

Article

Theme in English Native and Learner Writing

Ana Elina Martinez-Insua

Universidade de Vigo, Spain

Received: 30 September 2021/Accepted: 15 February 2022/Published: 1 April 2022

Abstract

This study explores Theme and its presence in English L1 and L2 writing. Adopting the Hallidayan definition as a point of departure, Theme is characterised by its position and its orienting function. Drawing on Berry's (2013) distinction between contentful and contentlight Subject Themes and Prince's (1981) scale of assumed familiarity, the study works on a scale of contentfulness to classify thematic components according to semantic weight. The research aims to test the applicability of Theme and contentfulness in the process of teaching and learning English and assumes that greater awareness of the thematic structure of native English may benefit learners and teachers alike. The learner data are drawn from the *Written Corpus of Learner English* (WriCLE), and the English L1 control sample from the *Louvain Corpus of Native English Essays* (LOCNESS). The findings reveal the presence of significant differences between Themes produced by L1 and L2 writers, as well as significant connections between the L1-user/L2-user dichotomy and certain features of the thematic components in the essays analysed. The findings support the importance of introducing not only corpus literacy, but also notions such as Theme and contentfulness into the process of teaching and learning English L2.

Keywords

Subject Theme, contentfulness, learner English

1 Introduction

Studies on the notion of Theme from the perspective of Systemic Functional Linguistics (SFL) are abundant in the literature (Berry, 1995, 2013; Forey, 2002; Fries, 1981, 1992, 1994, 1995a, 1995b; Halliday & Matthiessen, 2004, 2014; McCabe, 1999; Martin, 1992, 1995; Matthiessen, 1992, 1995; Whittaker, 1995, among others). Similarly, the literature features a rich array of more specific studies which look at Theme in relation to the process of teaching and learning languages (Alyousef, 2015; Drury, 1991; Fontaine & Kondratof, 2003; Hasselgard, 2009; Herriman & Boström Aronsson, 2009; Mauranen, 1996; Ventola & Mauranen, 1991, and others). The aim of this research is to contribute to this latter group, continuing the line of inquiry initiated in Martinez-Insua (2018) to provide a deeper understanding of Theme and its uses in native and learner writing, and to consider how a greater focus on Theme may help to improve the process of language learning and teaching in English.

The whole investigation rests on the certainty that the textual metafunction is a dynamic resource; and it is realized as a textual movement. It does not treat a text as a synoptic whole but as a constant process of unfolding ... the textual metafunction motivates ideational decisions (Matthiessen, 1992, pp. 73-74).

It seems logical, therefore, that attention should be paid to this metafunction in the process of learning and teaching a language. It is contended here that, even without a deep knowledge of the systemic functional framework, students wishing to learn how to speak and especially write a second language may benefit from a basic grounding in the concept of Theme and its features. In this respect, the research supports Huang et al.'s (2017) claim that a functional approach to teacher development can have a positive impact on both teachers' knowledge and skills and students' writing.

Thematic structure can be realised differently in different languages: not only do speakers of different languages often organise the same information in different ways (McCabe & Alonso-Belmonte, 2000, p. 78), but the logic expressed through the organisation of written text is culture-specific (Kaplan, 1995, p. 21). While the connection between thematic structure and coherence is widely recognised as a core factor of comprehension, "[t]here is no universal model of what constitutes a coherent text; each language (and each culture) has developed its own methods and writers will encounter cross-cultural interference when writing in a foreign language" (Fontaine & Kondratof, 2003, p. 2). Non-native academic writing, for example, may be correct as far as grammar and lexicon are concerned but "still appear somewhat deviant from native writing" (Herriman & Boström Aronsson, 2009, p. 101). In this respect, Mauranen (1996, pp. 226-227) notes that

texts which consist of impeccable sentences may appear unusual, even somewhat incoherent in their organization, but this does not necessarily result from faulty thinking. Language-specific discourse skills and culture-specific discourse strategies combine to yield texts which may flout expectations that readers have formed as part of their socialization into their own writing cultures.

In this respect, the present study is influenced by others which have shown that thematic structures in English and Spanish are not identical, and that these differences are at least partly the result of the different degrees of verbal inflection in each language (Arús-Hita, 2010, p. 30; Arús-Hita et al., 2012; McCabe, 1999).

Applying Biel's (2019, p. 51) advice that "[o]ne way to study the needs of language learners is to look at their performance and to thereby find ways to improve language learning and consequently language teaching", this research analyses small samples of native and learner written English (see Section 3.1) to identify where and how to focus language teaching and learning efforts.

The rest of this article is organised as follows: Section 2 outlines the theoretical framework used, focusing on the notions of Theme and contentfulness. Section 3 contains a description of the samples, analysis variables, hypotheses and methodology used. Section 4 presents the results in relation to aspects such as contentfulness, presence of textual and interpersonal elements, and Theme length (as measured in words). Section 5 discusses the findings and their implications for language teaching, and Section 6 concludes with some final remarks.

2 Theoretical Framework

2.1 Theme

The notion of Theme advanced in this study deviates slightly from the Hallidayan definition in that it is not restricted to "the first group or phrase that has some function in the experiential structure of the clause" (Halliday & Matthiessen, 2014, p. 91). Taking as its basis the idea of Theme as the point of departure of the message, the study is informed by Berry (1995, 2013), Matthiessen (1992) and

Firbas (1992a, 1992b) and argues that the Theme is clause initial but its boundary may extend beyond the first constituent of the clause. This characterisation of Theme as initial in the utterance is based on its orienting value as “the hearer’s starting point for creating the context of interpretation, and for understanding how the speaker facilitates the interaction by structuring the clause in a way that makes it easier for the hearer to infer certain aspects of the speaker’s intention” (Lapolla, 2019, p. 12). The Theme advances gradually towards the Rheme of the clause and the elements it contains “influence the hearer’s projection of what is to follow” (ibid.). While the Theme is rich in thematic prominence and poor in communicative dynamism (CD), the Rheme lacks thematic value and carries the higher degrees of CD of the core-constituting elements.¹ The transition from Theme to Rheme, from the foundation-laying elements to the core of the message, is gradual and realised by the verbal component of the message.

