, but it can be replaced by *pues, puesto que, ya que* and *al fin y al cabo*. Finally, the last most common type of relation, N-V_C_1_L, has the wider variety of signal choice. In the case of English, we can find it with the following signals: *to cause, to make, to lead, to come from, the fear, the reason* and *the cause*. On the other hand, Spanish texts illustrate this type of relation with signals such as *causar, conducer, llevar, hacer, provocar, empujar a, la causa, el miedo and los causantes*.

In conclusion, the results suggest that (1) the most frequent causal relations regarding to the cognitive approach and the functional approach are content and non-volitional relations respectively; (2) the type of signal most prominent to represent such relations is a conjunction, and it tend to appears in second position rather than first position; (3) the causal relations which are more recurrent in both languages are: N-V_C_2_CJ (non-volitional content second position relations expressed through conjunctions), V_C_2_CJ (volitional content second position relations expressed through conjunctions) and N-V_C_1_L (non-volitional content first position relations expressed through lexicon); (4) there are minor differences between English and Spanish texts; however, those differences are statistically significant to believe that they are due to chance; and (5) that when are looking for causal signals we are very likely to find *because* or *porque* since they are the most typical signals corresponding to the most common type of causal relations.

7. Summary and concluding remarks

This study was aimed at investigating the most common type of causal relations appearing in English essays and expository texts and their correspondent Spanish translations, as well as the most frequent signals used to express those causal relations

Using corpus annotation as the methodology to gather data from a bilingual data sample of thirty six English original texts and their corresponding translations into Spanish, a number of interesting translation patterns have emerged. First, in terms of RST relations, non-volitional relations predominate over volitional relations in the English dataset, and this tendency is preserved in the Spanish translations.. In terms of the Domain type, content relations predominate in the English dataset, and again, this

tendency is preserved in the Spanish translations.. As to the order of appearance, we notice that second position relations are more often used than first position relations in the English original texts and their corresponding Spanish translations. As to the types of signals preferred by each language, the category of 'conjunction' predominates in the English dataset and is preserved in the Spanish translations, followed by the use of lexical items.

As to possible patterns between both languages, the results of the corpus annotation tasks reveals that the most typical causal relation is non-volitional and content, and that it is a conjunction that appears in second position. Moreover, it is usually signalled by the use of the following linguistic expressions: *because* in English and *porque* in Spanish. At the final point, we have been aware that there are some type of causal relations that are absent in both languages, and these relations are usually associated with speech-act sub-category.

Although this investigation/ project has counted with some limitations, it has left a great number of open issues which could (and should) be studied in the future. For instance, it would be interesting to check whether these variables of causal relations work in the same way at large scale, where more genres are taken into account and when original Spanish texts and English translations are regarded, so that we can find a way to automatize the signalling of causality.

8. Acknowledgments

I would like to thank my supervisor, Dr. Julia Lavid, who has given me helpful guidance and advice at different phases of the work: from the initial planning and design of the methodology to be used to the development of the study, including the guidance with useful computational resources such as *Uam Corpus Tool* for the annotation of the corpus. She also revised the final manuscript thoroughly and provided substantial input to improve its final oral defense. Moreover, I am thankful to Dr. Julia Lavid for giving me permission to use a bilingual sample of texts from the MULTINOT corpus, a research effort of the FUNCAP Research Group, which she coordinates at Universidad Complutense de Madrid.

Clearly, possible mistakes, errors of judgment in the annotation process, misrepresentations of the facts or other failings of this work are only mine. It is my hope that this study, though modest and novice, will shed light on further stimulating insights and future research into the cross-linguistics analysis and annotation of causality.

9. References

Altenberg, B. (1984). Causal Linking in Spoken and Written English. *Studia Linguistica*, 38, 1, 20-69.

Bellés, B. (2006). *Discourse Markers within the University Lecture Genre: A contrastive study between Spanish and North-American lectures* (PhD Dissertation). Universitat Jaume I: Castellón. Retrieved from http://www.tdx.cat/bitstream/handle/10803/10442/tesis.pdf;jsessionid=5AF89DC5228D58F75FE9A42AED81D230.tdx1?sequence=2

Cenoposiciones. (2010). *The Expression of Cause, Reason, Purpose and Result*. Spain, Murcia.

