Paragraph boundaries and discourse genre: Applying Lexical Priming to Spanish written texts

Elena Martínez Caro
Universidad Complutense de Madrid ✉ ⓘ

EN Abstract: Lexical Priming (LP) is the name of the linguistic theory developed by the British linguist Michael Hoey (cf. Hoey 2005 and 2013), where corpus-linguistic concepts such as collocation and colligation are related to the experimental findings of psycholinguistics in processes of word association. The theory deals with lexical, grammatical and textual *priming* and is seen as particularly suitable for the investigation of some text-organisation issues as, for example, the boundaries across paragraphs.

Using basic ideas from LP and corpus linguistic methods, this contribution explores paragraph boundaries and their relation to the concept of discourse genre. The main aim is to discover linguistic patterns recurring in paragraph-initial position in some selected discourse genres of written Spanish, with the aid of the text processor tool Sketch Engine. A set of language patterns, lexical collocations and grammatical colligations, are explored that commonly occur in paragraph beginnings of the selected texts, both general and more specific, and some cohesive choices and tendencies noted for the genres examined.

Paragraph beginnings in writing are special and sensitive points in a text where the writer needs to consider what the most appropriate element is depending on a number of factors, among which the choice of genre is an important issue. This study is a modest attempt at demonstrating that genre is central in studies of language use based on corpora and disregarding a view of language description where languages are seen as a homogeneous whole (see Berber Sardinha 2017: 224).

Key words: lexical priming, paragraph, Spanish, written language, discourse genre

ES Límites de párrafo y género discursivo: aplicación de la teoría de la imprimación léxica a textos escritos en español

Resumen: Lexical Priming (LP, o imprimación léxica) es el nombre de la teoría lingüística desarrollada por el lingüista británico Michael Hoey (cf. Hoey 2005 y 2013), en la que ciertos conceptos de la lingüística de corpus como colocación y coligación se han puesto en relación con hallazgos experimentales de la psicolingüística en los procesos de asociación de palabras. La teoría trata sobre la imprimación léxica, gramatical y textual y puede ser considerada como particularmente adecuada para la investigación de ciertos fenómenos de la organización textual como, por ejemplo, los límites de párrafo.

Utilizando ideas básicas de LP y métodos de la lingüística de corpus, este estudio investiga los límites del párrafo y su relación con el concepto de género discursivo. El objetivo principal es descubrir patrones lingüísticos recurrentes en la posición inicial de párrafo en unos géneros discursivos seleccionados del español escrito y con la ayuda de la herramienta de procesamiento textual Sketch Engine. En esta contribución, se explora un conjunto de segmentos lingüísticos, colocaciones léxicas y coligaciones gramaticales, que aparecen de manera frecuente a principio de párrafo de los textos seleccionados, de carácter más general o específico, y se identifican algunos patrones de cohesión y demás tendencias en los géneros examinados.

Las posiciones de principio de párrafo en textos son puntos especiales y sensibles en los que el escritor necesita considerar cuál es el elemento más adecuado dependiendo de una serie de factores, entre los que la elección del género es uno de los más importantes. Este artículo pretende contribuir a demostrar que el género es un asunto central en los estudios del uso de la lengua basados en corpus y a descartar una visión de la descripción lingüística donde las lenguas son consideradas como un todo homogéneo (Berber Sardinha 2017: 224).

Palabras clave: teoría de la imprimación léxica, párrafo, español, lengua escrita, género discursivo

Contents / Sumario: 1. Introduction. 2. Background. 2.1. An outline of Lexical Priming. 2.2. Previous work on the paragraph. 3. Data, method and research goals. 4. The paragraph beginning across genres: grammatical

categories and functions (colligations). 5. The paragraph beginning across genres: selected patterns and lexical items (collocations). 6. Further markers opening paragraphs. 6.1. What punctuation indicates in paragraph-initial position. 6.2. Lexico-grammatical cohesive chains (textual collocations). 7. Concluding remarks. References.

1. Introduction

Lexical Priming (LP) is the name of the linguistic theory developed by the British linguist Michael Hoey (cf. Hoey 2005 and 2013), where the author relates corpus-linguistic concepts such as collocation and colligation to the experimental findings of psycholinguistics in processes of word association. The theory deals with lexical, grammatical and textual *priming* and is seen as particularly suitable for the investigation of some text-organisation issues as, for example, the boundaries across paragraphs.

The present paper is a contribution to discourse segmentation, especially focusing on paragraph boundaries, and its relation to the concept of discourse genre. The main aim is to discover linguistic patterns recurring in paragraph boundaries, and more specifically in paragraph-initial position, in some selected discourse genres of written Spanish. Since the main corpus used is admittedly modest, rather than expecting to provide very definite and clear results, this study should be seen as contributing to show the *potential* that the theory of LP can offer for investigating important aspects of text coherence and organisation, using tools such as the text processor software Sketch Engine.

The focus of the analysis is on recurrent language segments (words and word combinations), recognised as lexical collocations, grammatical patterns or colligations and other constructions, found in paragraph beginnings. Following LP, these language segments are seen as potential cues to the readers for recognising the boundary points of paragraphs of texts. The claim is that these words and word combinations are part of the list of items from which language users are *primed* to select when constructing paragraphs.

The structure of the paper is as follows. After this introduction, the background section introduces basic notions of the theory of LP and relevant ideas of the paragraph as a discourse unit. Section 3 describes the written Spanish data used, the use of Sketch Engine and the methodological steps followed in this study; it also presents the specific research goals of the study, following from the background and the methodology adopted. Sections 4 and 5 offer the main results in terms of colligations and collocations, respectively: grammatical categories and functions in the former, and some selected patterns and lexical items in the latter. In section 6 some further devices found in paragraph beginnings are presented, associated with punctuation and the flow of cohesive chains. The study closes with a conclusion which summarises the main findings, offers some further implications and points out some limitations of this research.

2. Background

2.1. An outline of Lexical Priming

In this section, an overview of the main ideas of LP will be presented, with special attention to the objectives of the present contribution.

LP makes use of certain well-established concepts from corpus linguistics, such as collocation and colligation, and relates them to phenomena and findings resulting from psycholinguistic experimentation.

The concept of collocation, central to LP, is usually attributed to Firth ([1951]1957), but can be found in the research of corpus linguists and other linguists working with authentic data from the 1960s to the 1990s, where the concept reaches its full potential, for instance Halliday (1966), Sinclair (1966, 1991), Leech (1974) and Stubbs (1995, 1996) (cf. Hoey 2013. See also Robbins 1980: 53-55 for an interesting overview of the notion).

Collocation can be broadly defined, as in Hoey (2013), as "the partly arbitrary tendency of words to frequently occur in each other's environment" (n.p.), for example in the frequent and natural combination of certain nouns with certain adjectives in the phrases *dark night* or *inevitable consequence*. In simple words, collocations are recurrent combinations of words (cf. Firth ([1951]1957; Sinclair 1991). Collocations are seen as pervasive in language and a key factor in accounting for the naturalness in language. For Hoey, collocation is a psycholinguistic phenomenon whose evidence can be found statistically in computer corpora. He stresses the psychological and statistical association between the words, "evidenced by their occurrence together in corpora more often than is explicable in terms of random distribution" (Hoey 2005: 3). For Hoey, some sentences are made up of interlocking collocations such that they could be said to reproduce stretches of earlier sentences.

Another fundamental notion in LP is naturally (semantic) priming. A 'priming' word may provoke a particular 'target' word. For example, a listener, previously given the word *body*, will recognise the word *heart* more quickly than if they had previously been given an unrelated word such as *trick*; in this sense, *body* primes the listener for *heart* (Hoey 2005: 8). This view has implications for language processing and child language acquisition.