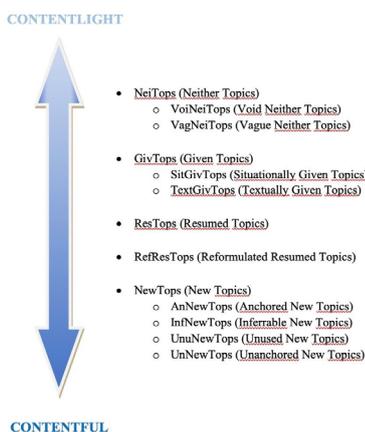
In her research on children’s writing, Berry (1995, 2013) argues that the Theme of a clause may go beyond the first constituent and cover “everything up to the main verb” (2013, p. 248). Such assumption is adopted here, just like Matthiessen’s (1992, p. 51) acknowledgement that “it is hard to impose a constituency boundary” between Theme and Rheme, there being a “diminuendo effect” whereby “the thematic prominence of the clause gradually decreases as the clause unfolds”. As explained by Matthiessen (1992, p. 52), a given configuration of constituents can serve as a carrier of a textual wave, where thematic prominence unfolds from the early thematic peak to the late thematic trough. This idea of a textual wave carried by the grammatical constituency of the message is combined in this study with Firbas’ (1992a, 1992b) view of the verbal sequence acting as a mediator between the foundation-laying elements of the clause in the Theme, and the core-constituting elements in the Rheme.

2.2 The contentfulness scale

In her analysis of the thematic options that lead to success in informal spoken and formal written English, Berry (2013) focuses on meaning and semantic choice in Subject Themes (SubjThs). More specifically, she discusses “content weight in terms of meanings, in terms of what the Subject Themes refer to” (2013, p. 249). She establishes a distinction between contentful and contentlight SubjThs, arguing that contentful SubjThs advance the subject matter of the message by introducing new material or giving indications of a new attention to old material (2013, p. 258). By contrast, contentlight SubjThs do not propose new information but rather present as given something which is already present in the discourse, easy to process or simply not relevant (2013, p. 262). According to Berry’s analysis, the lesser the meaning, the lighter or more contentlight the Theme, and vice versa: the greater the amount of semantic content, the more contentful the Theme.²

Figure 1

The Contentfulness Scale



NeiTops represent the user's choice for imprecision of reference, a kind of pass option in which the user chooses not to use the thematic position to foreground any particular meaning, either interpersonal or experiential. Two degrees of lack of precision can be distinguished: the absence of reference, on the one hand, and vague reference on the other (Berry, 2016, p. 181). Elements such as existential *theres* and dummy and anticipatory *its* make up the subcategory labelled as VoiNeiTops in Figure 1, while the subcategory VagNeiTops comprises *theys* without antecedents and generic *yous, wes* and *ones*.

1. *It*_{VoiNeiTop} seems we are being subconsciously and consciously led to believe that a woman can do anything a man can do without considering the consequences of that concept (ICLE-US-SCU-0001.4)
2. *As I see it, we*_{VagNeiTop} would need to reformulate the gun control issue so that it is less about the Second Amendment and more about the need to reduce gun violence (UAM/C33-1 C1)

The distinction between the pronominal SitGivTops (3) and TexGivTops (4) is informed by Prince's (1981, p. 236) distinction between evoked references recoverable from the text and those recoverable from the situation.

3. *I*_{SitGivTop} totally agree with their view (UAM/C33-1 C1)³
4. ... *but they*_{TexGivTop} fail to consider them as common citizens (UAM/C125-2 C2)

As in Berry (2013), the label ResTop (5) refers to nouns/nominal groups that, having been introduced into the discourse time ago, require reactivation. In the case of RefResTops (6), the noun/nominal group contains, in addition, reformulations that bring semantic nuances to the resumption of the topic.

5. *for British firms which rely on international trade to survive, the single market*_{ResTop} will be an incentive to increase dealings with other European firms (ICLE-BR-SUR-0006.3)
6. *Somehow, the prospect of treating Europe as one market*_{RefResTop} is more acceptable if the countries participating keep their individual currencies and hence their identities (ICLE-BR-SUR-0006.3)

Even though all four types of NewTops are presented as new when introduced into a message, AnNewTops and InfNewTops maintain a higher degree of connection with the preceding context and their contentfulness is, therefore, lighter. The reference of AnNewTops (7) has to be created by the hearer, but they contain an anchor that links them with some other entity in the discourse. It is this other entity in the discourse which lightens the contentfulness of the reference and makes anchored topics the least contentful subtype of NewTops.

7. ... *Whether transsexuals gain or not specific rights worldwide*_{AnNewTop} might depend on society's accepting and updating on this human group that, far from being voiceless, seems to be speaking out louder than ever (UAM/C125-2 C2)

Like in Prince's 'Inferrables' (1981, p. 236), the sender assumes the addressee can infer the discourse entity in InfNewTops (8) from others already present in the discourse through logical or plausible reasoning. The connection is logical in this case: there is no anchor, so the contentfulness of the reference increases with respect to AnNewTops:

8. *In this way, the National Rifle Association (NRA)*_{InfTop} has traditionally identified the purchase of a gun with the exercise of an inalienable right (UAM/C33-1 C1)⁴

UnuNewTops (9-10) refer to entities the sender assumes to be in the addressee's discourse model even though they are not present in the current discourse context. The addressee does not need to create them but simply place them into the particular discourse-model. The entity which is introduced into the message is treated as known by the addressee even if this is its first mention.

9. *Noam Chomsky*_{UnuNewTop} went to Penn (Prince, 1981, p. 233)
10. Getting back to the idea of hell the next question might come up when thinking about that place of punishment; who would go to hell in the Church's opinion? *Answering that question James Akin*_{UnuNewTop} says... (UAM/C125-3 C2)

UnNewTops (11), the most contentful category in the scale, includes references to discourse entities which are not present, anchored, or linked to anything in the preceding discourse, so the addressee has to create them *ex novo*.

11. *The trend of our society today*_{UnNewTop} is striving so hard to equalize men and women that the efforts perhaps have been taken a little too far when it comes to women in combat (ICLE-US-SCU-0001.4)

3 Spanish Learner and English Native Writing

A sample of 1,045 clauses drawn from academic essays by university students was used for this exploratory analysis of L1 and L2 academic writing. In the case of learners, the sample contains essays by students with two different levels of proficiency: B2 (upper-intermediate) and C2 (proficiency), according to the Common European Framework of Reference for Languages (CEFR). In the case of L1 users, the sample comprises essays written in American and British English.

3.1 Data

The data were drawn from two corpora, WriCLE (*Written Corpus of Learner English*) and LOCNESS (*Louvain Corpus of Native English Essays*). WriCLE (750,000 words) contains essays written in English by Spanish university students from Universidad Autónoma de Madrid (Rollinson & Mendikoetxea, 2010). LOCNESS (324,304 words), the control English L1 corpus, forms part of the ICLE corpus (*International Corpus of Learner English*), developed at the Université Catholique de Louvain (Granger, 1998), and contains essays written by university students using British and American English, and also by British A-level pupils.

The texts selected from WriCLE comprise: (1) five argumentative essays by B2 learners (average length: 50.4 clauses); and (2) five argumentative essays by C2 learners (average length: 53.4 clauses). In both cases, all students had Spanish as L1. The texts selected from LOCNESS include: (1) ten argumentative essays by British university students (average length: 26.5 clauses); and (2) ten argumentative essays by American university students (average length: 26 clauses). In both cases, all students had English as L1. Table 1 describes the sample in more detail.

