

Comparing textual genres in Spanish: the case of the tourism domain

Iria da Cunha, M. Amor Montané, Beatriz Fisas & M. Ángeles Escobar
Universidad Nacional de Educación a Distancia (UNED) (Spain), Universitat de Barcelona (UB) (Spain) & Universitat Pompeu Fabra (UPF) (Spain)
iriad@flog.uned.es, amor.montane@ub.edu, beatriz.fisas@upf.edu & maescobar@flog.uned.es

Abstract

Tourism is one of the most important specialized domains in Spain. For this reason, many linguistic studies related to tourism texts have been carried out in the last few years. In this work, we focus on three of the most frequent and difficult textual genres written in Spanish in this domain: the informative article, the travel blog post, and the rules and regulations (for tourist accommodation facilities). The goals of this research are two: a) to carry out a linguistic analysis of the three aforementioned genres from the tourism domain on the textual, lexico-grammatical and discourse levels to gain a comprehensive understanding of the features of these text types, and b) to compare these genres in order to identify statistically significant differences among them and show that some specific linguistic features are prototypical of each one. To achieve these goals, we compiled a corpus in Spanish comprised of texts corresponding to the analysed genres. Then, we analysed linguistic features from both a quantitative and qualitative point of view. Finally, we carried out a comparative analysis using statistical techniques that shed light on significant differences between the genres.

Keywords: textual genres, tourism, corpus, linguistic features, statistical analysis.

Resumen

Una comparación de géneros textuales en español: el caso del ámbito del turismo

Uno de los ámbitos especializados más importantes en España es el turismo. Por esta razón, durante los últimos años se han realizado una gran cantidad de

estudios lingüísticos relacionados con textos turísticos. En este trabajo, nos centramos en tres de los géneros textuales en español más frecuentes y difíciles de escribir en este ámbito: el artículo de divulgación, la entrada de blog de viajero y la normativa (de alojamiento turístico). Esta investigación persigue dos objetivos: a) llevar a cabo un análisis lingüístico de los tres géneros mencionados del ámbito del turismo teniendo en cuenta tres niveles (textual, léxico y discursivo), para obtener una caracterización global de los rasgos de este tipo de textos, y b) comparar estos géneros para identificar diferencias estadísticamente significativas entre ellos y probar que algunos rasgos lingüísticos son específicos de cada uno. Para lograr estos objetivos, hemos compilado un corpus en español formado por textos correspondientes a los géneros analizados. A continuación, hemos analizado rasgos lingüísticos, tanto desde un punto de vista cuantitativo como cualitativo. Finalmente, hemos llevado a cabo un análisis comparativo mediante técnicas estadísticas, que mostraron diferencias significativas entre los géneros.

Palabras clave: géneros textuales, turismo, corpus, rasgos lingüísticos, análisis estadístico.

1. Introduction

Nowadays, the accurate writing of specialised texts has become a relevant activity for professionals. These texts are not easy to write, since textual genres produced in the context of specialised communication have specific characteristics that authors must bear in mind (Cabré, 1999; Gotti, 2008). Furthermore, these characteristics vary across domains and genres (van Dijk, 1989; Swales, 1990; Bhatia, 1993).

Tourism is one of the most important specialised domains in Spain, since it is a key economic pillar of the country. For this reason, many linguistic studies related to this domain have been carried out. They focus both on traditional genres (such as travel itineraries and informative articles), and on emerging genres (such as travel websites and travel blog posts). Calvi (2006 & 2010) and Calvi and Mapelli (2011) proposed an approach for classifying them and offer a detailed analysis of some genres generated in this domain. Based on the list of genres gathered in these papers, da Cunha and Montané (2019) analysed the most frequently occurring genres representing the greatest writing difficulties for university students and tourism professionals in Spain. The informative article, the travel blog post, and the rules and regulations (for tourist accommodation facilities) were three of the most

frequent and difficult genres written in Spanish in the tourism domain. Moreover, the greatest writing difficulties detected were related to textual structure, selection of contents, lexical choices, text cohesion, and degree of formality.

Against this backdrop, this article aims to:

- (a) Carry out a linguistic analysis of the three aforementioned textual genres from the tourism domain on the textual, lexico-grammatical and discourse levels to gain a comprehensive understanding of the features of these text types.
- (b) Compare these textual genres in order to identify statistically significant differences among them and prove that some specific linguistic features are prototypical of each one, according to the assumptions exposed in Section 4.

To achieve these objectives, a corpus-based methodology was used, since as Breeze (2019: 79-80) states:

Corpus-based methods offer a useful way to approach specialised genres, since their strength lies in their ability to detect what is characteristic about texts of a conventionalised nature. By finding what is particularly frequent, or infrequent, in a particular corpus, it is possible to understand more about how language is used – and therefore how meaning is made – in those texts.

Thus, we compiled a corpus comprised of informative articles, travel blog posts, and rules and regulations written in Spanish. Then, we analysed different linguistic categories from both a quantitative and qualitative point of view. Finally, we carried out a comparative analysis using statistical techniques that shed light on significant differences between the analysed genres.

Section 2 includes a review of the main publications about tourism genres. Section 3 introduces our theoretical frameworks. Section 4 sets out the main assumptions of the work. Section 5 explains our methodology, while Section 6 offers a linguistic characterization of the genres according to the proposed corpus. Section 7 contains a statistical design to deal with our comparative analysis of the tourism texts under stake, and Section 8 discusses the results. Finally, Section 9 lays out conclusions and provides some pointers for future work.

2. Literature review

Traditional tourism genres, such as guidebooks, tourism brochures, travel itineraries or informative articles, have been the object of many studies over the years (Calvi, 2006). Calvi (2010) also stresses the need for focusing on new forms of communication that have been developed in the Internet age, such as institutional websites and blogs. Some authors have focused on these emerging genres in the tourism domain, including blogs and Internet forums (Mapelli & Piccioni, 2011), blogs and travel websites (Martínez Escalona, 2012), travel blogs written by non-professionals (Goethals, 2013), tourism newsletters (Rodríguez Abella, 2014), and hotel websites (Cheng, 2016; Suau-Jiménez & Piqué-Noguera, 2017). In this context, Calvi and Mapelli (2011) proposed a multi-functional and multi-dimensional classification that underscores the complexity of textual genres in this domain, which tend to be unconventional and of a relatively hybrid nature. For instance, tourism reports are related with economics, tourism rules and regulations with law, and tourism informative articles with journalism. Owing to this hybrid nature, in general, the analysis of tourism genres is challenging, since determining their macrostructure and formal features is difficult (Calvi, 2011).

The analysis of tourism genres in Spanish has been carried out from several approaches, taking different linguistic features into account. The vast majority of previous studies focused on specific genres. For example, in Mapelli and Piccioni (2011), 13 morphosyntactic features were analysed (adverbs, verbal tenses and persons, and pronouns, among others) in a corpus of nine genres, using Parodi's (2005) multi-dimensional methodology.