Danlos, L. (2001). Event coreference in causal discourses. In P. Bouillion & F. Busa (eds.) *The Language of the Word Meaning*. Cambridge: Cambridge University Press, pp. 216-242.

Duque, E. (2014). Signaling causal coherence relations. In *Discourse Studies*, 16, 1, pp. 25-46. Retrieved from http://dis.sagepub.com/content/16/1/25.abstract

Fraser, B. (2009). An account of discourse markers. *International Review of Pragmatics*, 1, 293-320.

Halliday, M. and Hasan, R. (1976). *Cohesion in English. London*: Longman.

Lagerwerf, L. (1998). *Causal Connectives Have Presuppositions. Effects on Coherence and Discourse Structure*. The Netherlands: Holland Academic Graphics. Retrieved from http://dare.ubvu.vu.nl/handle/1871/15721

Lavid, J. (1994). *Cognitive constraints on causal ordering strategies in English*, pp. 149-159. Universidad Complutense de Madrid: Madrid. Retrieved from https://dialnet.unirioja.es/descarga/articulo/1958138.pdf

Lavid, J., Arús, J., DeClerk, B. & Hoste, V. (2015). Creation of a high-quality, register-diversified parallel (English-Spanish) corpus for linguistics and computational investigations. In *Procedia – Social and Behavioral Sciences*, 198, pp. 249-256. Retrieved from http://www.sciencedirect.com/science/article/pii/S1877042815044444

Maat, H. & Sanders, T. (2001). Subjectivity in causal connectives: An empirical study of language in use. In *Cognitive Linguistics*, 12, 3, pp. 247-273. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.533.9893&rep=rep1&type=pdf

Meier, E. (2002). Causal Subordination in English and Norwegian. *Nordic Journal of English Studies*, 1, 1, 33-64.

Mendizábal, N. (1997). Expresión de la causa y finalidad en español como segunda lengua. In *REALE*, 8, pp. 77-87. Retrieved from http://dspace.uah.es/dspace/bitstream/handle/10017/7413/expresion_mendizabal_REALE_1997.pdf;jsessionid=F69343EE6106AAE125EFA24B6DE1BFEB?sequence=1

Moeschler, J. (2003). Causality, lexicon and discourse meaning. In *Rivista di Linguistica*, 15, 2, pp. 277-303. Retrieved from linguistica.sns.it/RdL/15.2/06.Moeschler.pd

O'Donnell, M. (2016). Uam corpus tool (version 3.3f) [Software]. Available from http://www.corpustool.com/download.html

Piñero, G. (2001). El valor de los marcadores del discurso que expresan causalidad en español. In Estudios Filológicos, 36, 153-171. Retrieved from http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0071-17132001003600011&lng=en&nrm=iso&ignore=.html

Pit, M. (2003). *How to Express Yourself with a Causal Connective: Subjectivity and Causal Connectives in Dutch, German and French*. The Netherlands, Amsterdam: Rodopi B. V.

Portolés, J. (1998). *Marcadores del discurso*. Barcelona: Ariel.

RST Rethorical Structure Theory. (n.d.) Retrieved Mars 15, 2016 from the RST official page: http://www.sfu.ca/rst/01intro/intro.html

Sanders, T., Wilbert, S. & Leo, N. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 4, 93-133.

Sanders, T. (2005). Coherence, Causality and Cognitive Complexity in Discourse. In *Proceedings/Actes SEM-05, First International Symposium on the Exploration and Modelling of Meaning*, pp. 31-44. Retrieved from www.let.uu.nl/~Ted.Sanders/personal/.../Sanders%20(2005).pdf

Sanders, T. & Sweetser, E. (2009). Causal Categories in Discourse and Cognition. Germany, Berlin: Walter de Gruyter.

Sweetser, E. (1990). *From Etymology to Pragmatics.* Cambridge: Cambridge University Press.