The use of corpus evidence as part of the toolbox for teachers and learners “can illuminate language teaching from many different angles” (Sinclair, 1991, pp. xii-xiii). More specifically, “[c]ombined with native-language corpora as positive evidence of language use, learner corpora can be used to provide negative evidence” (Callies, 2019, p. 253). While the samples used in this study are not large, small comparative samples like these can offer a useful illustration of how language is deployed in the foreign language as a way to help learners to achieve particular communicative goals (Thompson, 1991, p. 311). Indeed, small corpus analyses can be a source of additional knowledge about language that “could not have been discovered through the analysis of large corpora” (Ghadessy et al., 1991, p. xvii)

Foreign language and mother tongue comparisons can be highly effective for understanding how the foreign language functions in its environment, and may “lead to insights that were not even sought for at the outset of the investigation” (Hasselgard, 2009, p. 138). This, in turn, “will help learners cope better with the task of coming to terms with the language” (Thompson, 1991, p. 316). Encouraged by such evidence, this research seeks to detect possible differences in the Themes used by natives and advanced learners in their writing; which might explain specific difficulties encountered by Spanish learners on the road to proficiency.

Table 1
Analysis Sample

Corpus	Variety	Essays	# clauses analysed	# words
LOCNESS	BrE	BR-SUR-001.3	265	5,936
		BR-SUR-002.3		
		BR-SUR-003.3		
		BR-SUR-004.3		
		BR-SUR-005.3		
		BR-SUR-006.3		
		BR-SUR-007.3		
		BR-SUR-008.3		
		BR-SUR-009.3		
		BR-SUR-0010.3		
LOCNESS	AmE	US-SCU-0001.4	261	6,285
		US-SCU-0002.4		
		US-SCU-0003.4		
		US-SCU-0004.4		
		US-SCU-0005.4		
		US-SCU-0006.4		
		US-SCU-0007.4		
		US-SCU-0008.4		
		US-SCU-0009.4		
		US-SCU-00010.4		
WriCLE	B2	UAM/A11-2	252	6,185
		UAM/A101-4		
		UAM/A101-2		
	C2	UAM/A142-1	267	6,200
		UAM/A145-2		
		UAM/C125-2		
C2	UAM/C125-3	267	6,200	
	UAM/C15-1			
	UAM/C15-2			
		UAM/C15-3		

3.2 Variables

The dependent variables used in this analysis of Theme include: contentfulness, presence of textual and interpersonal components in the Theme, and number of words per Theme. The independent variables are: the writers' L1 (English or Spanish), geographical variety used by English-L1 subjects (BrE or AmE), and level of proficiency of learners (B2 or C2).

3.3 Hypotheses

Null Hypothesis:

- ~ There are no differences in the SubjThs in essays by L1 and L2 writers

If the null hypothesis is rejected, the following additional hypotheses will be tested:

Hypothesis (I): Contentfulness of SubjThs

- ~ The contentfulness of SubjThs in essays by L1 writers differs from that in essays by L2 writers

Hypothesis (II): Presence of Textual Themes

- ~ Themes in essays by L2 writers contain Textual Themes more frequently than those in essays by L1 writers

Hypothesis (III): Presence of Interpersonal Themes

- ~ Themes in essays by L2 writers contain Interpersonal Themes less frequently than those in essays by L1 writers

Hypothesis (IV): Number of words per Theme

- ~ Themes in essays by L1 and L2 writers differ in number of words

The rationale underlying these hypotheses relates to the fact that word order is not fixed in Spanish, and explicit Subjects are not compulsory. Fewer pronominal SubjThs can be expected, partly because of the Spanish tendency to “leave pronominal Subjects unexpressed as a by-product of verbal inflection” (Arús-Hita, 2010, p. 30). In addition, no strict end-weight principle seems to apply to Spanish, which could favour the presence of longer strings in thematic position. Finally, although the use of interpersonal components in thematic position is frequent in spoken Spanish, it is hypothesised that learners use them less profusely when they write in English, being mindful of possible word-order restrictions in English and of the fact that features of involvement and interaction with the audience are more appropriate for spoken than for written language (Herriman & Boström Aronsson, 2009; McCabe, 1999).

3.4 Research methodology and notational conventions

The methodology used was direct observation. Themes were retrieved manually from the 30 essays sampled, identifying 1,045 clauses.

To facilitate the retrieval and analysis of the Themes, each clause is assumed to represent a cline from thematicity to rhematicity mediated by transitionality. As described earlier, the area of thematicity (Theme) displays a lower degree of communicative dynamism (CD) than that of rhematicity (Rheme), but also realises the thematic peak that Rhemes lack. Despite this gradual transition from one to the other, for operational reasons, three separate parts are identified in each clause: Theme, Transition and Rheme. Being transitional, the verbal component is not completely devoid of thematic traits and helps to orient the message. However, to facilitate their analysis, Themes are assumed to entail everything up to the verbal component which, despite its partly orienting role, is interpreted as the Transition (except in the cases of interrogative, imperative and passive clauses, as discussed below). The function of all of the elements of the Theme is to orient, locate and lay the foundations of the clause. Except on rare occasions, Subjects belong to this part of the clause and the contentfulness of these SubjThs is the focus of the analysis presented in Section 4.1.⁵

The verbal group (including TMEs and the notional component) is analysed as the Transition, but always bearing in mind the diminuendo effect that links Theme and Transition, the lack of a clear-cut boundary between them, and the fading-out effect from the thematic peak towards the thematic trough. In clauses with a negative polarity, negative elements such as *not* are also included in the verbal group

(e.g. they *don't like ice-cream*). As mentioned, there are a number of exceptions to this, in which some or all the verbal components, including the lexical verb, belong to the Theme:

In the case of passives, all of the components up to and including the *be*-auxiliary marking the passive voice of the verbal expression are considered to form part of the Theme, as all of them share an orienting function. The *be*-auxiliary is considered part of the Theme rather than the Transition, since otherwise (12) and (13) below would have the same Theme, and the very significant portion of the orienting information it conveys would not be included in the Theme of (8).

12. *No doubt, and considering the present situation, they should be*_{Theme} given a chance.

13. *No doubt, and considering the present situation, they*_{Theme} should behave.

In the case of imperatives (as explained by Halliday and Matthiessen, 2014, p. 103), the unmarked Theme options are clauses with *let's* (14), and clauses “with the verb in thematic position” (15):

14. *Let's*_{Theme} go home now

15. *Keep*_{Theme} quiet

In the case of interrogatives, everything up to the notional verbal component is regarded as Theme. This facilitates the distinction between imperatives and interrogatives. For instance, the orienting part of (16) below (*Do you*) is clearly pointing towards interrogative word order, in contrast to (17), where the orienting function applies to *Do* only:

16. *Do you*_{Theme} want ice-cream?