A concept in LP which is central to the aims of this contribution is that of colligation, also first introduced by Firth ([1951]1957). Colligation is the grammatical pattern with which a word or word combination is associated. As Hoey (2005) notes, the basic idea of colligation is that,

> just as a lexical item may be primed to occur with another lexical item, so also it may be primed to occur in or with a particular grammatical function. [...] [C]olligation may be interpreted as going beyond traditional grammatical relations and embracing such phenomena as the postitioning of a word or word sequence within the sentence or paragraph, and even positioning within the text. (p. 43)

Put simply, colligation covers the following concepts: "the grammatical company a word or word sequence keeps (or avoids keeping) either within its own group or at a higher rank; the grammatical functions preferred or avoided by the group in which the word or word sequence participates; and the place in a sequence that a word or word sequence prefers (or avoids)" (Hoey 2005: 43).

One of the attractive aspects of the LP theory, beyond the use of familiar and well-established notions of collocation and colligation, is their interpretation of such notions from a psycholinguitic perspective. The claim is that the relevant words combinations used in language "have characteristic discourse and text-linguistic functions that listeners and readers subconsciously identify as they repeatedly hear or see them in context" (Hoey 2013: n.p.). Thus, for Hoey the textual collocations which are noted in repeated encounters are associated with particular cohesive strategies. And particular word combinations are also noted to regularly associate with certain positions in the discourse or text, such as the beginning of a paragraph, in text-initial position or in the closing turns of telephone talk (Hoey 2013; cf. also Goutsos 1997 and McGee 2014).

Another important feature of LP is the claim that, at the same time as the language user recognises the grammatical, semantic, pragmatic and textual/discourse contexts associated with the word or phrase in question, as mentioned above, s/he will also subconsciously identify other aspects of it as the genre or style it is characteristically used in. The implication is that the language user is thus primed to use certain words and word combinations in certain discourse contexts, text positions and specific genres.

With respect to the notion of genre but also to other concepts used in the theory (e.g. collocation, colligation, semantic and pragmatic associations), it is worth reproducing verbatim here the following extract with a specific example and its implications:

> Thus, if [readers] read a "serious" newspaper, they will note that *according* often co-occurs in the paper with *to a* (a collocation); this combination in turn often co-occurs in such newspapers with RESEARCH SOURCE (e.g., *according to a study*), a semantic association that newspapers manifest to a greater extent than does academic writing; they will note, too, that it is likely to occur in the context of reporting something bad (a pragmatic association not shared with academic writing). If, on the other hand, they are exposed to academic texts, they may subconsciously identify *according to* as being likely to occur with proper nouns (e.g., *according to Smith*), a colligation that academic texts manifest more commonly than newspapers. (Hoey 2013: n.p.)

LP is particularly suitable for studying issues of text production and processing. Hoey argues that lexis is intimately bound up with decisions of discourse organisation. Words or combinations of words may be textually primed in two ways and each of these is in turn associated with one of the two textual notions discussed above. Thus, textual collocation relates to words and their combinations being primed positively or negatively to participate in cohesive chains of different and distinctive types; and textual colligation relates to words and their combinations occurring (or avoid occurring) at the beginning or end of independently recognised discourse units, for instance, the sentence, the paragraph, the speech turn (Hoey 2005: 114-5). Hoey also distinguishes between cohesive chains and cohesive links: whereas a cohesive chain is defined as "three or more items linked by textual collocation, [...] including repetition", a cohesive link "occurs when just 2 items are so connected" (Hoey 2005: 117). Thus, cohesive links are understood as short cohesive chains.

Also within the textual perspective it is worth mentioning that priming for theme or rheme is common, that is paying special attention to what goes in the sentence initial element (theme) in contrast to the rest of the sentence (rheme).

Finally, a recent compilation of applications and advances in LP theory is Pace-Sigge & Patterson (2017). This includes the relevant study by Berber Sardinha (2017) on spoken and written American English, which combines the concept of collocation with the multidimensional framework for register variation by Biber (see, for instance, Biber & Conrad 2019), and where the author examines nine dimensions of collocations. Whereas research on lexical priming so far has been mostly on written discourse, Pace-Sigge (2013) is a good illustration of the theory being applied to the spoken (English) language.

2.2. Previous work on the paragraph

The literature on the paragraph is far from abundant, outside the scope of writing manuals and contributions on language teaching and learning. The studies that exist derive from different traditions, by and large work on discourse analysis and text linguistics, studies from the field of psycholinguistics and related research from the areas of computational and corpus linguistics.

The decades of the 1960s, 1970s and 1980s produced some important research on the paragraph from these different perspectives. In a previous contribution (Martínez Caro 2014: 193-197) important ideas on the paragraph from several of these studies were reported. Examples of the discourse and text analysis tradition are Brown & Yule (1983), van Dijk (1977), Givón (1983), Grimes (1975), Hinds (1977, 1978), Hoey (1983), Longacre

(1979, 1980), and the more recent work by Hannay & Mackenzie ([2002]2017), Ji (2002, 2008) and Siepmann et al. (2008). For Spanish, more specifically, the studies by Obdržálková (2013), Martínez Caro (2014) and Vera Luján (2012), for example, can be mentioned (see also Borreguero Zuloaga 2021). Studies in this tradition share the view that the paragraph is a structural unit, with syntactic, semantic and pragmatic features, which helps organise the discourse.

In the domain of psycholinguistics, the paragraph is recognised as a cognitive and psychological unit, as in the research by Young and Becker (see, for example, Koen, Becker & Young 1969). Several studies on the paragraph, for example Bustos Gisbert (2012), González Cobas (2014) and Hoey (2005), report on experiments with paragraphing with subjects who were asked to break an unparagraphed text and comment on the implications that the results of such experiments have for the psychological nature of the paragraph. According to Hoey (2005), the breaks that these informants favoured were not made because they recognised paragraph structural units "but because they were responding to their perceptions of the way the text as a whole was structured" (p. 134).

A few contributions on the paragraph prefer to adopt a comprehensive or multifunctional perspective to the paragraph, notably Bustos Gisbert (2012) and González Cobas (2014), since they consider that the definition of the paragraph cannot be complete unless this unit is considered from a number of different dimensions.

Hoey (2005) devotes his chapter 7 to ideas on textual priming, where he claims that certain lexical items may be primed (either positively or negatively) to appear "at the beginning or end of an independently recognised chunk of text", including naturally the paragraph (p. 129).

Hoey & O'Donnell (2008) and Bustos Gisbert (2012), among others, argue against the traditional view of the paragraph as being characterised by a topic sentence at the beginning of paragraphs, and which has prevailed in the language teaching literature for some years. Hoey & O'Donnell (2008), in particular, conclude that the topic sentence does not exist but, instead, writers make paragraphing decisions based on "statistical probablilities". For instance, the reintroduction of a full name (rather than an anaphoric reference to that name) or the presence of certain expressions (e.g. *as a [job], fundamentally*, etc.) are more likely to signify the beginning of a new paragraph rather than topic shift.

Goutsos (1997), in turn, notes that over half (54%) of paragraph beginnings occur with "potential signals of topic framing or topic introduction", for instance, elements such as adjuncts, conjunctions, predicators, interrogatives, and topic introduction techniques. The only remaining class (definite subjects) include 'encapsulating nominals' (both grammatical –*that*– or anaphoric nouns –*the idea*), which can also be related to topic framing and closure. Other contributions that have looked at the precise marking of the beginning of paragraphs and the way the thematic transition is indicated in that position for Spanish are Bustos Gisbert (2012), González Cobas (2010) and Martínez Caro (2014). The last of these, based on a corpus analysis of news magazine texts in Spanish, offered evidence for the recognition of paragraphs as structural units of the discourse and concluded that the paragraph tends to be marked in a more evident and conclusive way in its beginning than its end.