Other studies have also analysed morphosyntactic features, such as deixis and discourse modalisation in travel itineraries (Ares, 2011), and adjectives, verbal tenses and persons in tourism webpages (Rodríguez Abella, 2011). Pragmatics and discourse features have been studied as well, for example, discourse voices in travel reports (Canals & Liverani, 2011), interpersonal metadiscourse in institutional webpages (Suau-Jiménez, 2011), linguistic elements expressing ideology in Spanish touristic laws (Carpi, 2011), interpersonal discourse in travel forums (Suau-Jiménez, 2014), pragmatic and textual strategies in hotel reviews (Goethals, 2016), and hedging in research articles and hotel websites (Suau-Jiménez & Piqué-Noguera, 2017).

Other relevant features analysed in the literature are related with the lexicon used in tourism genres, such as verbs of perception in guidebooks and

tourist advertising (Pérez Vázquez, 2011), specialised collocations in tourism webpages (Navarro & Miotti, 2011), lexical variation in hotel promotion websites (Sanmartín, 2011), lexical choices in guidebooks (Mapelli, 2013), and lexicon of cultural heritage in travel blogs and online travel reports (Pano, 2019). Other research that has also focused on the lexicon in tourism, especially on terminology and neology, is found in Calvi (2016), where a bilingual and comparable Spanish-Italian corpus was used. This included travel itineraries, reports, advertising posts, travel blog posts and travel forum posts, among other genres.

Despite the studies carried out in the tourism domain so far, Calvi (2011) pointed out that more research on the structural configuration and rhetorical moves of tourism genres is lacking. In this context, our work is innovative in various ways. It simultaneously analyses three sets of linguistic features (textual, lexico-grammatical and discourse features) from both a qualitative and quantitative point of view. It also compares three genres, using a corpus of texts written in Spanish, automatic Natural Language Processing (NLP) tools, and statistical techniques. Concerning the selected genres, the informative article has been analysed before (Calvi, 2006), as well as the travel blog post (Mapelli & Piccioni, 2011). Nevertheless, to our knowledge, a comparative corpus-based linguistic analysis between these two genres has not been carried out from a quantitative point of view and taking the textual, lexico-grammatical and discourse features proposed in this study into account. Moreover, the rules and regulations genre (in the context of tourist accommodation facilities) has not been previously analysed or compared to other genres in the same domain.

3. Theoretical framework

This section summarises the key features of the theoretical frameworks adopted for the corpus analysis. Each one deals with a linguistic level: textual, lexico-grammatical and discourse levels.

In the first place, we discuss two approaches regarding the textual level. Following van Dijk (1977), we may identify different textual genres according to a clearly codified and widely accepted pattern. For example, the research article usually consists of different sections: Introduction, Problem, Solution and Conclusions. We also follow van Dijk (1989) and assume a textual superstructure, which provides the general organisation of each type

of text. The superstructure shows different sections, and some of them include titles and subtitles. In addition, according to Swales (1990), some passages of the text can be considered as moves. A move “represents a stretch of text serving a particular communicative (that is, semantic) function” (Upton & Cohen, 2009: 589). We use these moves to characterise the textual structure of the genres of our corpus, along the same lines as the corpus-based analysis proposed by Biber et al. (2007).

In the second place, concerning the lexico-grammatical level, we follow the Communicative Theory of Terminology (CTT) by Cabré (1999). According to the CTT, textual genres produced in specialised domains have some global characteristics, such as precision, concision, systematicity, impersonality and objectivity. These characteristics are expressed through different linguistic features. For example, the passive voice is used to express impersonality, while the absence of subjectivity markers implies objectivity, and so on (Cabré et al., 2010).

In the third place, with regard to the discourse level, we use Mann and Thompson’s (1988) Rhetorical Structure Theory (RST). This theory describes how texts are organised in terms of discourse relations between the text’s discourse segments (sometimes marked explicitly using discourse connectors). We follow the definition of discourse segment put forward by Tofiloski et al. (2009: 77):

Discourse segmentation is the process of decomposing discourse into elementary discourse units (EDUs), which may be simple sentences or clauses in a complex sentence, and from which discourse trees are constructed.

In sum, to carry out the linguistic analysis of our corpus, we assume the main proposals of the aforementioned frameworks. First, superstructure and moves are used in the textual level. Second, the global characteristics of texts produced in specialised domains are employed in the lexico-grammatical level. Third, discourse segments, relations and connectors characterise the discourse level.

4. Assumptions

This study starts from several assumptions related to the textual, lexico-grammatical and discourse levels, following the theoretical frameworks mentioned in Section 3. The three general assumptions are the following:

1. The three genres have dissimilar textual, lexico-grammatical and discourse features, on the assumption that “rules and regulations” is the most different genre because of its hybrid nature bridging the legal and the tourism domain.
2. Within the digital media, new genres (such as the travel blog post), which may share target reader and function with traditional genres (such as the informative article), are emerging. These genres might be similar in some specific aspects. Nevertheless, as their communication channel is not the same, they could also differ in some lexico-grammatical and discourse features that would make it possible to differentiate them.
3. The textual, lexico-grammatical and discourse features in informative articles are fixed, as much as in rules and regulations, as they are considered traditional genres. In contrast, there is more room for creativity in the travel blog posts.

The specific assumption related to the textual level is the following:

1. The texts corresponding to the travel blog post and the informative article contain many different sections, titles and moves, depending on the subjects discussed in each text. The texts corresponding to the rules and regulations genre include fewer different sections, titles and moves.

The assumptions concerning the lexico-grammatical level are the following:

1. In rules and regulations, subjective elements are scarce. In travel blog posts and informative articles, subjective elements are usually employed.
2. In the three genres, active voice is mainly employed, as it is usual in Spanish. This voice is more frequent in the travel blog post.
3. Verbs in the first person singular and plural are used in all three genres. The travel blog post is the genre in which the first person singular is used more frequently.

Finally, the assumptions related to the discourse level are the following:

1. Travel blog posts include more sentences than informative articles, and rules and regulations.

2. In travel blog posts, complex sentences are usually divided into discourse segments, which are marked through connectors. In informative articles, and rules and regulations, discourse segmentation and connectors are not so frequent.
3. In informative articles and in travel blog posts, argumentative strategies are often used. Therefore, authors usually include connectors to mark discourse relations aiming to contrast (such as Antithesis, Concession and Contrast), to express reasons to visit a place or make a trip (such as Cause) or to explain the aim of a trip (such as Purpose). Reformulation strategies are also used in these two genres. Thus, authors use connectors to mark relations of Restatement and Summary, in order to offer additional information to the readers to clarify the text. In the rules and regulations genre, all those connectors are scarcely used.
4. In rules and regulations, and in travel blog posts, connectors expressing Condition are usually employed. In informative articles, the Condition relation is used less as a discourse strategy.