17. *Do*_{Theme} your homework

The rest of the message is the Rheme, which is no longer orienting but central as it contains the core-constituting elements. As thematicity diminishes in the Transition, CD increases and reaches its highest point in the Rheme. Exceptions to this definition include intransitive messages with no post-verbal complements. Even though the transitional nature of the components implies the presence of a certain rhematic content, strictly speaking, there is no Rheme proper in these cases:

18. *It*_{Theme} *is raining*_{Transition-Rheme}

4 Findings

The results of the variable analysis explained in Section 3.2 are presented as follows: Section 4.1 offers detailed findings regarding the contentfulness of SubjThs; Sections 4.2 and 4.3 report on the presence or absence of textual and interpersonal components, respectively; finally, Section 4.4 presents the results obtained in relation to Theme length.

It is important to preface this discussion with the preliminary finding that the null hypothesis was rejected in the light of the evidence gathered. Themes in L1 and L2 essays do differ from each other and the following sections illustrate how.

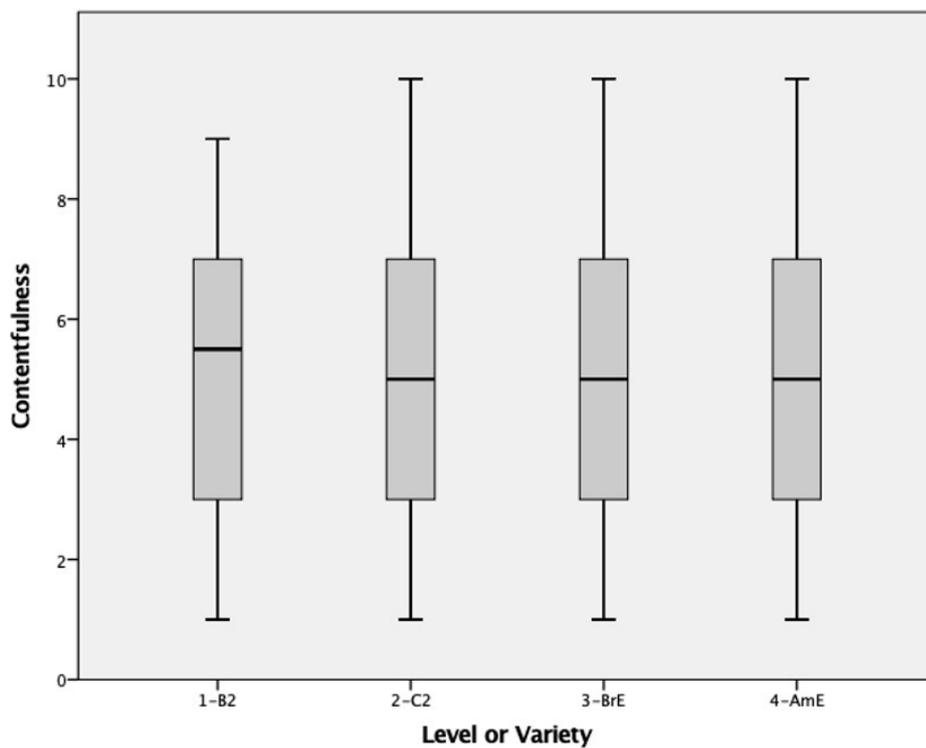
4.1 Contentfulness of subject themes

Table 2 shows the distribution of the ten categories of contentfulness in the samples analysed. In the boxplot representation of these results (Figure 2), numbers 1 to 10 on the Y axis refer to the ten categories of the contentfulness scale arranged in ascending order from most contentlight (VoiNeiTop = 1) to most contentful (UnNewTop = 10).

Table 2
Contentfulness of SubjThs

Contentfulness	B2	C2	BrE	AmE	Total
VoiNei	33 13.1%	32 12%	37 14%	19 7.3%	121 11.6%
VagNei	10 4%	13 4.9%	18 6.8%	6 2.3%	47 4.5%
SitGiv	34 13.5%	50 20.2%	39 14.7%	44 16.9%	171 16.4%
TexGiv	29 11.5%	17 6.4%	23 8.7%	45 17.2%	114 10.9%
Res	20 7.9%	37 13.9%	39 14.7%	30 11.5%	126 12.1%
RefRes	53 21%	40 15%	32 12.1%	43 16.5%	168 16.1%
AnNew	50 19.8%	38 14.2%	33 12.5%	48 18.4%	169 16.2%
InfNew	9 3.6%	19 7.1%	19 7.2%	15 5.7%	62 5.9%
UnuNew	14 5.6%	15 5.6%	16 6%	10 3.8%	55 5.3%
UnNew	0 0%	2 0.7%	9 3.4%	1 0.4%	12 1.1%
Total	252	267	265	261	1045

Figure 2
Contentfulness of SubjThs



The figures reveal that the B2 and C2 learners in the samples do not always use these types of topics equally. The distribution of NeiTops is very similar in both groups, and similar to that in L1 writing. The distribution is similar for most NewTops. However, when it comes to the distributions of given, resumed, and anchored topics, differences above 5% appear between the frequencies of use observed in intermediate and proficient writing, and also with respect to the frequencies attested in L1 writing.

When using given topics, B2 learners resemble L1 writers using AmE in that their SubjThs display even distributions of SitGiv and TexGiv topics. The SubjThs written by C2 learners, however, are closer to those observed in L1 BrE writing, where the givenness of more topics comes from the situation (SitGivTops) than from the text (TexGivTops). Regarding those topics which are recovered from the co-text, either literally (Res) or somehow reformulated (RefResTops), C2 learner writing displays a rather even distribution of both types, just like L1 BrE essays do. In the case of B2 learner writing, the distribution is noticeably uneven with a clear preference for RefResTops (21%) over ResTops (7,9%) which, again, resembles what is observed in L1 AmE writing. Finally, in the case of NewTops, the same clear preference for AnNewTops that is attested in L1 AmE is observable in B2 learner writing. When introducing new topics in SubjThs, C2 learner and L1 BrE writing also show preference for those anchored to the previous discourse, but the frequencies of inferred and unanchored topics are higher than in B2 learner and L1 AmE writing.

4.2 Presence/absence of textual themes

The aim of the second hypothesis is to test the validity of the assumption that learners use textual Themes more frequently than L1 users.

As mentioned before, long clause complexes containing more than two clauses tend to be more frequent in Spanish formal writing than in English. At the same time, the syllabi of most (if not all) ESL/EFL programmes pay significant attention to the use of conjunctions and connecting expressions from even very basic levels. Spanish learners of English devote much time to this array of connectors and are frequently encouraged to use them in order to make their output (particularly their written output) consistent and coherent. It is hardly surprising, therefore, that their essays tend to be rich in textual components.