Finally, McGee's (2014) research on argumentative texts concludes that the paragraph break tends not to occur alone in this type of text but that multiple signal marking appears to be the norm. O'Donnell et al. (2012) exploit new corpus-based methods to identify textual colligations, lexical items and combinations in common textual locations, including the beginning or end of texts, paragraphs and sentences.

3. Data, method and research goals

The Spanish data on which this contribution is based derives from two main sources: (a) my own, and admittedly modest, compilation of Spanish written texts from a selection of different genres, comprising news magazines, travel blogs, film reviews and cooking blogs (including recipes); and (b) a subcomponent of the European Spanish Web 2011 Corpus (eseuTenTen11) containing newspaper reports, and obtained from Sketch Engine.

The software used to exploit and manage the texts is the corpus management tool Sketch Engine. More specifically, this tool was used to digitalise the self-compiled corpus tagged for paragraph boundaries, to carry out the extraction of the relevant dataset, and to operate different searching combinations regarding the lexical items, collocations and colligations.

In Sketch Engine, the main function used was Concordance, and within that, CQL (*Corpus Query Language*); in CQL the query labels used were *word*, *lemma* and *tag*, and the structure tag for paragraph. The search structures in CQL are structures which refer to any parts or sections into which a corpus might be divided, such as sentences, paragraphs, and documents. Thus, for instance, a corpus can be labelled for direct speech through CQL to make it possible to limit the search to only direct speech in that corpus. Most corpora in Sketch Engine will use at least structures such as *s*, *p* and *doc*, corresponding to sentences, paragraphs and documents, respectively. Apart from this, CQL is useful in general for searching for complex lexical structures and grammatical grammatical categories, in a single search.

The methodological steps followed for this investigation were as follows. I first manually tagged my self-compiled texts with the tag label <p> for beginning of paragraph. Second, I created the corpora by manually adding the texts to each of the corresponding corpus per genre (film, travel, etc.). Once the paragraph-tagged texts were uploaded, I did some (limited) searches always introduced by <p>. Finally, a subcorpus of the eseuTenTen11 corpus was also created on Sketch Engine to do some manageable searches for comparison with the self-compiled texts.

It has to be noted that this subcorpus from eseuTenTen11 is not as homogeneous as one would have wished and expected, in that, representing the language of one of the national newspapers (*El Mundo*), it

contains different subgenres within this general genre, not only news reports but also interviews and more descriptive texts. All in all, it is also true that the majority of the texts in TenTen11-ElMundo are news reports.

In the following table the volume of the self-compiled corpus in terms of the number of paragraphs is provided.

Table 1. Volume of multi-genre corpus in number of paragraphs

	Corpus and genre	No. of paragraphs
Own compilation of Spanish written texts	News magazines (news) Travel blogs (travel) Film reviews (film) Cooking recipes (cooking)	547 392 120 109
	Total self-compiled texts	**1,168**

The number of paragraphs in the subcorpus from TenTen11 from *El Mundo* newspaper was 585,551.

Taking into consideration a number of issues raised in the background section and the methodology presented in the present section, the main objectives of this contribution are the following:

a) To identify lexical and grammatical patterns (colligations and collocations) recurring in paragraph beginnings in some selected genres of written Spanish;
b) to reach useful generalisations as to which patterns (colligations and collocations) characterise these boundaries in the different selected genres of Spanish writing, and stress the differences found; and
c) using central ideas from LP, to observe the relevance and interest of this theory for text segmentation issues and in particular exploring the beginning of the paragraph.

This paper offers an overview of paragraph-initial words, collocations and colligations in four selected genres of Spanish writing, without intending to provide a full quantitative account of these elements.

As an illustration of how a search looks like in Sketch Engine, with the selected corpora, the following figure is given where the subcorpus searched was the eseuTenTen11 with the tag for verb in paragraph-initial position (see section 4).
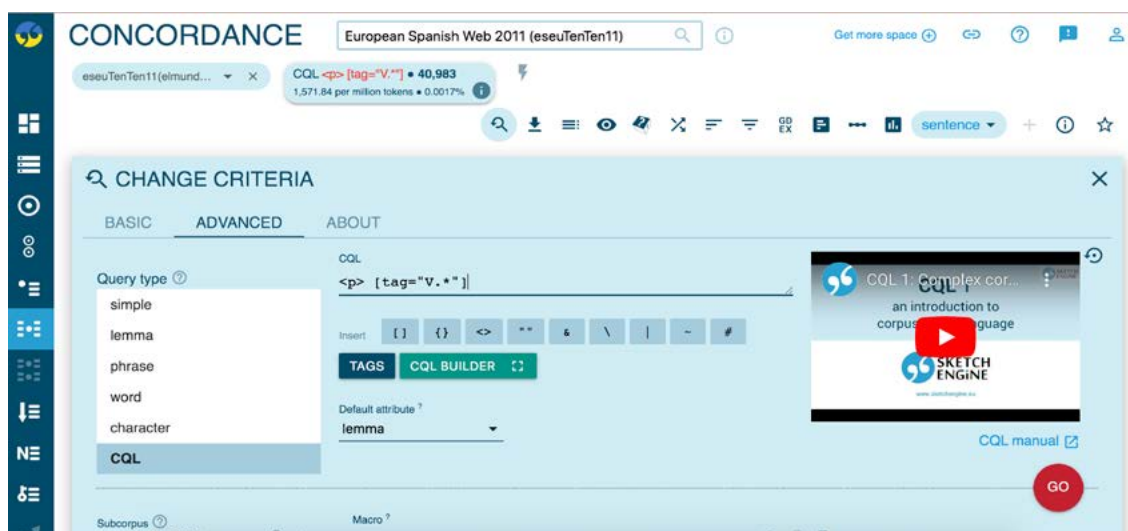

Figure 1. Example of a concordance search with Sketch Engine

More specific searches can be carried out in Sketch Engine combining the search of a grammatical category (for instance, conjunction) in a certain corpus with a specific item from that category (for example, the conjunction *pero*). These have proved to be indeed useful when searching for specific collocations and colligations in paragraph beginnings.

4. The paragraph beginning across genres: grammatical categories and functions (colligations)

The starting point of the analysis was examining the corpus for the category of words or combination of words occurring in paragraph beginnings and trying to correlate this with the corresponding syntactic functions. In the LP theory, this kind of search has to do with the notion of colligation, which as indicated above, deals with

the grammatical pattern with which a word or word combination is associated, where 'grammatical pattern' refers to grammatical position or grammatical function (cf. 2.1).

The first corpus to be looked at was the selected component of the eseuTenTen11 Corpus, containing news discourse from the newspaper *El Mundo*, as we saw. This seemed a good move for a first analysis of the results, given the compact format of this subcorpus.

This first search gave the following results for 70% of the paragraphs analysed for the grammatical categories representing the first element at the beginning of paragraph (the other 30% offering less definite results): nouns 21.78%, determiners (followed by nouns) 15.56%, prepositions 10.33%, verbs 7%, adverbs 5.64%, conjunctions 3.75%, pronouns 2.11% and adjectives 1.79%. A frequently recurring unit in the remaining 30% (almost 14%) was numerals, which were discarded as they often corresponded to numbering of sections or points.

A more simplified version of this set of results, once the most important categories had been grouped together, is presented below in Table 2. Here, the category NOUN comprises not only nouns, but also pronouns and noun phrases containing determiners; and the function ADVERBIAL contains both adverb and prepositional phrases.