5. Corpus methodology

5.1 Building the corpus

As noted above, this study considers three tourism-related genres: informative articles, travel blog posts, and rules and regulations. According to Calvi's (2010) classification, an informative article falls within the publishing genre, and would normally be included in a travel or tourism magazine, providing tourist information about a place. A travel blog post lies within the informal genre, derived from the interactive modalities of digital communication (Dann & Liebman, 2007), offering information about a trip. Rules and regulations are a legal genre stemming from the legal domain and applied to regulate tourism-related activities (Calvi, 2010).

The corpus comprises 20 texts per genre making a total of 60 texts from the tourism domain (see Table 1).

Textual genre	Number of texts	Number of words
Informative article	20	34,532
Travel blog post	20	25,440
Rules and regulations	20	40,277
Total	60	100,249

Table 1. Corpus statistics.

Texts were selected from a wide variety of sources:

- 20 informative articles from five Spanish travel magazines: *Viajar*, *El Viajero*, *Horizontes*, *Top Viajes* and *Revista Ibérica* (four articles each).
- 20 travel blog posts written by specialised Spanish travel bloggers, including Paco Nadal, Nani Arenas and Lucía Sánchez.
- 20 sets of rules and regulations from tourist accommodation facilities placed in different Spanish regions, such as Hotel Route 42 (Castilla-La Mancha), Hotel Celta Galaico (Galicia) and Hotel Atarazanas (Andalucía).

5.2. Categories of analysis

In Appendix 1, the categories corresponding to the linguistic features analysed in this work are listed. On the textual level, the following three categories were considered:

- “Sections” used to structure the text, that is, each of the sections that could be clearly identified, either because section breaks were included or because their topic was distinct. For example, in an informative article, sections included “Heading”, “Introduction”, “Body”, and “Appendix”.
- “Titles” included at the beginning of each section. For example, the “Appendix” section of an informative article includes the title “A Practical Guide.”
- “Moves” found in each section, that is, semantic and functional textual units with a specific communicative purpose.¹ For example, the “Introduction” section of a travel blog post includes “Photo”, “Photo caption”, “Blogger’s thoughts (reason for writing)”, and “links to websites”.

On the lexico-grammatical level, two categories serve to analyse the corpus. On this level, we combine both morphosyntactic features related to verbal aspects (such as voice and number) and specific lexical units that indicate subjectivity. Details are given below:

- Morphosyntactic features, specifically, the use of the “active and passive voices”, and the use of the “first person singular and plural”.
- “Subjectivity markers”, based on a clearly defined group of subjective units (Otaola, 1988), such as superlatives and nouns, adjectives and adverbs conveying opinion. For example: *desgraciadamente* (“unfortunately”), *quizás* (“perhaps”) or *peor* (“worse”).

Finally, on the discourse level, the following three categories are relevant:

- “Sentences”, considering that they are word sequences starting from an upper-case letter and ending with a period, question mark, or exclamation mark.
- “Discourse segments” (Tofiloski et al., 2009). Specifically, we use the criteria for discourse segmentation most used in Spanish (Iruskieta, da Cunha & Taboada, 2015). Examples 1 and 2 include sentences in Spanish extracted from two different rules and regulations in our corpus (with their corresponding translation into English). In these sentences, the different discourse segments are indicated in square brackets.

1. [*En caso de que la salida del pasajero se efectúe antes de la fecha estipulada en la reserva,*] [*no se procederá a efectuar ningún reembolso.*]

[In case a guest departs before the date indicated in the booking confirmation,] [no refund will be issued.]

2. [*A los clientes que no pongan a disposición sus habitaciones durante esas horas no se les realizará la limpieza de la habitación,*] [*aunque se les podrán reponer consumibles.*]

[Should you not make your room available during the stipulated cleaning period, your room will not be cleaned,] [although consumable goods may be replaced.]

- “Discourse connectors” related to eight RST discourse relations: Antithesis, Cause, Concession, Condition, Contrast, Purpose,

Restatement and Summary.² These relations were selected because they occur most frequently in the *RST Spanish Treebank*³ and are also regularly indicated in this corpus using connectors. The connectors associated with the aforementioned relations and considered for the sake of this study were extracted from da Cunha et al. (2011). For instance, in reference to the examples set out earlier: in example 1, the connector *en caso de* (“in case”), which expresses Condition, is detected at the beginning of the first segment of the sentence, and, in example 2, the connector *aunque* (“although”), which expresses Concession, is detected between the two segments included in the sentence.

5.3. Semi-automatic corpus analysis

The linguistic analysis of the aforementioned levels of the corpus took place in two phases. First, data were manually extracted for sections, titles and moves (from the textual level). Second, the remaining data were extracted automatically (from the lexico-grammatical and discourse levels), by using the following automatic NLP tools: A morphosyntactic analyser (*Freeling*; Atserias et al., 2006), and a discourse segmentation system (*DiSeg*; da Cunha et al., 2012).

5.4. Model structures elaboration

In order to characterise genres from a textual point of view, similarities in the sections, titles and moves of each of the three analysed genres were measured. A frequency threshold of 50% was established to determine relevance. Consequently, if the same section, title or move appeared in at least 50% of texts corresponding to a genre, it was deemed to be a relevant feature of that genre. This information was utilised to design a model structure including the sections, titles and moves that tend to appear most frequently in texts for each genre.

5.5. Linguistic characterisation of the genres

The study also analysed the lexico-grammatical and discourse features mentioned in Section 5.2. To characterise the genres, the averages of the different analysed linguistic features were used. As the texts included in the corpus had different sizes (see Table 1), the data were normalised by using the frequency per million words (fpmw, Biber et al., 1998).⁴ The fpmw is

calculated dividing the absolute frequency of the analysed linguistic characteristic by the total number of words in the corpus; the result is multiplied by 1 million. The equation can be represented in this way:

$$\text{fpmw} = (\text{absolute frequency}/\text{number of words in the corpus}) \times 1,000,000$$

6. Linguistic analysis of the corpus

This section includes our proposal of model structures of the three genres from the tourism domain. Moreover, we provide the results of the linguistic analysis on the textual, lexico-grammatical and discourse levels, giving a comprehensive overview of the most outstanding features of each genre.

6.1. Model structures

As explained in Section 5.4, sections, titles and moves were selected taking a frequency threshold into account, in order to design a model structure including those that tend to appear most frequently in each genre. Appendices 2, 3 and 4 offer the model structures for the three genres: informative article, travel blog post, and rules and regulations, respectively.⁵

6.2. Linguistic characterisation

In this section, we provide and discuss the normalised averages of the analysed features for each genre. In the first place, the quantitative results related to the informative article are shown in Table 2.