Table 3 and Figure 3 show the distribution of Textual Themes across the samples.

Table 3

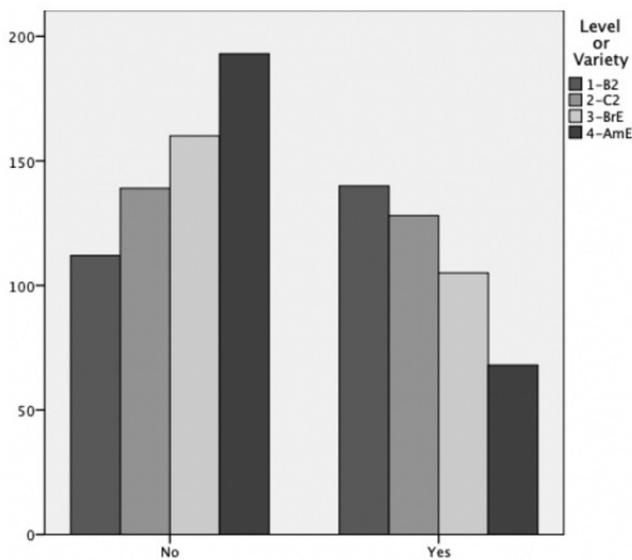
Presence or Absence of Textual Themes

Textual Theme	Level or Variety				
	B2	C2	BrE	AmE	Total
No	112	139	160	193	604
	44.5%	52%	60.4%	74%	58%
Yes	140	128	105	68	441
	55.5%	48%	39.6%	26%	42%
Total	252	267	265	261	1045

The results confirm the hypothesis, with L2 essays showing a greater presence of Textual Themes (51.63%) than L1 essays (32.88%). As expected, use of Textual Themes is higher among learners with a B2 level (55.5%) than among more proficient (C2) learners (48%), as the Themes in C2 learner essays are more similar to those produced by L1 users. While the difference between learners and L1 users of the British variety is noticeable, the difference between learners and L1 users of American English is even more significant, the latter being the group with the lowest frequency of Textual Themes, as illustrated by Figure 3:

Figure 3

Presence or Absence of Textual Themes

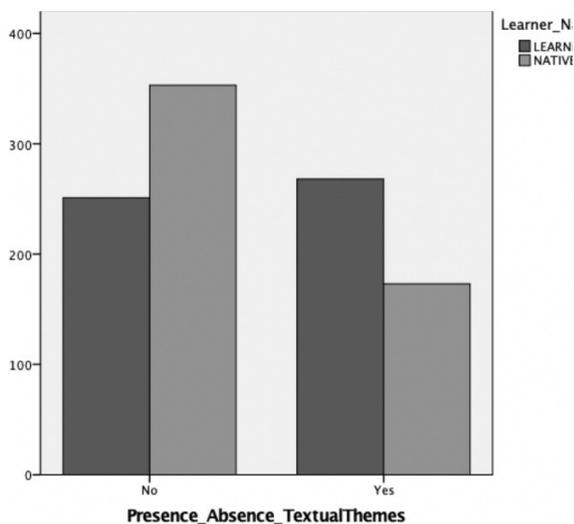


Since chi-square and Cramer’s V tests are not suitable for complex tables with multiple linguistic and explanatory variables, L1 and L2 samples were merged to create a single group of each and thereby reduce the number of variables. The merged data from Table 3 and Figure 3 are presented below as Table 3b and Figure 3b, respectively.

Table 3b

Presence or Absence of Textual Themes

Textual Theme	Level or Variety		
	L2	L1	Total
No	251 48.4%	353 67.1%	604 57.8%
Yes	268 51.6%	173 32.9%	441 42.2%
Total	519	526	1045



The difference in occurrences between the merged samples of essays by L1 and L2 writers is highly significant ($\chi^2(1)=37.645$, $p=0.000$; Cramer's V: 0.19, $p=0.000$). While more than 51% of the Themes in learner essays contain textual components, less than 33% of those in L1 essays do so. This corroborates both the statistically significant connection between the distribution of Textual Themes and the L1 or L2 profile of the writer, and the learners' greater tendency to express involvement "by interfering in the text to show how different parts of the text are connected" (Herriman & Boström Aronsson, 2009, p. 114).⁶

4.3 Presence/absence of interpersonal themes

The third hypothesis posits that Spanish learners introduce fewer interpersonal components in thematic position in their essays, with the extended assumption that presence of interpersonal Themes will be even lower in essays by learners with a lower degree of proficiency than in those of more proficient learners.

Table 4

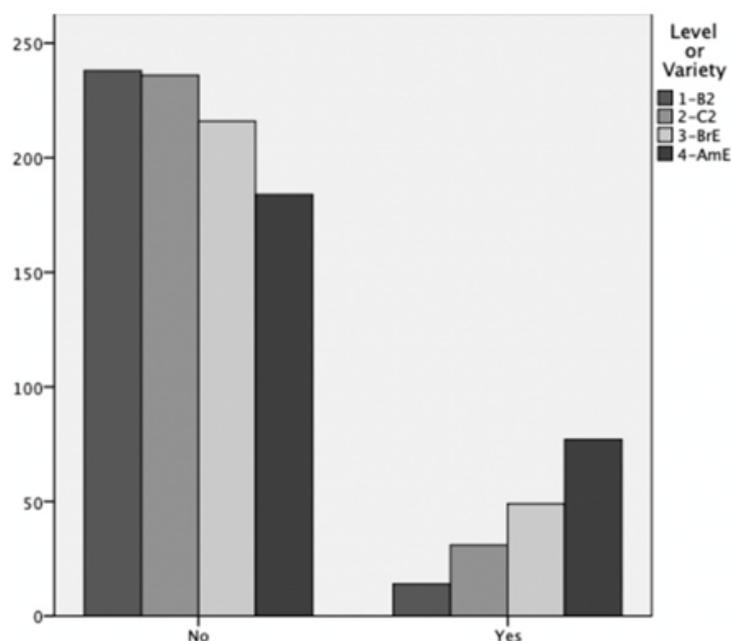
Presence or Absence of Interpersonal Themes

Interpersonal Theme	Level or Variety				Total
	B2	C2	BrE	AmE	
No	238	236	216	184	874
	94.5%	88.4%	81.5%	70.5%	83.6%
Yes	14	31	49	77	171
	5.5%	11.6%	18.5%	29.5%	16.4%
Total	252	267	265	261	1045

Table 4 shows that the frequency of interpersonal components in thematic position is remarkably lower in learner essays than in L1 essays, and even lower in B2 learner essays (5.5%) than in C2 learner essays (11.6%). Figure 4 illustrates graphically the very significant difference in distribution.