Note that some of these categories had to be filtered to discard unwanted elements or else add some new element to them. For instance, the category PRONOUN was cleaned of the examples with initial particle *se*, which I treated as part of a (pronominal) verb with *se* (as in *Se quedaron*), and thus added to the category VERB. The examples of the category NOUN in paragraph beginnings were cleaned of the abbreviations P. and R. (standing for *pregunta [question]* and *respuesta [reply]*) characteristic of the genre interview.

Table 2. Three main grammatical categories in paragraph-initial position in newspaper discourse (TenTen11-ElMundo subcorpus): Frequency and proportion. Simplified version

Nominal (noun / det + noun / pronoun)	231,032	39.46 %
Adverbial	93,501	15.97 %
Verbal element	41,004	7.00 %
Total no. of paragraphs	585,551	100 %

A look at both the general results and those in the simplified version in Table 2 indicates that a good proportion of paragraphs in newspaper discourse (almost 40%) are introduced by nominal elements, which are initial-paragraph nouns (nearly 22%; cf. example 1a), noun phrases taking one determiner (15.40%), or much less frequently two (see 1b); and a pronoun as paragraph-initiating element (only over 2%; cf. 1c). The next most frequently found element was adverbials (or adjuncts), with a frequency of nearly 16%, in which both initial adverbs (over 5.5%; cf. 2a) or prepositions (that is prepositional phrases, over 10%; cf. 2b) were included. Thirdly, with less than half the frequency of adverbial elements, paragraphs in this corpus are introduced with a verbal element (cf. 3a), including as said above those introduced by the pronominal particle *se* (3b).

(1) a. <p> **Benítez**, ahora con 52 años, tiene que ser asistido por Ivonne en su vida diaria, que transcurre en una humilde vivienda de un barrio de las afueras de San Juan cercano al municipio de Carolina. (TenTen11-ElMundo, 9)
     b. <p> **Toda mi carrera** profesional se ha desarrollado en el ámbito de la alta tecnología, casi siempre relacionada de alguna manera con el marketing. (TenTen11-ElMundo, 968)
     c. <p> **Esto** es lo que el consumidor de alimentos ecológicos suele encontrarse en numerosos herbolarios y centros de venta especializados. (TenTen11-ElMundo, 8)
(2) a. <p> **Así**, el guitarrista Keith Richards se lesionó seriamente en la cabeza al caerse de un árbol en las Islas Fiyi. (TenTen11-ElMundo, 2)
     b. <p> **En** los últimos años, su régimen ha sido criticado por "autoritario" por las organizaciones de defensa de los derechos del hombre. (TenTen11-ElMundo, 8)
(3) a. <p> **Extraigo** esta curiosa anotación del segundo viaje de Darwin a Tierra de Fuego donde el perspicaz naturalista observó el 'peligroso' desdén de los nativos hacia la propiedad privada. (TenTen11-ElMundo, 326)
     b. <p> **Se** podría hacer un comedor para banquetes, bautizos y bodas. Le daría empaque al hospital. (TenTen11-ElMundo, 15)

The same procedure used for the TenTen11-ElMundo subcorpus was employed in the multi-genre corpus. A decision was made to apply the main three categories found for the former corpus, that is nominal, adverbial and verbal elements at the beginning of the paragraph. The results showed that these three elements represented between 72% and 82% of all the elements in paragraph-initial position, which justified focusing on these three categories. The specific frequencies in each genre are provided in Table 3.

One limitation of these searches has to be mentioned. As we know, grammatical categories do not always match the obvious corresponding syntactic function. For instance, a nominal combination (noun phrase) may be used as an adverbial (with adjunct function), and not with the more obvious function of subject or object/complement, as in the example ***Un año más tarde***, *en 1994, fue el Príncipe de Asturias el que viajó a Sudáfri-*

ca para participar en la ceremonia del nombramiento de Mandela como presidente sudafricano. (News, Mandela). That is, there is a certain mismatch between the categories used in here as the three main types. This was solved in the more manageable self-compiled corpus by examining manually all examples of the three categories and discarding those which did not correspond to the *prototypical* grammatical category. This includes the nominal elements with adverbial function mentioned and verbal elements not functioning as main verb of the clause.

For obvious reasons, the same procedure could not be carried out in the much larger TenTen11-ElMundo subcorpus. In this case, however, it is believed that, given that the examples represent a small proportion within the large total, they do not distort the general figures much.

Table 3. Main grammatical categories in paragraph beginnings across genres

	Film	Travel	News	Cooking
Nominal (det + noun / noun / pronoun)	52.50 %	40.82 %	49.18 %	37.61 %
Adverbial	18.33 %	24.23 %	25.78 %	15.60 %
Verb	11.66 %	12.76 %	7.13 %	22.02 %
Total no of paragraphs	100 % (120)	100 % (392)	100 % (547)	100 % (109)

A series of conclusions can be extracted from the results in Table 3. First, nouns and noun phrases are predominant as paragraph beginnings in all four genres with higher frequencies in the film and news texts (around 50%). Adverbials are the second main grammatical colligation, with a higher use (of around 25%) in news and travel texts. Finally, it is noticeable that in the cooking blogs the verb is the second most used category, over adverbials; the instructional character of this genre can explain this preference of initial verbs in imperative sentences over initial adverbs. As can be seen, the results for the news genre in this multi-genre corpus are not very different from the ones obtained in the TenTen-El Mundo corpus, which belongs to the same genre.

Examples of nominal elements are provided in (4), both a single noun or a full noun phrase introduced by a determiner. And (5) and (6) illustrate the use of an adverbial and a verb in paragraph-initial position.

(4)   a. <p> **Chazelle** vuelve a contar con su fiel Justin Hurwitz para componer una banda sonora omnipresente y distintiva. (Film, Babylon)
      b. <p> **El Mirador de la Albufera** es un lugar para disfrutar del bonito paisaje y la bella estampa que nos deja los muelles de madera de su embarcadero, con la preciosa luz que tenemos al atardecer. (Travel, Albufera)

(5)   a. <p> **Realmente**, el cerebro tiende más hacia el chocolate que hacia la manzana. (News, Alimentos y cerebro)

(6)   b. <p> **Pelar** las cebollas, cortarlas en tiras finas y sofreír en una olla más bien grande (la lombarda al principio abulta bastante) con 3 cucharadas de aceite de oliva y un poco de sal hasta que empiecen a dorarse. (Cooking, Lombarda)

Variation in the domain of the noun phrase

In relation to the category NOUN in paragraph beginnings, it is interesting to examine the more specific semantic sets to which some of these nouns belong. For instance, Hoey (2005: 139ff) mentions the sets SURNAME (NAME OF PERSON) and PLACE NAME. A look at the paragraph-initial nouns in the data from this perspective indicates some interesting differences between the genres. In the news texts, there was an important amount of nouns both expressing surnames, alone or preceded by first names (when introduced for the first time), and place names, with a similar proportion (nearly 40%). Some examples are presented in (7), for surnames, and (8), for place names.

(7)   a. <p> **Sanz** reconoce que después de cuatro ediciones completadas sigue conservando el mismo "respeto" por la prueba. (News, Reina desierto)
      b. <p> **Arafat** era el único dirigente con la legitimidad suficiente para aglutinar el conjunto del pueblo palestino, [...] (News, Muere Arafat)
      c. <p> **Luis García**, responsable del equipo de delitos de pederastia infantil, alerta además de una práctica que está cada vez más extendida, el 'sexting'. Este fenómeno, consiste en la distribución de este tipo de contenidos entre los propios menores. (News, Anonimato en Internet)

(8)   a. <p> **España** continúa compitiendo contra sus propias puntuaciones en los transplantes de órganos. (News, Transplantes en ESP)
      b. <p> **Arabia Saudí** es uno de los países más estrictos en lo que a cuestiones religiosas se refiere. (News, La Meca)

Whereas in film reviews the tendency is for surnames to be primed positively at the beginning of paragraphs (in nearly 36% of instances), in travel blogs it is, as expected, place names that are more predomi-

nantly found than other nouns with further semantic sets in a clear proportion (over 43%). In cooking both surnames and place names are negatively primed to occur in paragraph-initial position. Examples of surnames from film writing are given in (9) and of place names from travel writing in (10).