Taboada, M. & Das, D. (2013). Annotation upon Annotation: Adding Signalling Information to a Corpus of Discourse Relations. In *Dialogue and Discourse*, 4, 2, pp. 249-281. Retrieved from https://www.sfu.ca/~mtaboada/docs/Taboada_Das_Dialogue_and_Discourse_2013.pdf

Van Dijk, T. (1979). Pragmatic Connectives. *Journal of Pragmatics*, 3, 5, 447-456.

Zufferey, S. & Cartoni, B. (2012). English and French Causal Connectives in Contrast. In *Languages in Contrast*, 12, 2, pp. 232-250.

List of tables

Table 1 Bilingual sample used in this study

Table 2: Inventory of English and Spanish signals of causal relations

Table 3: Description of causal relations according to the functional approach

Table 4: Description of causal relations according to the cognitive approach

Table 5: Description of causal relations according to the order of appearance

Table 6: Description of causal relations according to the type of signal.

Table 7: Frequency distribution of type of causal relation in bilingual dataset

Table 8: Frequency distribution of signal types in the bilingual dataset

Table 9: Contingency table of type of causal relation in bilingual dataset

Table 10: Chi-Square results of type of causal relation in bilingual dataset

Table 11: Contingency table of type of signal in bilingual dataset

Table 12: Chi-Square results of type of causal relation in bilingual dataset

Table 13: Types of causal relations in English and Spanish

Table 14: Types of causal relations standardized results in English and Spanish

Table 15: Causal relations with conjunctions in English

Table 16: Causal relations with conjunctions in Spanish

Table 17: Causal relations with prepositions in English

Table 18: Causal relations with prepositions in Spanish

Table 19: Causal relations with prepositional phrases in English

Table 20: Causal relations with prepositional phrases in Spanish

Table 21: Causal relations with lexicon in English

Table 22: Causal relations with lexicon in Spanish

Table 23: Causal relations with non-finite clauses in English

Table 24: Causal relations with non-finite clauses in Spanish

Table 25: Causal relations with syntactic constructions in English

Table 26: Causal relations with syntactic constructions in Spanish

Table 27: Causal relations with it-clefts in English

Table 28: Results of the descriptive analysis in English

Table 29: Results of the descriptive analysis in Spanish

List of figures

Figure 1: Preliminary annotation scheme for causal relations in English and Spanish

Figure 2: Annotating an English text

Figure 3: Annotating a Spanish text

Figure 4: Sample of an English essay annotation

Figure 5: Sample of a Spanish essay annotation

Figure 6: Distribution of RST causal relations in English data set

Figure 7: Distribution of RST causal relations in Spanish data set

Figure 8: Distribution of causal relations by Domain type in English

Figure 9: Distribution of causal relations by Domain type in Spanish

Figure 10: Order of appearance results summary in English

Figure 11: Order of appearance results summary in Spanish

Figure 12: Types of causal relations results in English and Spanish

Figure 13: Types of causal relations standardized results in English and Spanish

Figure 14: Type of signal results summary in English

Figure 15: Type of signal results summary in Spanish

Appendix

- Data results in English

CONJUCTIONS		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
because	1st position	1	4		1		
	2nd position	26	33	2	2		
for	1st position						
	2nd position	4	3	1			
as	1st position		2	1		1	
	2nd position		1				
However much	1st position		1				
	2nd position						
if	1st position					1	
	2nd position						
since	1st position	2	4	3	1	1	
	2nd position		3	1			
After all	1st position						
	2nd position	1					

PREPOSITIONS		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
for	1st position	5	5		1		
	2nd position	2	4	1	1		
without	1st position	1	1				
	2nd position		1				

from	1st position	1					
	2nd position	1	10				
Through	1st position						
	2nd position		1				
on	1st position		1				
	2nd position						
with	1st position	2					
	2nd position	1					
by	1st position		1				
	2nd position						

PREPOSITIONAL PHRASES		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
In the face of	1st position					1	
	2nd position						
Because of	1st position		1				
	2nd position	3	3				
From the fact	1st position						
	2nd position		1				
For fear of	1st position						
	2nd position	1					
Due to	1st position				1		
	2nd position						