Figure 4

Presence or Absence of Interpersonal Themes



As hypothesised, learners use much fewer interpersonal components in their Themes: only 8.7% of the Themes written by learners contain interpersonal information, in contrast to 24% of the Themes written by L1 users. This scarcity is particularly noticeable in essays written by learners with lower degrees of proficiency (5.5%). Once again, a substantial difference is observed between users of the British and American varieties. While British essays showed greater frequency of textual components within the thematic area of the clause, American essays were found to be more likely to use interpersonal components in thematic position (see Table 3). To a certain extent, it comes as no surprise that Spanish learners resemble British L1 users more. Despite increased access to the American variety through globalisation, television and the internet, the British variety is still the officially taught standard in most ESL/EFL programmes in Spain.

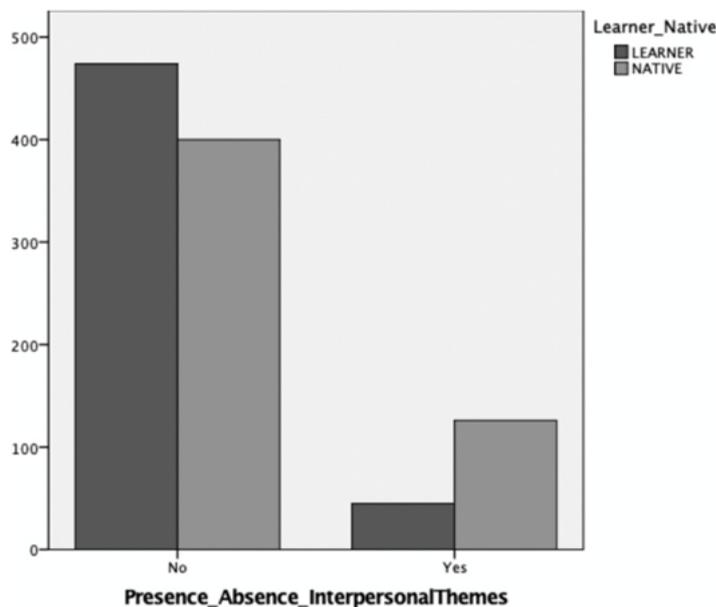
As before, in an attempt to obtain more reliable statistics in relation to interpersonal components, the data in Table 4 and Figure 4 were merged to just one linguistic and one explanatory variable.

Table 4b

Presence or Absence of Interpersonal Themes

Interpersonal Theme	Level or Variety		Total
	L2	L1	
No	474	400	874
	91.3%	76%	83.6%
Yes	45	126	171
	8.7%	24%	16.4%
Total	519	526	1045

Figure 4b

Presence or Absence of Interpersonal Themes

The statistical analysis confirms that the difference in frequency of Interpersonal Themes in essays by L1 and L2 writers is highly significant ($\chi^2(1)=44.589$, $p=0.000$; Cramer's V: 0.207, $p=0.000$; Phi: 0.207). This corroborates the statistically significant connection between the characteristics of the thematic share of the clause and the L1 or L2 profile of the writer.

4.4 Number of words per theme

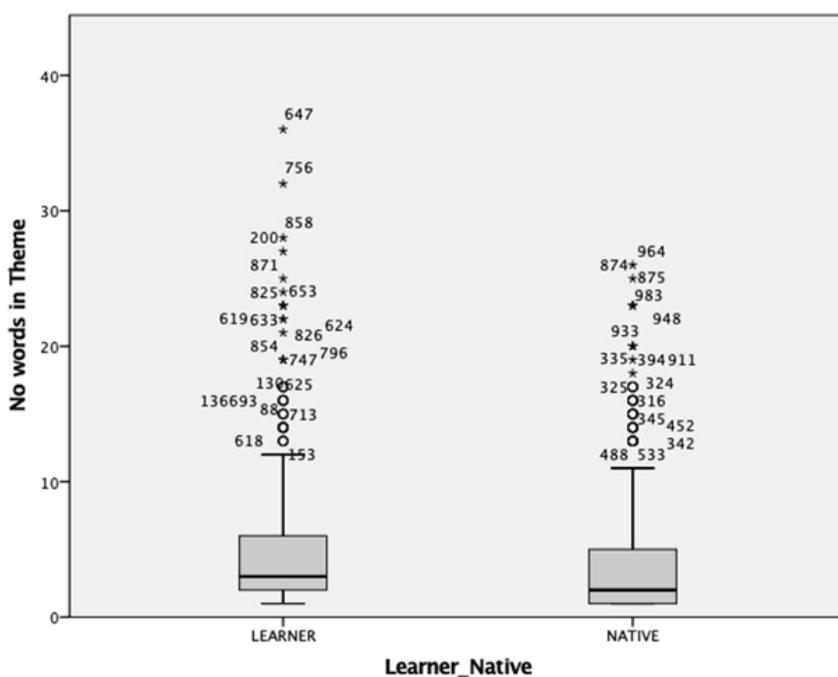
Hypothesis (IV) predicts that Spanish learners will produce longer Themes than L1 users of English. This assumption is based on the greater abundance of textual components in learner Themes and the fact that the English tendency to maintain the initial part of the clause as short as possible is not as generalised in Spanish.

Table 5

Number of Words per Theme

# words	L2	L1	Total
1 word	124 24%	173 33%	297 28.4%
2-3 words	156 30%	162 30.8%	318 30.4%
4-10 words	191 36.8%	145 27.5%	336 32.2%
+ 10 words	48 9.2%	46 8.7%	94 9%
Total	519 100%	526 100%	1045 100%

Figure 5

Number of Words in Theme

As Table 5 and the boxplot in Figure 5 illustrate, the frequency of one-word Themes is noticeably high in L1 essays (33%), and the Themes of almost 64% of the clauses in this sample contain 1-3 words. In the case of learner essays, fewer single-word Themes were attested (24%) and almost 67% of the Themes measured were 2-10 words long. The variability in number of words found in learner essays is ampler than the variability attested in L1 essays. However, overall differences for this variable were not found to be significant.⁷

5 Discussion and Implications for Language Teaching

Despite its exploratory nature, the study reveals significant differences between Spanish learner and English L1 essays, and between B2 and C2 learner essays. This concurs with reports by L2 writing instructors of student compositions which are grammatically correct but in which the overall effect is one of incoherence (Alonso-Belmonte & McCabe, 2003; Herriman & Boström Aronsson, 2009). The results seem to support the proposition that introducing the Theme/Rheme dichotomy in the language learning process may be helpful for the teacher to evaluate L2 writing at the level of discourse (Alonso-Belmonte & McCabe, 1998, p. 15).