(9) a. <p> **Aronofsky** adapta para su película una obra de teatro homónima de Samuel D. Hunter. (Film, The whale)
b. <p> **Channing Tatum** debuta en la dirección con una historia basada parcialmente en una historia real: la suya propia. (Film, Dog)

(10) a. <p> **Japón** no es un país barato, pero los costes de viajar al país nipón ya no son lo que eran hace años, [...] (Travel, Japón)
b. <p> **Santander** es una de las ciudades más bonitas de España [...] (Travel, Cantabria)

5. The paragraph beginning across genres: selected patterns and lexical items (collocations)

The next step in the corpus analysis was to try and identify certain collocations recurring in paragraph beginnings which completed the grammatical patterns presented in the previous section. Partly inspired by previous research and intuition (cf. Hoey 2005: 133), I carried out some random searches of words and collocations to examine their frequency in the data. Because the focus was on the contrast among genres and because of the difference in the volume of the texts, this part of the analysis was limited to the multi-genre corpus.

The present section presents the results of this analysis for the small selection of items that proved to be more important in frequency resulting from those searches. Given the specific character of the words and collocations, resulting in more modest frequencies than the categories described in section 4, the perspective adopted will be qualitative rather than quantitative.

The first few patterns considered here are drawn from previous research in Martínez Caro (2014), which used the same subcorpus for the news genre as in this paper. In similar lines to Hoey's ideas on LP and their proposal of a local grammar of English, Hunston & Sinclair (1999) notice the importance of patterns in the grammar and lexis of languages, in that "every word can be described in terms of the patterns it typically occurs in", and "words that share a particular pattern typically also share a meaning" (p. 83).

The first collocational pattern to consider here in paragraph beginnings is formed of a COPULAR VERB collocating with a STANCE ADJECTIVE (for the notion of stance, cf. Biber et al. 1999: 966ff) expressing some evaluative or attitudinal meaning. In my dataset, this evaluative pattern was found to be more important in news magazines, mainly, and film reviews. A few examples have been also found in the travel texts, whereas it is practically nonexistent in cooking. Some examples of this pattern in the news and film texts are presented in (11).

(11) a. <p> **Es incuestionable** la manera en que estas aplicaciones están cambiando el concepto de ver las cosas. (News, El móvil)
b. <p> **Es muy difícil** imaginar el dolor de las madres a las que les arrebataron a sus hijos, porque alguien decidió que era mejor entregárselos a otras familias. (News, Robo bebés)
c. <p> **No es habitual** que un cineasta apueste tanto por su propia valía, y ahí reside la valía de esta primera secuela, [...] (Film, Avatar)
d. <p> En este sentido, **parece evidente** que estamos ante una cinta de autor. (Film, Tomorrow)

This pattern is used by the writer, not only to evaluate some idea related to the discussion, but also as an introductory device of an element which wants to be given a certain degree of focus, and consequently appears towards the end of the sentence.

Secondly, I looked at the CONDITIONAL CLAUSE pattern introduced by the SUBORDINATOR *si*. This was searched in the present corpus in possible combination with a preceding adverb (*entonces si*) or another conjunction (*pero si*). The data indicates that this pattern is positively primed to open a paragraph especially in the travel genre, followed by the film texts, and news magazines. It is, however, negatively primed in cooking texts. The examples in (12) illustrate its use in the first three genres.

(12) a. <p> **Si** tienes curiosidad, puedes visitar la Catedral ortodoxa griega de San Andrés y San Demetrio, cerca del metro Colombia, y la Iglesia Ortodoxa Rumana de España y Portugal [...], en Carabanchel. (Travel, Madrid)
b. <p> **Y si** eres de los que repiten, quizás quieras planear una visita por el Japón más desconocido, el sur, con ciudades como Fukuoka. (Travel, Japón)
c. <p> **Si** convenimos en que el boxeo es el deporte que más lecciones de vida nos ha dado (siempre a través de su dramatización fílmica, claro está), entonces no se debe ignorar la importancia que ha tenido el peso de la derrota en dichas enseñanzas. (Film; Small, slow but steady)
d. <p> **Pero si** existen unos auténticos protagonistas en la Fundación También estos son los más pequeños. (News, Deporte adaptado)

Since paragraphs often introduce a new topic or subtopic in the narration, description, argumentation or other, identifying constructions used to introduce an important and relevant idea in the discussion are good devices at the beginning of paragraphs. One productive instance of these constructions is the one intro-

duced by the indefinite *otro (/otra/s)* in the collocations *otro de* + DETERMINER / *otro* + NOUN. As with the subordinator *si*, this device has been found to be positively primed in the film, travel and news genres, and negatively primed in cooking. (13) provides a few examples.

(13)　a. <p> **Otro** de los elementos más atractivos de la película es la hermosa carta de amor al cine que Mendes traslada desde su guión [...] (Film, Imperio de la luz)
　　　b. <p> **Otra** de las opciones más utilizadas por los turistas que visitan Egipto y quieren combinar cultura con descanso es hacer un crucero por el Nilo. (Travel, Egipto)
　　　c. <p> **Otro** restaurante conocido por sus famosos reservados y por donde pasaban políticos y financieros es Balzac, [...] (News, Clásicos lujo)

Evidential expressions, expressing the source of the information contained in the relevant sentence, were also searched in the opening of paragraphs. A recurrent pattern with this meaning is the prepositional phrase *según* before a noun phrase indicating the source of the information expressed. In this case, as the data indicates, it is only in the news-magazine genre that the collocation *según* + SOURCE is positively primed to occur in paragraph beginnings. In the other three genres, *según* + SOURCE is negatively primed, not offering any instance of this collocation. The head nouns in the phrase tend to be general nouns not referring to a particular person or institution. See the examples in (14).

(14)　a. <p> **Según** datos del Comisionado para el Mercado de Tabacos (CMT), [...] (News, Cigarrillo electrónico)
　　　b. <p> **Según** los expertos del IAQ, [...] (News, Anonimato en Internet)
　　　c. <p> **Según** el director, [...] (News, Palace)
　　　d. <p> **Según** el informe anual que hace las Naciones Unidas [...] (News, Haití)
　　　e. <p> **Según** esta asociación, [...] (News, Cáncer en ESP)

Thematic adverbs and coordinators

The following discussion deals with the use of certain adverbs and conjunctions (coordinators) in paragraph beginnings. These elements are naturally primed to appear in the *theme*-sentence position and, where they serve to connect with previous paragraphs, are particularly useful in paragragraph-initial boundaries. A search of various of these elements was carried out in the multi-genre corpus and the results indicate that the most productive elements are the adverbs *también* and *además*, and the coordinators *y* and *pero*.

In terms of the genres where they are mostly found, *además* and *también* are primed more in travel texts. *También* also occurs in films and news, and *además* in news. Both these two connecting adverbs are only occasionally found in the cooking genre. Whereas *también* can occur in combination with other preceding elements, *además* is always paragraph-initial. (15) and (16) give some examples of these two elements in these genres.