Textual level		Lexical level		Discourse level	
Feature	Average	Feature	Average	Feature	Average
Nº of sections	3,352.1	Nº of subjective units	7,736.8	Nº of sentences	42,830.0
Nº of titles	4,030.8	Nº of verbs in the active voice	98,506.5	Nº of discourse segments	68,544.0
Nº of moves	24,020.9	Nº of verbs in the passive voice	840.6	Total Nº of connectors	6,298.9
		Nº of verbs in the first person singular	3,102.7	Nº of connectors expressing Antithesis	2,142.2
		Nº of verbs in the first person plural	1,685.1	Nº of connectors expressing Cause	1,054.4
				Nº of connectors expressing Concession	964.1
				Nº of connectors expressing Condition	1,801.9
				Nº of connectors expressing Contrast	41.6
				Nº of connectors expressing Purpose	230.0
				Nº of connectors expressing Restatement	64.7
				Nº of connectors expressing Summary	0.0

Table 2. Normalised averages related to analysed data for the informative article genre.

On the textual level, taking the averages correlation between sections and titles into account, we observe that each section of the informative article genre usually contained more than one title. Moreover, each section tended to include several moves (more than seven per section). On the lexicogrammatical level, the occurrences of subjective units were noteworthy, taking into account the number of sentences of the corpus: one out of five sentences included at least one subjective unit. In this genre, the active voice was primarily used rather than the passive voice. Moreover, verbs in the first person singular were used more frequently than verbs in the first person plural. On the discourse level, sentences comprised one or two discourse segments. The most frequent connectors expressed Antithesis and Condition. There were no connectors expressing Summary.

In the second place, the quantitative results related to the travel blog post are included in Table 3.

Textual level		Lexical level		Discourse level	
Feature	Average	Feature	Average	Feature	Average
Nº of sections	4,691.7	Nº of subjective units	11,441.1	Nº of sentences	51,432.3
Nº of titles	4,742.7	Nº of verbs in the active voice	129,272.3	Nº of discourse segments	87,596.6
Nº of moves	26,616.3	Nº of verbs in the passive voice	592.8	Total Nº of connectors	11,525.0
		Nº of verbs in the first person singular	10,685.9	Nº of connectors expressing Antithesis	3,749.1
		Nº of verbs in the first person plural	3,223.0	Nº of connectors expressing Cause	1,872.5
				Nº of connectors expressing Concession	1,128.1
				Nº of connectors expressing Condition	4,158.7
				Nº of connectors expressing Contrast	108.2
				Nº of connectors expressing Purpose	508.4
				Nº of connectors expressing Restatement	0.0
				Nº of connectors expressing Summary	0.0

Table 3. Normalised averages related to analysed data for the travel blog post genre.

In this genre, on the textual level, each section contained around one title in average. In addition, more than five moves per section were included. On the lexico-grammatical level, the occurrences of subjective units were remarkable, taking into account that one out of four sentences of the corpus showed at least one of these units. Furthermore, this genre predominantly used the active voice, as passive voice was very unusual in this kind of texts. Finally, the analysis revealed that first person singular verbs appeared more frequently than first person plural verbs. On the discourse level, sentences often comprised almost two discourse segments in average. The most frequent connectors expressed, again, Condition and Antithesis. There were no connectors expressing Summary and Restatement.

In the third place, the quantitative results related to the rules and regulations are shown in Table 4.

Textual level		Lexical level		Discourse level	
Feature	Average	Feature	Average	Feature	Average
Nº of sections	1,405.7	Nº of subjective units	3,961.3	Nº of sentences	52,146.3
Nº of titles	6,111.6	Nº of verbs in the active voice	94,439.3	Nº of discourse segments	76,791.2
Nº of moves	19,252.3	Nº of verbs in the passive voice	2,030.8	Total Nº of connectors	6,576.2
		Nº of verbs in the first person singular	5,715.0	Nº of connectors expressing Antithesis	482.6
		Nº of verbs in the first person plural	1,513.5	Nº of connectors expressing Cause	176.8
				Nº of connectors expressing Concession	243.5
				Nº of connectors expressing Condition	5,046.5
				Nº of connectors expressing Contrast	0.0
				Nº of connectors expressing Purpose	472.1
				Nº of connectors expressing Restatement	154.7
				Nº of connectors expressing Summary	0.0

Table 4. Normalised averages related to analysed data for the rules and regulations genre.

In this genre, on the textual level, sections contained more than four titles in average. Moreover, each section tended to include about 13 moves. On the lexico-grammatical level, we highlight that subjective units were barely used in the corpus, since only one out of 13 sentences of the corpus included these units. Again, in this genre the active voice was primarily used rather than the passive voice. Finally, first person singular verbs appeared more frequently than first person plural verbs. On the discourse level, sentences rarely comprised more than one discourse segment. The most frequent connectors expressed Condition. There were no connectors expressing Summary and Contrast.

7. Statistical design

In order to check that the proposed features allow us to characterise the different types of texts analysed in this work, this section includes a statistical design given the following hypotheses:

- Null Hypothesis (H0): There are no differences between the textual genres considering the textual, lexico-grammatical and discourse features proposed in this work (see Sections 4 and 5.2).
- Alternate Hypothesis (our claim): There are differences between

textual genres considering the aforementioned features proposed in this work.

As Cantos (2002) explains, the quantitative analysis of linguistic features allows us to characterise specialised domains and genres. Nevertheless, the same author (Cantos, 2002: 241) highlights the relevance of statistical significance:

We cannot decide just by looking at these figures; we need to perform a further calculation: a test of statistical significance and determine how high or low the probability is that the difference between the two corpora on these features is due to chance.

Thus, in this study, several statistical tests were conducted to compare the normalised data across the three linguistic levels (following Barón & Téllez, 2004). To do so, the categories corresponding to the analysed features are used as statistical variables. Levene's test was used to analyse normal distribution. Later, different tests were run to compare averages. An analysis of variance (ANOVA) test was conducted to determine if it was possible to differentiate among genres for parametric variables,⁶ while the Kruskal-Wallis test for independent samples was utilized for non-parametric variables.⁷ Subsequently, in order to determine if there were statistically significant differences among specific genres, post hoc multiple comparison tests were conducted. Based on the results of Levene's test, either Tukey's HSD test (for variables with equal within-group variance) or Dunnett's T3 test (for variables with unequal variances) was utilised. Finally, a discriminant analysis was conducted to describe significant differences among the genres.

Concerning the textual level, the statistical tests revealed statistically significant differences among genres for two textual variables. For the "N° of moves" variable, when genre pairs were compared, the travel blog post and the rules and regulations had significant differences ($p = .018$). Hence, we reject the H_0 , and validate the alternate hypothesis. In fact, the travel blog post included the most moves, whereas the latter, fewer moves. For the "N° of sections" variable, the rules and regulations genre differed the most from the other genres, informative article ($p = .001$) and travel blog post ($p = .000$), because it contained fewer sections than the other two genres, as expected. Regarding the "N° of titles", no significant differences were found among genres. Therefore, the alternate hypothesis related to the textual features was partially validated.