To a certain extent, it makes sense that the features of the Themes written by more proficient learners were observed to be closer to those of the Themes written by natives. This is an encouraging finding in that it suggests that learners become aware of the importance of Themes and their features progressively as their proficiency grows, even if they are not (fully) aware of the notion of Theme. At the same time, the finding supports the desirability of addressing the notions of Theme and contentfulness before learners reach B2 level, since the sooner they become familiar with them, the sooner the improvement in their command of English stylistic features will show in their writing. Otherwise, learners' writing risks resembling native informal speech more than native formal writing, as described by Berry (2013, p. 264).

Considering the scale and the findings above, a very general summary is that the contentfulness of SubjThs in native writing differs considerably depending on the geographical variety employed, and that the contentfulness of SubjThs attested in more proficient learners' writing tends to stick to tendencies observed in L1 BrE writing. This coincidence is observed all along the scale of contentfulness, from the most contentlight topics (VoiNeiTops and VagNeiTops) to the most contentful ones (UnuNewTops), including those with an intermediate degree of contentfulness (TexGivTops). As previously pointed out, such proximity between L2 English and L1 BrE may be the logical consequence of the British variety being the one included in most official syllabi in Spain.⁸

Although this is just a tentative contribution to the subject, intended to highlight areas for further research rather than to draw any generalisable conclusions, the findings open up an interesting new didactic approach for teachers of academic writing classes. This research corroborates that the SFL framework and the use of language corpora may help in the process of language teaching and learning. The findings support the claim that raising awareness of certain concepts and providing labels for describing texts and clauses in functional terms “enables teachers to make visible and explicit to students (where relevant) how texts make meaning – both the texts that students need to read and the texts they need to write” (Coffin, 2010, p. 2). SFL and the SFL metalanguage offered here provide an array of categories for the analysis of how language constructs ideas or experiences, reflects, and enacts relationships between interlocutors, and manages the flow of information within a text and a communicative context (Gebhard & Britton, 2014, p. 107).

As Fontaine & Kondratof (2003, p. 19) argue, learners need to be taught the textlinguistic skills that can help them to manage their texts and the scientific community need an increased awareness concerning this issue. The samples analysed illustrate differences, not only between L1 and L2 users of English, but also between less proficient L2 writers and other more proficient writers. This suggests teaching materials and resources might benefit from including concepts and metalanguage such as those described here.

6 Concluding Remarks

This study examines the notions of Theme and contentfulness as possible tools for language teachers and learners. Theme is treated as the starting point of the message, the orienting portion of a wave where the

message starts moving towards the kernel of the communication, which is the Rheme. Contentfulness is construed as content weight and divided into ten categories ranging from most contentlight to most contentful. Significant differences were observed in the analysis of thematic options in L1 and L2 English writing regarding length, presence of textual and interpersonal components in thematic position, and contentfulness of the thematic share of their clauses, even in the case of learners with a reasonably advanced command of English. The results cannot be taken as definitive, not only because of the small size of the samples, but also because of the very nature of the concepts of Theme and contentfulness proposed. Further attention to them and the analysis of longer samples of L1 and L2 writing will permit a more comprehensive description of the topic.

The scope and purpose of this article is not to exhaust the possibilities of study but rather to highlight areas for future development and improvement. While it focuses on SubjThs, further research is expected to expand the categorisation to other thematic components and sharpen the definition of both Theme and contentfulness.

Acknowledgements

I am deeply grateful and indebted to Margaret Berry for her kind and unfailing support towards me and my work throughout the years. Always helpful and enthusiastic about discussion, she provided me with most illuminating comments on this study. Of course, I alone am responsible for any errors or misinterpretations.

This work was supported by the Spanish State Research Agency and the European Regional Development Fund (grant no. FFI2016-77018-P), and the Regional Government of Galicia (grant no. ED431C 2017/50).

Notes

1. Firbas (1992a, p. 168) defines CD as “the relative extent to which the element contributes towards the further development of the communication.”
2. The dichotomy between contentful and contentlight SubjThs was used in previous studies to propose what has been termed a ‘scale of contentfulness’ (Martinez-Insua, 2017, 2018), a fine-grained version of which is found in Author (2019). In its initial form, the scale of contentfulness (Martinez-Insua, 2017, 2018) was based on a reduced version of Berry’s (2013) network. The richer, fine-grained version of the scale in Martinez-Insua (2019) combines features of Berry’s (2013) contentful/contentlight dichotomy and Prince’s (1981) familiarity scale. As explained there, while Berry’s (2013) approach offers more detail in relation to contentlight SubjThs, Prince’s proposal offers richer distinctions in the contentful end of the scale (Martinez-Insua, 2019, p. 222).
3. Space limitations make it impossible to provide longer context strings. The complete texts are identified in between brackets.
4. Although ‘In this way’ forms part of the Theme, it does not belong to the Subject of the clause and so does not form part of the topic analysed in this paper.
5. This does not mean that SubjTh is the only component with contentfulness within the Theme, however. Although other components within the Theme (e.g. circumstantial Adjunct) carry a certain semantic weight that affects the Theme as a whole, their analysis is beyond the scope of the present study.
6. In their comparison of learner and native writing, Herriman & Boström Aronsson (2009, p. 114) conclude that Swedish learners’ overuse of conjunctive textual Themes makes “their style of writing

more appropriate for spoken than written language”. Similarly, in her study of native and learner summaries, Drury (1991, p. 443) observes that textual Themes “are found more commonly in the second-language summaries” and concludes that, this, “together with frequent thematic reference to the author of the source, suggests that this text is more like a spoken summary than a written summary”.

7. The statistical analysis confirms that the overall differences between the length of Themes written by L1 and L2 users are not significant ($\chi^2(3)=14.491$, $p=0.02$; Cramer’s V: 0.118, $p=0.02$; Phi: 0.118).
8. The fact that the contentfulness of SubjThs in learner writing differs occasionally depending on the writers’ degree of proficiency and that B2 learner writing seems to be closer to L1 AmE, rather than BrE, deserve further attention, which lies beyond the scope of this paper.