(15)　a. <p> **También** se hace imprescindible visitar Nara desde Osaka, y [...] (Travel, Japón)
　　　b. <p> Por último, **también** se puede visitar la iglesia de Ciudad Lineal, [...] (Travel, Madrid)
　　　c. <p> **También** su largo clímax final plantea hasta qué punto funciona todo el dispositivo, [...] (Film, Avatar)
　　　d. <p> **También** han pasado a la historia las camareras con cofia del recordado Príncipe de Viana, [...] (News, Clásicos lujo)
(16)　a. <p> **Además**, junto al mirador encontramos el Embarcadero de la Gola de Pujol, [...] (Travel, Albufera de Valencia)
　　　b. <p> **Además**, otra de las ventajas de la recuperación del plástico es que este puede ser utilizado como materia prima de nuevo, [...] (News, Plástico vertedero)

The coordinators *y* and *pero* must initiate the paragraph, not allowing other preceding elements. Whereas *y* is lexically primed in paragraph-initial position in all four genres, including exceptionally the cooking texts, *pero* is very evidently found in news texts, mainly, but also the travel texts. (17) and (18) provide examples of these two coordinators.

(17)　a. <p> **Y** para continuar os recomiendo probar los aceites de Sierra Magina [...] (Cooking, Croquetas de mi madre)
　　　b. <p> **Y** como buen mago, el hacedor de sueños imposibles le otorga a la música de sus películas un espacio narrativo fundamental. (Film, The Fabelman)
　　　c. <p> **Y** ya visitando los extremos de Japón, en el norte de la isla de Hokkaido encuentras una impresionante zona volcánica de numeantes onsen, [...] (Travel, Japón)
　　　d. <p> **Y** es que la falta de actividad física o lo que es lo mismo, el sedentarismo se ha convertido ya en el cuarto factor de riesgo de mortalidad en todo el mundo. (News, Deporte para no morir)
(18)　a. <p> Paco de Lucía era un músico admirado en todo el planeta cuando grabó su primer gran éxito. [...]
　　　<p> **Pero** Paco también tenía carácter. [...] (News, Paco de Lucía)

b. <p> **Pero** el desfavoreceder contexto económico de estos años no es el único factor que ha influido en el declive de esta industria. (News, Cigarrillo electr.)

c. <p> **Pero** la gamba roja es escasa, y cara, y Denia no lo podía fiar todo a ese producto, [...] (Travel, Denia)

It is interesting to observe that, whereas paragraph-initial *pero* is primed to occur followed by the sentence subject, the coordinator *y* is followed instead by non-subject elements, for instance phrase or clause adjuncts, as can be seen in the examples in (17) and (18).

6. Further markers opening paragraphs

6.1. What punctuation indicates in paragraph-initial position

Previous work has noted the use of rhetorical questions in paragraph beginnings in the news-magazine genre (Martínez Caro 2014), and this construction was also examined here. To this end, I used the punctuation tag in Sketch Engine, formulated as *F*, to identify rhetorical questions initiated by the punctuation mark ¿, characteristic of interrogatives in Spanish. At the same time, the formula *F* also yielded examples of reported speech, introduced by the double quotation symbol (").

A number of rhetorical questions were found in my data, mostly in travel blog texts, followed in frequency by the film and news genres. Examples are given in (19).

(19)   a. <p> **¿Sabías que** en Madrid hay un lugar de esos poco conocidos, que alberga el que quizá sea el monumento más antiguo de la ciudad de Madrid? (Travel, Madrid)

b. <p> **¿Hará algo Rajoy** para evitarlo? Esperemos que sí, [...] (News, España 2013)

c. <p> **¿Si un árbol** se desploma en un bosque donde no hay nadie, entonces dicha caída emite algún tipo de sonido? (Film, Decision to leave)

Martínez Caro (2014: 209) argued that rhetorical questions are found to indicate a more important break in the discourse flow than other markers of the beginning of paragraphs, often to open a sequence of paragraphs (or episode; cf. Dik 1997, Butler 2003: 310-311) about a certain discourse (sub)topic.

As for quoted speech, as expected given the characteristics and purpose of the genre, news texts make an important use of this device in opening paragraphs. In contrast, this same pattern is negatively primed in film, travel and cooking texts.

(20)   a. <p> "**Cuando se** come chocolate, se dispara el nivel de azúcar en sangre y el cerebro produce con rapidez serotonina ... ", añadió Kiefer. (News, Alimentos y cerebro)

b. <p> "**5 kilos** de comida es algo simbólico para nosotros que estamos medio bien --concluye Pipi--, pero para la gente que lo está pasando realmente mal es algo importante". (News, Solidaridad ciudadana)

c. <p> "**He llamado** mil veces para preguntar por qué aún no me ha llegado el dinero de la beca que me concedieron en septiembre. ... " comenta Álvaro, alumno de la Facultad de Comunicación de Sevilla. (News, becas)

These two most found signs of punctuation at the beginning of paragraph, ¿ and ", may also coincide, indicating a rhetorical question which has been reported by someone else, rather than the writer him/herself.

6.2. Lexico-grammatical cohesive chains (textual collocations)

As we saw in section 2.1, Hoey argues that words and their combinations may be positively or negatively primed to intervene in cohesive chains of different types, which he calls textual collocations.

A lexical cohesive device potentially marking the beginning of paragraphs is the (total or partial) repetition of a name or another noun, rather than the omission of the noun altogether once its referent has been introduced in the previous discourse. Naturally, one must remember that Spanish, being a so-called pro-drop language, does not make use of anaphoric subject pronouns referring to nouns, unless they express contrast or emphasis (cf. Martínez Caro 1999).

Pronominalisation can be used instead of repetition, but generally speaking, it is avoided in paragraph beginnings, as it is not considered explicit enough for this textually significant position (Hannay & Mackenzie [2002]2017: 160). Conversely, the repetition of an explicit noun phrase, rather than the use of the corresponding anaphoric cohesive pronoun of that phrase, is a more common indicator of paragraph beginnings, especially in some genres, as mentioned by Hoey (2005) and McGee (2014) (cf. also Hannay & Mackenzie [2002]2017: 159-161 on the 'reference cohesion scale').

Noun phrases in paragraph beginnings can refer to human referents related to the story; this is commonly found in the news-magazine genre, especially, and the film genre, less obviously. The beginnings of paragraphs in these texts often contain textual collocations with these human referents and serve as cohesive chains.

A good example of this 'continuous thematic progression' in texts (cf. e.g. Downing 2015: 228) is in one of the news-magazine texts, about the passing of the British philosopher and writer Christopher Hitchens. Here the referent is *maintained* and repeated in the paragraph beginning in a few of the thirteen paragraphs of which the text consists, by an explicit repetition or partial repetition of the noun phrase in question (cf. Hannay

& Mackenzie [2002]2017: 159-160). These paragraphs coincide with the two "extreme positions" of the text: at the beginning of the text, where Hitchens is introduced, and in the paragraph before the last one, which starts the closure of the text. In this near-to-last paragraph (<p12>), the writer resumes the main topic reference in paragraph-beginning position by choosing to express it with the full noun phrase containing its first name and surname. (21) shows the use of a few paragraph beginnings from this text.

(21) <p1> **El pensador británico Christopher Eric Hitchens**, filósofo, escritor y hasta periodista, [...] murió el 15 de diciembre de 2011 [...] Tenía 62 años.
<p2> *Hitch,* como le conocía muchísima gente, era ateo. [...]
<p3> **Pero el gran *Hitch*** hizo algo más que eso. [...]
[...]
<p12> **Christopher Hitchens** murió sin la muerte horrible que le deseaban los compasivos *fieles;* murió [...], y se fue después de regalarnos a todos este libro inaudito, *Mortalidad,* [...]
<p13> Y, sobre todo, la lección definitiva de alguien que murió como había vivido: combatiendo, pero en paz. (News, Hitchens)

In travel blogs, in contrast, referents are more commonly used to indicate place names, as we saw above (section 4), thus expressing inanimate referents, which also commonly begin paragraphs and where repetition of the noun phrase is used as part of a cohesive chain. An example is (22).