With regard to the lexico-grammatical level, statistical tests did identify statistically significant differences for four out of the five analysed variables. For the “N° of subjective units” variable, the three genres differed significantly ($p < .05$), as expected, given that the number of these units in each genre varied. In the rules and regulations genre, they were very scarce. Moreover, we found that subjectivity is a key feature of the travel blog post, since it contains more subjective units than the other two genres.

Concerning morphosyntactic verbal features, for the “N° of verbs in the passive voice” and “N° of verbs in the active voice” variables, the H0 was rejected, since we obtained significant differences ($p < .05$). This result may be due to the fact that the travel blog post rarely exhibits passive voice, since it prioritizes the use of active voice. With respect to the variables related to verbal persons, the significant differences were only found between the informative article and the travel blog post ($p = .01$) concerning the first person singular. The latter is the genre where the first person singular was used more frequently. Surprisingly, no differences regarding this variable were found between the travel blog post, and rules and regulations. Hence, the alternate hypothesis related to the lexico-grammatical features was partially validated.

In relation with the discourse level, statistical differences were found for six variables. No significant differences were found among genres regarding the number of sentences. The informative article and the travel blog post differed significantly between them for the variables “N° of discourse segments” ($p = .002$) and “total N° of connectors” ($p = .001$), the second being the genre where more segments and connectors were detected. Regarding the variable “total N° of connectors,” the travel blog post also contained significantly more connectors than rules and regulations ($p = .002$). These two results allowed us to reject the H0.

In the case of the variables related to specific connectors, the rules and regulations genre had more differences from the others concerning the use of connectors expressing Antithesis, Concession and Cause ($p < .05$). These connectors were scarcely used in this genre, as expected. Nevertheless, we did not find significant differences among genres in the use of connectors expressing Purpose, Contrast, Restatement and Summary. In the case of Purpose connectors, they were used in the texts of the three genres and their frequency was similar. In the case of the other connectors, their frequency in the corpus was so low that it was not possible to obtain statistical

significance. Regarding the use of connectors expressing Condition, the statistical results showed that the informative article had differences with the travel blog post ($p = .01$), and rules and regulations ($p = .000$). In these two genres, this kind of connectors was more frequent, as we had anticipated. Consequently, the alternate hypothesis concerning the discourse features was also partially validated.

Finally, the results of a discriminant analysis revealed statistically significant differences between the three genres. The classification of genres proved to be rather suitable, since 93.3% of texts were assigned to the correct genre. Cross validation results were also acceptable, since 90.0% of texts were correctly classified. Results are depicted graphically in Figure 1.

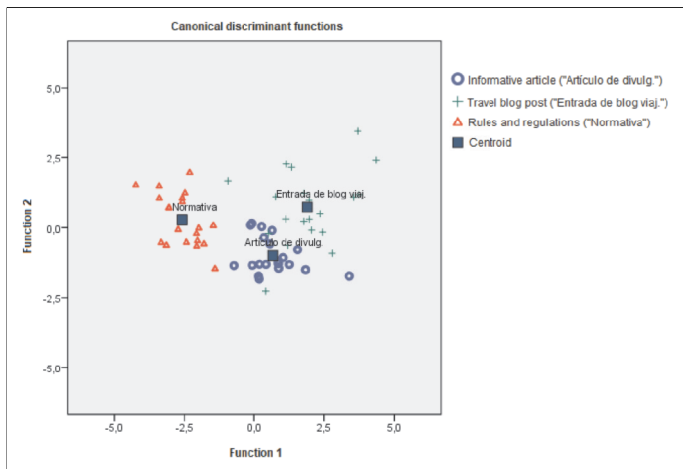


Figure 1. Results of the discriminant analysis in graphic format.

Centroids (that is, the mean discriminant score for each group) were clearly distinct for all the genres, which fell far from one another. The distance was furthest for the rules and regulations, implying that this genre has more dissimilar textual, lexico-grammatical and discourse features. In addition, texts from different genres hardly overlapped. Only a few texts from the travel blog post genre overlapped with texts from the informative article genre, since they are similar in some specific aspects.

Finally, the largest dispersion was found for the travel blog post, implying that texts in this genre most differ from each other. Conversely, the informative article and the rules and regulations genres had the least degree

of dispersion, indicating that the features of the texts corresponding to each genre tend to be similar. Thus, the discriminant analysis supported that the three genres analysed have linguistic characteristics that allow us to differentiate them from each other, which validates our hypothesis and confirms the three general assumptions included in Section 4.

8. Discussion

In this section, we discuss the significant differences obtained from the statistical analysis, which concern the assumptions mentioned in Section 4. With regard to the textual level, we proved that there are significant differences between the travel blog post and the rules and regulations regarding the inclusion of moves, and between the three genres concerning the number of sections. The texts corresponding to the travel blog post genre include many different moves, depending on the subjects discussed in each post. By contrast, the texts corresponding to the rules and regulations genre include fewer different moves, since these moves coincide among the texts. In the case of sections, something similar happens: rules and regulations also include many sections, but these sections are common to most of the texts in our corpus. Nevertheless, in the travel blog post and in the informative article there is a smaller number of sections, but these sections do not usually coincide among the different texts of the corpus. These differences are due to the hybrid nature between the legal and the tourism domain of rules and regulations, which follows a prototypical textual structure, based on legal articles (for example, “Article 1”, “Article 2”, “Article 3”). Thus, the features of the legal domain would affect the global characterisation of this tourism genre. In fact, Calvi (2011) indicates that not all genres commonly used in tourism communication are specific to the sector: some of them belong rather to related fields, and hardly undergo changes when moving to the sphere of tourism. This is what happens with rules and regulations, which properly belong to the legal field.

Concerning the lexico-grammatical level, we proved that the presence of subjective units is a lexical feature that allows us to differentiate between genres. In rules and regulations, subjective units are scarce, since the texts of this genre must be objective, because they have some characteristics of legal texts, as mentioned. Moreover, they should tend to impersonality, like legal contracts (Alcaraz, Hugues & Gómez, 2014). However, in the travel blog

post, subjective units are very frequent: as authors express their own travel experiences, the content is more personal. In informative articles, subjective units are also included, since they provide tourist information about a place. Therefore, these genres are not so objective. As Calvi (2011) states, the discourse of tourism tends to move away from the subjective vision of the travel stories, but it recovers this vision in genres such as the travel blog, which is related to autobiographical genres. Nevertheless, our findings indicate that subjective units are more frequent in travel blog posts than in informative articles. This result makes us think that, as informative articles are published in a travel or tourism magazine, they are expected to be more objective than travel blog posts.