LEXICON: VERBAL PHRASES		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
To cause	1st	3	10		1		

	position						
	2nd position		6				
To make	1st position	4	6				
	2nd position						
To lead	1st position	4	9	3			
	2nd position						
To come from	1st position						
	2nd position		2				

LEXICON: NOMINAL PHRASES		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
The fear of	1st position		1				
	2nd position						
The reason	1st position		1	1			
	2nd position						
The cause	1st position		1				
	2nd position		1				

NON-FINITE CLAUSES		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
provided	1st position	1		1			
	2nd position				1		
By + gerund	1st position	2	2	1			
	2nd position		1				
gerund	1st position	2					
	2nd position	1					
For + gerund	1st position	1					

	2nd position	1					
Given (the)	1st position	1	1		2	1	
	2nd position			1			

SYNTACTIC CONSTRUCTIONS		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
This is one reason that/for	1st position	1	1				
	2nd position						
The/one reason … is that	1st position						
	2nd position	1	3				
why	1st position						
	2nd position	4	7	1	1		
So + adj + that	1st position	3	1	1	2		
	2nd position						
That/this is because	1st position						
	2nd position		4				
Such + adj/noun + that/to	1st position	1	1				
	2nd position						
That is why	1st position	2	1				
	2nd position						

IT-CLEFT		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
It is why	1st position	1					
	2nd position						
It is because	1st position		1				
	2nd						

	position						
It is for	1st position	1					
	2nd position						

- Data results in Spanish

CONJUCTIONS		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
si	1st position		1			2	
	2nd position		1	1			
porque	1st position						
	2nd position	18	20	3			
Debido a que	1st position		1				
	2nd position		1				
pues	1st position						
	2nd position	4	3				
Puesto que	1st position	1	1	2		1	
	2nd position	3	1	1			
Ya que	1st position						
	2nd position	7	7	1	1		
Con tal de que	1st position	1					
	2nd position						
mientras	1st position			1			
	2nd position						
A medida que	1st position		1				
	2nd position						
Por cuanto	1st position						

	2nd position		1				
Dado que	1st position		4				
	2nd position						
como	1st position		1	1	1		
	2nd position						
Al fin y al cabo	1st position						
	2nd position	1					

PREPOSITIONS		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
con	1st position	2					
	2nd position	1					
por	1st position	5	7	2			
	2nd position	3	11	1	1		
de	1st position						
	2nd position		6				
según	1st position		2				
	2nd position						
sin	1st position	1					
	2nd position		1				
desde	1st position	1					
	2nd position						
ante	1st position					1	
	2nd position		1				

PREPOSITIONAL PHRASES	Content		Epistemic		Speech-act	
	Volitional	Non-	Volitional	Non-	Volitional	Non-

			volitional		volitional		volitional
Debido a	1st position	1	1		1		
	2nd position	2	5				
Gracias a	1st position						
	2nd position		1				
A causa de	1st position		1				
	2nd position						
Por miedo a	1st position						
	2nd position	1					

LEXICON: VERBAL PHRASES		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
causar	1st position	2	6		1		
	2nd position		2				
conducir	1st position		3				
	2nd position						
llevar	1st position	3	5	2			
	2nd position						
hacer	1st position	2	6		1		
	2nd position						
provocar	1st position		3				
	2nd position						
Deberse a	1st position						
	2nd position		4				
Inducir a	1st position			1			
	2nd position						
Empujar	1st		1				

a	position						
	2nd position						
Proceder de	1st position						
	2nd position		1				

LEXICON: NOMINAL PHRASES		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
La causa	1st position		4				
	2nd position						
El miedo	1st position		1				
	2nd position						
El motivo	1st position	2					
	2nd position						
Los causantes	1st position		1				
	2nd position						

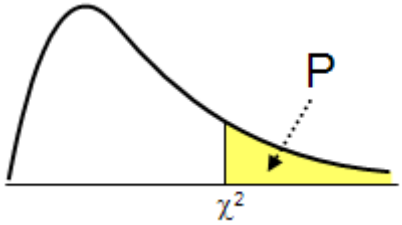
NON-FINITE CLAUSES		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
gerundio	1st position	2	1		1		
	2nd position	1					
Al + gerundio	1st position	3	1	1			
	2nd position						
participio	1st position	1					
	2nd position						
infinitivo	1st position	1					
	2nd position				1		
Por + infinitivo	1st position	1					