(22) <p>¿Dónde está Nowa Huta?
<p> **Nowa Huta** se encuentra al este de Cracovia, algo retirada del centro, así que para llegar hasta allí, os recomendamos utilizar el transporte público. (Travel, Polonia)

We can compare the cohesive effect created between the question and the answer in the previous interaction with the less cohesive and effective extract that follows, where the repetition has been removed:

(22') <p>¿Dónde está Nowa Huta?
<p> Se encuentra al este de Cracovia, algo retirada del centro, así que para llegar hasta allí, os recomendamos utilizar el transporte público.

The repetition of the noun phrase in (22) with the explicit mention of the topic, rather than the zero reference in (22'), is a more appropriate option for paragraph beginnings, which are expected to be more explicitly marked than other paragraph positions (cf. González Cobas 2014: 181ff).
Textual collocations include other, non-nominal, patterns, like the parallel sequences in the beginning of the two following subsequent paragraphs from the Hitchens text in news, as in example (23).

(23) <p> **Lo asombroso no es** que haya gente como la que ha escrito esa salvajada, que contradice una por una todas las virtudes cristianas. [...]
<p> **Lo asombroso es** el cachondeíto que el propio Hitchens, [...], se gasta para comentar ese y otros párrafos igualmente impagables. [...] (News, Hitchens)

Paragraph beginnings can contain nominal elements with an inferential association with a previously mentioned referent in the cohesive chain. In (24), for example, paragraph initial *[e]sta etapa prehistórica* refers to the previous paragraph beginning *[e]l Mesolítico*.

(24) <p> **El mesolítico** es el periodo de transición entre el paleolítico y el neolítico que duró desde hace 10.000 años hasta hace 5.000. Los hábitos de las culturas del Mesolítico eran básicamente nómadas, con asentamientos estacionales de invierno y campamentos de verano.
<p> **Esta etapa prehistórica** terminó con la llegada de la agricultura y la ganadería, procedente del Próximo Oriente. (News, Cazador Mesolítico)

Another device to create a cohesive chain in which an explicit referent, rather than the omission of that referent, is involved is the so-called 'linear thematic progression' (cf. e.g. Downing 2015: 228), where a topic element mentioned towards the end of the previous paragraph is (totally or partially) repeated in the paragraph-initial position of the following. Three examples of this phenomenon from the news-magazine genre are given in (25-27).

(25) [...] La vitamina B se encuentra en la levadura, trigo, hígado de ternera y vacuno, así como en aguacates y *plátanos*.
<p> **Los plátanos** son un arma eficaz contra el estrés. Contiene la proteína triptófano, que es convertida por el cuerpo en serotonina, la hormona de la felicidad. (News, Alimentos y cerebro)
(26) [...] "Ha pasado casi un año de la reunión que mantuvimos con el Ministerio de Justicia y ya se ha demostrado que *la oficina* que se habilitó aquí es inoperante totalmente, se creó para quitarnos de encima".
<p> **Esta oficina** fue creada por parte del ministerio para censar a los afectados y montar un banco de ADN, las pruebas por supuesto corren a cuenta de cada familiar. [...] (News, Robo bebés)

(27) <p> No sirvió de nada. *Hitch* [...] confió con muchísima más vehemencia en el trabajo de *los médi-cos*, sobre todo de algunos a los que se refiere con verdadero cariño.
      <p> **Médicos** que al final no lograron salvarlo, como él mismo dice, [...] (News, Hitchens)

Finally, linear progression need not involve repetition but the choice of a semantically-inferred noun phrase, with respect to the previously-introduced information at the end of the previous paragraph, as in (28).

(28) [...] todas ellas producto del desgaste óseo de un hombre que ha hecho siempre mucho deporte y que, a principios de este mes de enero, *cumplió 75 años*.
      <p> **Una fecha redonda** la del cumpleaños real, que la televisión pública española ha querido aprovechar para rendir un homenaje al jefe del Estado con una entrevista realizada [...] (News, Rey 75)

7. Concluding remarks

This contribution has attempted to demonstrate that paragraph beginnings in writing are special and sensitive points in a text where the writer needs to consider what the most appropriate element is depending on a number of factors, among which the choice of genre is an important issue. As I hope to have shown in the previous sections, not every element is suitable in the paragraph-initial position and, as Hoey (2005: 150) points out, "[w]riting effectively involves using appropriate text and paragraph beginnings". Corpus analyses like the one presented in this study are important, not only for examining specific aspects of a language, but are "central to a proper understanding of discourses as a whole" (Hoey 2005: 150).

A set of language patterns, collocations and colligations, have been explored in this paper that commonly occur in paragraph beginnings of the selected texts, both general patterns (grammatical colligations) and more specific ones (collocations such as *según* + SOURCE, or single words such as the coordinator *pero* or the adverb *también*). Some cohesive choices have also been indicated.

A number of tendencies noted for the genres examined are as follows. All genres are characterised by an important tendency to have nouns or noun phrases opening paragraphs. These nouns are primed to refer to place names in travel blogs, not surprisingly, and surnames in film reviews. With half of the frequency for nouns, adverbs are also found in paragraph beginnings in all genres with the exception of the cooking texts, where verbs are the second most found grammatical category, which can be explained by the usual instructional orientation of the cooking genre.

The news texts represent the genre with the greates range of variation of markers in paragraph beginnings. Apart from the linguistic patterns also found in the film and travel genres, the use of the coordinator *pero*, quotation sequences and lexical cohesive chains are important markers of paragraph beginnings. The cooking blog texts are the most distinct from the other three genres considered for the marking of paragraph beginnings, in that they do not share many of the markers found for the other three.

With the multi-genre perspective adopted and the application of the LP theory in combination with corpus-based methods, the study has offered a contribution to the investigation of an important aspect of text segmentation and organisation. The paper is also a modest attempt at demonstrating that genre is central in studies of language use based on corpora and disregarding a view of language description where languages are seen as a homogeneous whole (see Berber Sardinha 2017: 224).

At various points in this contribution, I have noted the modest character of the data considered. Indeed, one limitation of the present study is the restricted number of paragraphs that have been examined in my small self-compiled corpus, and especially in the genres of cooking blogs and film reviews. Hoey himself also acknowledges this limitation: "[o]ne problem, though, must be faced from the outset and that is the difficulty of accessing sufficient data. ... In a written corpus of nearly 100 million words, the number of paragraph boundaries and, of course, the number of texts are far fewer. The evidence for the claims I make in this chapter [on lexical priming] should be seen as suggestive rather than conclusive" (Hoey 2005: 131). Likewise, O'Donnell et al. (2012) note for their investigation of textual colligation, aiming at identifying lexical items with a tendency to occur at particular points in a text, that it needs "a corpus made up of a larger number of homogeneous whole texts". In this respect, the present contribution must be considered a first attempt at examining issues on paragraph boundaries across a selection of genres, and a preliminary investigation into the application of certain notions in LP that have relevance for paragraph segmentation.

One aspect of paragraph beginnings which has fallen outside the scope of this contribution is the examination of how the beginning of *initial* or *final* paragraphs of texts are different from the beginning of the *central* paragraphs. Thus, one could investigate paragraph-initial elements such as *en definitiva* or *y, sobre todo*, indicating that the text is coming to an end, as in the following examples of text final paragraphs:

(29) <p> **En definitiva**, si las administraciones han cerrado el grifo a las subvenciones para proyectos culturales, el micromecenazgo y una forma de entender la creación en red y al servicio de todos puede ser la mejor solución para que las ideas no dejen de fluir. Y en España tenemos ya espacios en internet con la calidad y seguridad suficiente para que el crowdfunding despegue con fuerza. (News, Crowdfunding)

(30) <p> **Y, sobre todo**, la lección definitiva de alguien que murió como había vivido: combatiendo, pero en paz. (News, Hitchens)

In this respect, O'Donnell et al. (2012) offer a taxonomy of their examined sentences in a newspaper corpus including headline and subheadline sentences, text-initial sentences, paragraph initial sentences and single-sentence paragraph sentences.