We also statistically evidenced that the active voice is a morphosyntactic feature that characterizes the travel blog post. This result seems to indicate that travel bloggers tend to use active voice in order to explain their own trip more dynamically and to emphasise the grammatical subject. We proved too that the first person singular allows differentiation between the travel blog post and the informative article. This feature is more frequent in the travel blog post since it resembles the genre of autobiography (Calvi, 2011). In addition, as Mapelli and Piccioni (2011) have pointed out, on the one hand, the travel blog post is associated with a situational context of continuum (that is, chronological storytelling) and, on the other hand, it shares features with informal communication. In the informative article, the use of the first person singular is not as recurrent, because it is a more formal genre.

Regarding the discourse level, our results showed that the travel blog post is the genre with most connectors. These results confirm that travel bloggers try to write clear and dynamic texts, and, thus, they tend to divide complex sentences into segments (usually, marked through connectors), in order for the recipient to easily understand the texts. In the case of connectors corresponding to argumentative strategies, our results showed that connectors of Antithesis, Concession and Cause are scarcely used in rules and regulations: first, because of the schematic format of this genre and, second, because these argumentative strategies using connectors would be unnecessary in this kind of texts. By contrast, in informative articles and in travel blog posts, argumentative strategies are often used. Therefore, authors usually include discourse relations aiming to contrast (Antithesis and Concession) and to express reasons to visit a place or make a trip (Cause).

We also demonstrated that connectors expressing Condition are usually employed in rules and regulations and in travel blog posts, while they are not common in informative articles. In the case of rules and regulations, this would be due to the inclusion of different legal articles that usually contain conditions, as mentioned in Alcaraz, Hugues and Gómez (2014), which must be met by both the guests and the tourist accommodation (for example, “If the guest leaves the room after 12 noon, a penalty must be paid”). In the case of the travel blog post, as bloggers try to involve the readers, the use of the Condition relation is one of the strategies employed to motivate readers to visit a specific place (for example, “If you go to that city, don’t miss visiting its cathedral”). On the contrary, informative articles are usually more descriptive and the Condition relation would be less used as a discourse strategy.

It is interesting to note that, although in this work we assume that there are differences between rules and regulations and the other two genres, the results show that they are not so different regarding some linguistic features. Some of these features are the number of sentences included in the texts, the number of segments, the total number of connectors, and the connectors expressing Condition.

Lastly, the discriminant analysis supported the notion that the three genres analysed have linguistic characteristics that allow us to differentiate between them. Rules and regulations turned out to be the genre that can be most easily distinguished because, as mentioned, it is a hybrid genre that shares features of the legal and tourism domains. The travel blog post and the informative article are similar genres, but they are also distinguishable. Actually, although these two genres share target reader and function, their communication channel is different, which has an impact on their linguistic features. The differences found in our research concerning lexicogrammatical and discourse features between the two genres reinforce this idea. In addition, the travel blog post had more dispersion among the texts, because it is the most creative and flexible genre of those analysed, with the most personal writing style.

9. Conclusions and future work

This article focuses on three textual genres in the tourism domain: the informative article (about a tourist destination), the travel blog post (about a travel experience), and the rules and regulations (for a tourist

accommodation facility). Its first goal was to perform an analysis of some linguistic features of these genres on the textual, lexico-grammatical and discourse levels, in order to gain an in-depth understanding of them. For its second goal, a statistical comparison among genres was carried out to prove that some specific linguistic features are prototypical of each one, according to the assumptions set out in Section 4, using a corpus-based methodology. Our corpus contained 60 sample texts in Spanish from these genres.

This study makes two major contributions. First, we provide a new corpus with a linguistic characterisation of the three aforementioned genres. We also elaborate the most frequent model structures utilised in these genres. These model structures include the sections, titles and moves that appear regularly in each one. In addition, the study offers quantitative information about their textual, lexico-grammatical and discourse features. We analyse morphosyntactic features including verbs in the active and passive voice, verbs in the first person singular and plural, and lexical units indicating subjectivity. We also consider discourse features such as sentences, discourse segments, and discourse connectors.

Second, this article provides a statistical design that has allowed us to test our hypothesis: the genres under discussion are different from each other, considering the textual, lexico-grammatical and discourse features proposed here. The corpus size seems to be suitable for texts statistical processing, taking into account the fact that significant statistical differences emerged from the corpus analysis.

Our current research is related to the design of an automatic tool to help to write tourism-related texts, bearing in mind the analysis included in this article. This tool, named arText, is available online and free of charge at: <http://sistema-artext.com/>. In the future, this tool could be implemented in different languages, such as English.

Acknowledgements

This work was supported by a research project that received a Beca Leonardo 2015 from the Fundación BBVA (IN[15]_hMS_IIN_0371), by a research project funded by the Ministerio de Ciencia, Innovación e Universidades at the call Proyectos I+D del Subprograma Estatal de Generación de Conocimiento 2018 (PGC2018-099694-A-I00), and by a Ramón y Cajal contract (RYC-2014-16935), all of them associated with the

Department of Foreign Philologies and their Linguistics at Universidad Nacional de Educación a Distancia (UNED) in Spain. This research was developed in the framework of the ACTUALING and IULATERM research groups. We would like to thank Josh Goldsmith for the translation of the text, and Sheila Queralt for the statistical advice.

Article history:

Received 25 June 2020

Received in revised form 31 March 2021

Accepted 27 June 2021

References

- Alcaraz, E., B. Hughes & A. Gómez (2014). *El español jurídico*. Barcelona: Ariel.
- Ares, A. (2011). "La deixis y la modalidad discursiva en los itinerarios de viaje periodísticos" in M. V. Calvi & G. Mapelli (eds.), 131-151.
- Atserias, J., B. Casas, E. Comelles, M. González, Ll. Padró, & M. Padró (2006). "FreeLing 1.3. Syntactic and semantic services in an open-source NLP library" in *LREC 2006 Proceedings. 5th Edition of the International Conference on Language Resources and Evaluation*, 48-55. European Language Resources Association (ELRA). URL: <<https://aclanthology.org/volumes/L06-1/>> [20/06/2020].
- Barón, F. J. & F. Téllez (2004). *Apuntes de bioestadística*. Universidad de Málaga.
- Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. Longman.
- Biber, D., U. Connor, & T. A. Upton (2007). *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. John Benjamins.
- Biber, D., S. Conrad, & R. Reppen (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
- Breeze, R. (2019). "Part-of-speech patterns in legal genres: Text-internal dynamics from a corpus-based perspective" in T. Fanego & P. Rodríguez-Puente (eds.), *Corpus-based Research on Variation in English Legal Discourse*, 79-103. John Benjamins.
- Cabré, M. T. (1999). *La terminología. Representación y comunicación*. IULA.
- Cabré, M. T., C. Bach, I. da Cunha, A. Morales & J. Vivaldi (2010). "Comparación de algunas características lingüísticas del discurso especializado frente al discurso general: el caso del discurso económico", in *Proceedings of the XXVII Congreso Internacional de AESLA*, 453-460. Universidad de Castilla-La Mancha.
- Calvi, M. V. (2006). *Lengua y comunicación en el español del turismo*. Arco Libros.
- Calvi, M. V. (2010). "Los géneros discursivos en la lengua del turismo: una propuesta de clasificación". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 19: 9-32.
- Calvi, M. V. (2011). "Pautas de análisis para los géneros del turismo" in M. V. Calvi & G. Mapelli (eds.), 19-45.
- Calvi, M. V. (2016). "Léxico de especialidad y lengua del turismo" in A. L. Duffé (ed.), *Estudios sobre el léxico. Puntos y contrapuntos*, 187-214. Peter Lang.
- Calvi, M. V. & G. Mapelli (eds.) (2011). *La lengua del turismo. Géneros discursivos y terminología*. Peter Lang.
- Canals, J., & E. Liverani (2011). "Voces del discurso en los reportajes de viajero" in M. V. Calvi & G. Mapelli (eds.), 113-130.
- Cantos, P. (2002). "Do we need statistics when we have linguistics?". *DELTA* 18(2): 233-271. DOI: <<https://doi.org/10.1590/s0102-44502002000200003>>.
- Carpi, E. (2011). "Lengua e ideología: huellas polifónicas en los preámbulos de la legislación turística española" in M. V. Calvi & G. Mapelli (eds.), 201-224.
- Cheng, F.-W. (2016). "Constructing hotel brands: A multimodal analysis of luxury hotel homepages". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 31: 83-108.
- da Cunha, I. & M. A. Montané (2019). "Textual

- genres and writing difficulties in specialized domains". *Revista Signos* 52(99): 4-30. doi: <dx.doi.org/10.4067/S0718-09342019000100004>.
- da Cunha, I., E. San Juan, J.-M. Torres-Moreno, M. Lloberes & I. Castellón (2012). "DiSeg 1.0: The First System for Spanish Discourse Segmentation". *Expert Systems with Applications* 39(2): 1671-1678. doi: <https://doi.org/10.1016/j.eswa.2011.06.058>.
- da Cunha, I., J.-M. Torres-Moreno & G. Sierra (2011). "On the Development of the RST Spanish Treebank". *Proceedings of the 5th Linguistic Annotation Workshop. 49th Annual Meeting of the ACL*, 1-10. ACL. URL: <https://aclanthology.org/volumes/W11-04/> [20/06/2020].
- Dann, G. M. S. & G. Liebman (2007). "From Travelogue to Travelblog: (Re-)negotiating Tourist Identity". *Acta Turistica* 19(1): 7-29.
- Goethals, P. (2013). "Travel blogs written by non-professionals: An exploratory analysis of a tourism genre". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 25: 147-170.
- Goethals, P. (2016). "Multilingualism and international tourism: a content- and discourse-based approach to language-related judgments in web 2.0 hotel reviews". *Language and Intercultural Communication* 16(2): 235-253. doi: <https://doi.org/10.1080/14708477.2015.1103249>.
- Gotti, M. (2008). *Investigating Specialized Discourse*. Peter Lang.
- Iruskieta, M., I. da Cunha & M. Taboada (2015). "A qualitative comparison method for rhetorical structures: Identifying different discourse structures in multilingual corpora". *Language Resources and Evaluation* 49: 263-309. doi: <https://doi.org/10.1007/s10579-014-9271-6>.
- Mann, W. C. & S. A. Thompson (1988). "Rhetorical structure theory: Toward a functional theory of text organization". *Text* 8(3): 243-281. doi: <https://doi.org/10.1515/text.1.1988.8.3.243>.
- Mapelli, G. (2013). "El léxico de las guías descriptivas" in L. Chierichetti & G. Garofalo (eds.), *Discurso profesional y lingüística de corpus. Perspectivas de investigación*, 125-138. CELSB.
- Mapelli, G. & S. Piccioni (2011). "Taxonomía de los textos turísticos: factores lingüísticos y factores contextuales" in M. V. Calvi & G. Mapelli (eds.), 47-74.
- Martínez Escalona, V. (2012). "Análisis de los portales de viajes de los periódicos españoles de referencia. Los casos de 'elviajero.com' y 'ochologuas.com'" in *Actas IV Congreso Internacional Latina de Comunicación Social*, 1-17. Sociedad Latina de Comunicación Social.
- Molina, C. & G. E. Sierra (2015). "Hacia una normalización de la frecuencia de los corpus CREA y CORDE". *Revista Signos* 48(89): 307-331. doi: <http://dx.doi.org/10.4067/S0718-09342015000300002>.
- Navarro, M. C. & R. Miotti (2011). "La combinatoria preferente en los sitios web de los lugares patrimonio de la UNESCO españoles e italianos" in M. V. Calvi & G. Mapelli (eds.), 315-334.
- Otaola, C. (1988). "La modalidad (con especial referencia a la lengua española)". *Revista de Filología Española* 68(1): 97-117. doi: <https://doi.org/10.3989/rfe.1988.v68.i1/2.414>.
- Pano, A. (2019). "El léxico del patrimonio cultural florentino en blogs y reportajes de viaje en línea". *Círculo de Lingüística Aplicada a la Comunicación* 78: 49-62. doi: <http://dx.doi.org/10.5209/CLAC.64371>.
- Parodi, G. (2005). "Lingüística de corpus y análisis multidimensional: Exploración de la variación en el corpus PUCV-2003". *Revista Española de Lingüística* 35(1): 45-76.
- Pérez Vázquez, M. E. (2011). "Verbos de percepción sensible e intelectual en las guías y en la publicidad turística" in M. V. Calvi & G. Mapelli (eds.), 295-313.
- Rodríguez Abella, R. M. (2011). "La lengua de la promoción turística en el sitio www.turismodecanarias.com" in M. V. Calvi & G. Mapelli (eds.), 153-175.
- Rodríguez Abella, R. M. (2014). "Reflexiones en torno al género boletín digital turístico: situación comunicativa y análisis lingüístico". *Normas* 4: 117-133.
- Sanmartín, J. (2011). "Unidad y variación en el español de turismo: las páginas web de promoción de hoteles de España y Chile" in M. V. Calvi & G. Mapelli (eds.), 335-357.
- Suau Jiménez, F. (2011). "La persuasión a través del metadiscursos interpersonal en el género Página Web Institucional de Promoción Turística en inglés y español" in M. V. Calvi & G. Mapelli (eds.), 177-199.
- Suau-Jiménez, F. (2014). "Dialogic voices of writers and readers in traveller forums through interpersonality" in L. Gil & C. Soler (eds.), *Dialogicity in Written Specialised Genres*, 137-163. John Benjamins.
- Suau-Jiménez, F. & C. Piqué-Noguera (2017). "Hedging in tourism discourse: the variable genre in academic vs professional texts" in *Proceedings*

of the IX Congreso Internacional de Lingüística de Corpus. Paris.

Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.

Tofloski, M., J. Brooke & M. Taboada (2009). "A syntactic and lexical-based discourse segmenter" in *Proceedings of the 47th Annual Meeting of ACL*, 77-80. ACL.

Upton, T. A. & M. A. Cohen (2009). "An approach to corpus-based discourse analysis: The move analysis as example". *Discourse Studies* 11(5): 585-605. doi: <<https://doi.org/10.1177/1461445609341006>>.

van Dijk, T. A. (1977). *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. Longman.

van Dijk, T. A. (1989). *La ciencia del texto*. Paidós.

Iria da Cunha holds a PhD in Language Sciences and Applied Linguistics (*Universitat Pompeu Fabra*, 2008). She is an Associate Professor at the Department of Foreign Philologies and their Linguistics at the *Universidad Nacional de Educación a Distancia* (UNED, Spain). She is a member of the ACTUALING and IULATERM research groups. Her main research lines are Specialized Discourse, Terminology and Natural Language Processing.

M. Amor Montané holds a PhD in Language Sciences and Applied Linguistics (*Universitat Pompeu Fabra*, 2012). She is a Serra Hünter Fellow at *Universitat de Barcelona*. She is a member of the IULATERM research group at *Institut de Lingüística Aplicada* of *Universitat Pompeu Fabra*. Her main research lines are Terminology, Specialized Discourse and Neology.

Beatriz Fisas has a BA in Linguistics with Honors by the *Universitat de Barcelona*. She is a member of the Natural Language Processing Research Group (TALN) at *Universitat Pompeu Fabra*. Her main research lines are Applied Linguistics, Natural Language Processing, Specialized Discourse and Corpus Linguistics.

M. Ángeles Escobar holds a PhD in Linguistics (University of Utrecht, 1995). She is an Associate Professor at the Department of Foreign Philologies and their Linguistics at the *Universidad Nacional de Educación a Distancia* (UNED, Spain). She is a member of the ACTUALING research group. Her main research lines are Contrastive Grammar (English/Spanish), Second Language Acquisition and Psycholinguistics.

NOTES

¹ All the moves were validated by the authors of this article.

² See RST website to obtain the definitions of these relations: <<http://www.sfu.ca/rst/index.html>> [22/06/20].

³ <http://www.corpus.unam.mx/rst/index_es.html> [22/06/20].

⁴ This method is frequently used in corpus linguistics, for example, in the *British National Corpus* and the *Corpus of Contemporary American English* (Molina & Sierra, 2015).

⁵ Literal English translations of the Spanish model structures are included in this article. We are aware that sections, titles and moves for the same textual genres may differ in English, and that a textual analysis of an English corpus would be necessary to linguistically characterise texts written in this language.

⁶ The parametric variables of the study are “N° of moves”, “N° of subjective units”, “N° of verbs in the passive voice”, “N° of sentences”, “total N° of connectors”, “N° of connectors expressing Condition” and “N° of connectors expressing Purpose”.

⁷ The non-parametric variables are “N° of sections”, “N° of titles”, “N° of verbs in the active voice”, “N° of verbs in the first person singular”, “N° of verbs in the first person plural”, “N° of discourse segments”, “N° of connectors expressing Antithesis”, “N° of connectors expressing Cause”, “N° of connectors expressing Contrast”, “N° of connectors expressing Concession” and “N° of connectors expressing Restatement”.

Appendix 1. Categories corresponding to the linguistic features analysed in this work.

DOMAIN: Textual genre		Text 1: [Text name]
Textual level	Number (N°)	Extracted data from texts
N° of sections		Name of section 1 Name of section 2 Name of section <i>n</i>
N° of titles		Title 1 Title 2 Title <i>n</i>
N° of moves		Move 1 Move 2 Move <i>n</i>
Lexical level	Number (N°)	
N° of subjective units		
N° of verbs in the active voice		
N° of verbs in the passive voice		
N° of verbs in the first person singular		
N° of verbs in the first person plural		
Discourse level	Number (N°)	
N° of sentences		
N° of discourse segments		
Total N° of connectors		
N° of connectors expressing Antithesis		
N° of connectors expressing Cause		
N° of connectors expressing Concession		
N° of connectors expressing Condition		
N° of connectors expressing Contrast		
N° of connectors expressing Purpose		
N° of connectors expressing Restatement		
N° of connectors expressing Summary		

Appendix 2. Model structure for the genre “informative article”.

Sections	Titles	Moves
Heading	<i>No title provided</i>	Title Author
Introduction	<i>No title provided</i>	Why you should visit
Body	<i>No title provided</i>	Photos Photo caption Description, list or recommendation of tourist attractions Most noteworthy features of selected tourist attractions Suggested itinerary or itineraries References to historical events References to noteworthy figures connected to the place and/or to a tourist attraction Suggested activities Awards or prizes won by tourist attractions
Appendix	A Practical Guide	Links to websites with information about accommodations, facilities, and transportation

Appendix 3. Model structure for the genre “travel blog post”.

Sections	Titles	Moves
Heading	<i>No title provided</i>	Title
Introduction	<i>No title provided</i>	Photo Photo caption Blogger's thoughts (reason for writing) Links to websites
Body	<i>No title provided</i>	Graphics: photos, photo gallery, videos and/or maps Graphic caption Description or recommendation of tourist attractions Links to websites Emotional reaction
Epilogue	<i>No title provided</i>	Wrap-up

Appendix 4. Model structure for the genre “rules and regulations”.

Sections	Titles	Moves
Heading	<i>No title provided</i>	Main document title
Legal framework	Article [Nº] [Nº].	Information about the applicable legal provisions
Bookings	Article [Nº] [Nº].	Facility booking information
Pricing	Article [Nº] [Nº].	Room pricing information
Payment	Article [Nº] [Nº].	Payment information
Cancellation policy	Article [Nº] [Nº].	Cancellation information
Check-in	Article [Nº] [Nº].	Check-in information
Check-in and check-out times	Article [Nº] [Nº].	Information about check-in and check-out times
Length of stay	Article [Nº] [Nº].	Information related to length of stay
Rooms	Article [Nº] [Nº].	Information about rooms
Hours	Article [Nº] [Nº].	Information about hours
Additional services	Article [Nº] [Nº].	Information about additional services
Customer rules and responsibilities	Article [Nº] [Nº].	Information about customer rules and responsibilities
Pets	Article [Nº] [Nº].	Pets policy
Premises liability	Article [Nº] [Nº].	Premises liability information
Complaints	Article [Nº] [Nº].	Information about complaints
Right to refuse admission	Article [Nº] [Nº].	Information about the right to refuse admission