	2nd position	1					

SYNTACTIC CONSTRUCTIONS		Content		Epistemic		Speech-act	
		Volitional	Non-volitional	Volitional	Non-volitional	Volitional	Non-volitional
Por qué	1st position						
	2nd position	5	8				
Es por ello/eso que	1st position	1	1				
	2nd position						
Es por esta razón que	1st position	1					
	2nd position						
Éste es uno de los motivos de que	1st position	1					
	2nd position						
La/una razón de… es/sea que	1st position						
	2nd position	1	3				
Tan/tanto + adj/noun + que	1st position	4	1		3		
	2nd position						
El motivo… es que	1st position						
	2nd position		1				
Ésta es una de las razones por las que	1st position	1					
	2nd position						
Por la razón de que	1st position						
	2nd position		1				
Tal + noun+ que	1st position		1				
	2nd position						

Por lo mucho que	1st position						
	2nd position		1				
Habida cuenta de que	1st position						
	2nd position	1					
Razón por la cual	1st position		1				
	2nd position						

- Chi-Square values of distribution table



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728

24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179
25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620
26	11.160	13.844	31.795	35.563	38.885	41.923	42.856	45.642	48.290	51.627	54.052
27	11.808	14.573	32.912	36.741	40.113	43.195	44.140	46.963	49.645	53.023	55.476
28	12.461	15.308	34.027	37.916	41.337	44.461	45.419	48.278	50.993	54.411	56.892
29	13.121	16.047	35.139	39.087	42.557	45.722	46.693	49.588	52.336	55.792	58.301
30	13.787	16.791	36.250	40.256	43.773	46.979	47.962	50.892	53.672	57.167	59.703
31	14.458	17.539	37.359	41.422	44.985	48.232	49.226	52.191	55.003	58.536	61.098
32	15.134	18.291	38.466	42.585	46.194	49.480	50.487	53.486	56.328	59.899	62.487
33	15.815	19.047	39.572	43.745	47.400	50.725	51.743	54.776	57.648	61.256	63.870
34	16.501	19.806	40.676	44.903	48.602	51.966	52.995	56.061	58.964	62.608	65.247
35	17.192	20.569	41.778	46.059	49.802	53.203	54.244	57.342	60.275	63.955	66.619
36	17.887	21.336	42.879	47.212	50.998	54.437	55.489	58.619	61.581	65.296	67.985
37	18.586	22.106	43.978	48.363	52.192	55.668	56.730	59.893	62.883	66.633	69.346
38	19.289	22.878	45.076	49.513	53.384	56.896	57.969	61.162	64.181	67.966	70.703
39	19.996	23.654	46.173	50.660	54.572	58.120	59.204	62.428	65.476	69.294	72.055
40	20.707	24.433	47.269	51.805	55.758	59.342	60.436	63.691	66.766	70.618	73.402
41	21.421	25.215	48.363	52.949	56.942	60.561	61.665	64.950	68.053	71.938	74.745
42	22.138	25.999	49.456	54.090	58.124	61.777	62.892	66.206	69.336	73.254	76.084
43	22.859	26.785	50.548	55.230	59.304	62.990	64.116	67.459	70.616	74.566	77.419
44	23.584	27.575	51.639	56.369	60.481	64.201	65.337	68.710	71.893	75.874	78.750
45	24.311	28.366	52.729	57.505	61.656	65.410	66.555	69.957	73.166	77.179	80.077
46	25.041	29.160	53.818	58.641	62.830	66.617	67.771	71.201	74.437	78.481	81.400
47	25.775	29.956	54.906	59.774	64.001	67.821	68.985	72.443	75.704	79.780	82.720
48	26.511	30.755	55.993	60.907	65.171	69.023	70.197	73.683	76.969	81.075	84.037
49	27.249	31.555	57.079	62.038	66.339	70.222	71.406	74.919	78.231	82.367	85.351
50	27.991	32.357	58.164	63.167	67.505	71.420	72.613	76.154	79.490	83.657	86.661