Acknowledgements

References

Berber Sardinha, Tony. 2017. Lexical priming and register variation. In Michael Pace-Sigge & Katie J. Patterson (Eds.), *Lexical Priming: Applications and Advances* (pp. 189-230). Amsterdam & Philadelphia: John Benjamins.

Biber, Douglas & Conrad, Susan. 2019. *Register, Genre and Style*. 2nd ed. Cambridge: Cambridge University Press.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan & Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English.* London: Longman.

Borreguero Zuloaga, Margarita. 2021. Las unidades del discurso y el texto escrito. In Óscar Loureda & Angela Schrott (Eds.), *Manual de lingüística del hablar* (pp. 373-391). Berlin: De Gruyter.

Brown, Gillian & Yule, George. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.

Bustos Gisbert, José Manuel. 2012. Interpretación multifuncional del párrafo. *Estudios de Lingüística Universidad de Alicante*, 26, 55-88.

Butler, Christopher S. 2003. *Structure and Function. A Guide to Three Major Structural-Functional Theories*. Vol. 2: *From Clause to Discourse and Beyond*. Amsterdam and Philadelphia: John Benjamins.

van Dijk, Teun. 1977. *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. London: Longman.

Dik, Simon C. 1997. *The Theory of Functional Grammar*. Part 2. *Complex and Derived Constructions*. 2nd revised ed., Kees Hengeveld (Ed.). Berlin and New York: Mouton de Gruyter.

Downing, Angela. 2015. *English Grammar: A University Course.* 3rd ed. London: Routledge.

Firth, John R. [1951]1957. A synopsis of linguistic theory, 1930-1955. In Frank Palmer (Ed.), *Selected Papers of John R. Firth 1952-59* (pp. 168-205). London: Longman.

Givón, Talmy. 1983. Topic continuity in discourse: an introduction. In Talmy Givón (Ed.) *Topic Continuity in Discourse: A Quantitative Cross-Language Study* (pp. 1-41). Amsterdam and Philadelphia: John Benjamins.

González Cobas, Jacinto. (2010). Causas del cambio de párrafo. Estudio de un texto narrativo. *Verba Hispanica*, 18, 117-133.

González Cobas, Jacinto. (2014). Aproximación a la unidad párrafo: un enfoque plural. *Estudios de Lingüística del Español*, 35, 161-188.

Goutsos, Dionysis. 1997. *Modeling Discourse Topic: Sequential Relations and Strategies in Expository Text*. Greenwood Publishing Group.

Grimes, Joseph. 1975. *The Thread of Discourse*. The Hague: Mouton.

Halliday, Michael A.K. 1966. Lexis as a linguistic level. In C.E. Bazell, J.C. Catford, Michael A.K. Halliday & Rossell H. Robins (Eds.) *In Memory of John R. Firth* (pp. 148-62). London: Longman.

Hannay, Mike & Mackenzie, J. Lachlan. 2002/2017. *Effective Writing in English: A Sourcebook*. 3rd ed. Bussum: Coutinho.

Hinds, John. 1977. Paragraph structure and pronominalization. *Papers in Linguistics*, 10. 77-99.

Hinds, John. 1978. Levels of structure within the paragraph. *Proceedings of the 4h Annual Meeting of the Berkeley Linguistics Society*, 598-609.

Hoey, Michael. 1983. *On the Surface of Discourse*. London: George Allen & Unwin.

Hoey, Michael. 2005. *Lexical priming. A New Theory of Words and Language.* Oxford: Routledge.

Hoey, Michael. 2013. Lexical Priming. In Carol A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 3342–3347). Oxford: Blackwell.

Hoey, Michael & O'Donnell, Matthew B. 2008. Death to the topic sentence: How we really paragraph. In Leung, Yiu-nam (Ed.) *Selected Papers from the Sixteenth International Symposium on English Teaching* (pp. 60-76). Taipei, Republic of China: English Teachers' Association/ROC.

Hunston, Susan & Sinclair, John. 1999. A local grammar of evaluation. In Susan Hunston & Geoff Thompson (Eds.) *Evaluation in Text. Authorial Stance and the Construction of Discourse* (pp. 74-101). Oxford: Oxford University Press.

Ji, Shaojun. 2002. Identifying episode transitions. *Journal of Pragmatics*, 34, 1257-1271.

Ji, Shaojun. 2008. What do paragraph divisions indicate in narrative texts*? Journal of Pragmatics,* 40, 1719-1730.

Koen, Frank, Becker, Alton & Young, Richard. 1969. The psychological reality of the paragraph. *Journal of Verbal Learning and Verbal Behaviour*, 8(1), 49-53.

Leech, Geoffrey. 1974. *Semantics*. Harmondsworth: Penguin.

Longacre, Robert E. 1979. The paragraph as a grammatical unit. In Talmy Givón (Ed.) *Syntax and Semantics*. Vol. 12. *Discourse and Syntax* (pp. 115-134). New York: Academic Press.

Longacre, Robert E. 1980. An apparatus for the identification of paragraph types. *Notes on Translation*, 15, 5-22. Summer Institute of Linguistics.

Martínez Caro, Elena. 1999. *Gramática del discurso. Foco y énfasis en inglés y en español*. Barcelona: PPU.

Martínez Caro, Elena. 2014. El párrafo como unidad discursiva: consideraciones de forma y contenido relativas a su demarcación y estructuración. *Estudios de lingüística del español*, 35, 189-213.

McGee, Iain. 2014. The pragmatics of paragraphing English argumentative text. *Journal of Pragmatics*, 68, 40-72.

Obdržálková, Vanda. (2013). El párrafo como unidad sintáctico-formal y temática de textos expositivo-argumentativos en checo y español: enfoque contrastivo y traductológico. *Sendebar*, 24, 127-150.

O'Donnell, Matthew Brook, Scott, Mike, Mahlberg, Michaela & Hoey, Michael. 2012. Exploring text-initial words, clusters and concgrams in a newspaper corpus. *Corpus Linguistics and Linguistic Theory*, 8(1), 73-101.

Pace-Sigge, Michael. 2013. *Lexical Priming in Spoken English Usage*. London: Palgrave Macmillan

Pace-Sigge, Michael & Patterson, Katie J. (Eds.) 2017. *Lexical Priming: Applications and Advances*. Amsterdam & Philadelphia: John Benjamins.

Robbins, Rossell H. 1980. *General Linguistics. An Introductory Survey*. 3rd ed. London: Longman.

Siepmann, Dirk, Gallagher, John D., Hannay, Mike, Mackenzie, J. Lachlan. 2008. *Writing in English: A Guide for Advanced Learners*. Tübingen: A. Francke.

Sinclair, John M. 1966. Beginning the study of lexis. In C.E. Bazell, J.C. Catford, Michael A.K. Halliday & Rossell H. Robins (Eds.) *In Memory of John R. Firth* (pp. 410-30). London: Longman.

Sinclair, John M. 1991. *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Stubbs, Michael. 1995. Corpus evidence for norms of lexical collocation. In Guy Cook & Barbara Seidlhofer (Eds.) *Principle and Practice in Applied Linguistics* (pp. 245-56). Oxford: Oxford University Press.

Stubbs, Michael. 1996. *Text and Corpus Analysis.* Oxford: Blackwell.

Vera Luján, Agustín. 2012. El párrafo como unidad discursiva. *ELUA, Estudios de Lingüística*, 26, 343-358.