References

- Alonso-Belmonte, I. & McCabe, A. (1998). Theme-Rheme in L2 writing. *Didáctica*, 10, 13-31.
- Alonso-Belmonte, I. & McCabe, A. (2003). Improving text flow in ESL learner compositions. *The Internet TESL Journal*, IX (2). <http://iteslj.org/Articles/Alonso-ImprovingFlow.html>
- Alyousef, H.S. (2015). A study of Theme and information structure in postgraduate business students’ multimodal written texts: A SF-MDA of Management accounting texts. *The Asian Conference on Language Learning. Official conference proceedings. The International Academic Forum*, 302-323.
- Arús-Hita, J. (2010). On Theme in English and Spanish: A comparative study. In E. Swain (Ed.) *Thresholds and potentialities of Systemic Functional Linguistics: Multilingual, multimodal and other specialised discourses* (pp. 23-48). EUT.
- Arús-Hita, J., Lavid, J. & Moratón, L. (2012). Annotating thematic features in English and Spanish: A contrastive corpus-based study. *Linguistics and the Human Sciences*, 6, 173-192.
- Berry, M. (1995). Thematic options and success in writing. In M. Ghadessy (Ed.), *Thematic development in English texts* (pp. 55-84). Pinter.
- Berry, M. (2013). Contentful and contentlight subject themes in informal spoken English and formal written English. In G. O’Grady, T. Barlett & L. Fontaine (Eds.), *Choice in language. Applications in text analysis* (pp. 243-268). Equinox.
- Berry, M. (2016). Systemic Functional Linguistics and teachers’ knowledge about students’ writing. In M. Giovanelli, Clayton & Dan (Eds.), *Knowing about language: Linguistics and the secondary English classroom* (pp. 173-185). Routledge.
- Biel, A. (2019). Complexity and qualitative lexical knowledge: A corpus-based study on the use of TAKE in German Learner English. In S. Götz & J Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 51-73). John Benjamins.
- Callies, M. (2019). Integrating corpus literacy into language teacher education: The case of learner corpora. In S. Götz & J Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 245-263). John Benjamins.
- Coffin, C. (2010). Language support in EAL contexts. Why systemic functional linguistics? (Special Issue of NALDIC Quarterly). *NALDIC*, 2-5.
- Drury, H. (1991). The use of systemic linguistics to describe student summaries at university level. In E. Ventola (Ed.), *Functional and Systemic Linguistics. Approaches and uses* (pp. 431-456). Mouton de Gruyter.
- Firbas, J. (1992a). On some basic problems of Functional Sentence Perspective. In M. Davies & L. Ravelli (Eds.), *Advances in Systemic Linguistics* (pp. 167-188). Pinter.
- Firbas, J. (1992b). *Functional Sentence Perspective in written and spoken communication*. Cambridge

University Press.

- Fontaine, L. & Kondratof, Y. (2003). The role of thematic and concept texture in scientific text: comparing native and non-native writers of English. English version of “Comparaison du rôle de la progression thématique et de la texture conceptuelle chez des scientifiques anglophones et francophones s’exprimant en Anglais”. *Special Issue Asp, la revue du GERAS*.
- Forey, G. (2002). *Aspects of Theme and their role in workplace texts*. Unpublished PhD Dissertation. University of Glasgow.
- Fries, P. (1981). On the status of Theme in English: Arguments from discourse. *Forum Linguisticum*, 6, 1-38.
- Fries, P. (1992). The structuring of information in written English text. *Language Sciences*, 14(4), 461-488.
- Fries, P. (1994). On Theme, Rheme and discourse goals. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 229-249). Routledge.
- Fries, P. (1995a). A personal view of Theme. In M. Ghadessy (Ed.), *Thematic development in English texts* (pp. 1-19). Pinter.
- Fries, P. (1995b). Themes, methods of development, and texts. In R. Hasan & P. Fries (Eds.), *On subject and Theme* (pp. 317-359). John Benjamins.
- Gebhard, M., Chen, I. & Britton, L. (2014). ‘Miss, nominalization is a nominalization’: English language learners’ use of SFL metalanguage and their literacy practices. *Linguistics and education*, 26, 106-125.
- Ghadessy, M., Henry, A. & Roseberry, R.L. (1991). Introduction. In M. Ghadessy, A. Henry & R.L. Roseberry (Eds.), *Small corpus studies and ELT. Theory and practice* (pp. xvii-xxiii). John Benjamins.
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3-18). Addison Wesley Longman.
- Halliday, M.A.K. & Matthiessen, C. (2004). *Introduction to Functional Grammar* (3rd ed.). Arnold.
- Halliday, M.A.K. & Matthiessen, C. (2014). *Halliday’s Introduction to Functional Grammar*. Routledge.
- Hasselgard, H. (2009). Thematic choice and expressions of stance in English argumentative texts by Norwegian learners. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 121-139). John Benjamins.
- Huang, J., Berg, M., Siegrist, M. & Damsri, C. (2017). Impact of a functional linguistic approach to teacher development on content area student writing. *International Journal of Applied Linguistics*, 27(2), 331-362.
- Herriman J. & Boström Aronsson, M. (2009). Themes in Swedish advanced learners’ writing in English. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 101-120). John Benjamins.
- Kaplan, R.B. (1995). Contrastive rhetoric. In T. Miller (Ed.), *Functional Approaches to Written Text: Classroom Applications. Paris: The Journal of TESOL France in association with U.S. Information Service*, 21-39.
- Lapolla, R. (2019). Arguments for seeing Theme-Rheme and Topic-Comment as separate functional structures. In J.R. Martin, G. Figuedero & Y. Doran (Eds.) *Systemic functional language description: Making meaning matter*. Routledge.
- Martin, J.R. (1992). Theme, method of development and existentiality – the price of reply. *Occasional Papers in Systemic Linguistics*, 6, 147-184.
- Martin, J.R. (1995). More than what the clause is about: English Theme. In M. Ghadessy (Ed.), *Thematic development in English texts* (pp. 223-258). Pinter.

- Martínez-Insua, A. E. (2017). On the contentfulness of Themes in English historical medical texts. In S. Neumann, R. Wegener, J. Fest, P. Niemiets & N. Hützen (Eds.), *Challenging boundaries in Linguistics. Systemic Functional Perspectives* (pp. 111-132). Peter Lang.
- Martínez-Insua, A. E. (2018). On the relevance of the textual metafunction for Spanish learners/teachers of English. In A. Sellami & L. Fontaine (Eds.), *Perspectives from Systemic Functional Linguistics* (pp. 271-289). Routledge.
- Martínez-Insua, A. E. (2019). Scientific writing and the contentfulness of Subject Themes. How science was explained to (lay) audiences. *Journal of Pragmatics*, 139, 216-230.
- Matthiessen, C. (1992). Interpreting the textual metafunction. In M. Davies, & L. Ravelli (Eds.), *Advances in Systemic Linguistics* (pp. 20-36). Pinter.
- Matthiessen, C. (1995). Theme as an enabling resource in ideational 'knowledge' construction. In M. Ghadessy (Ed.), *Thematic development in English texts* (pp. 20-54). Pinter.
- Mauranen, A. (1996). Discourse competence: Evidence from thematic development in native and non-native texts. In E. Ventola & A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp. 195-231). John Benjamins.
- McCabe, A. (1999). Theme and thematic patterns in Spanish and English history books. PhD Dissertation. Aston University.
- McCabe, A. & Alonso-Belmonte, I. (2000). Theme, transitivity and cognitive representation in Spanish and English written text. *Revista canaria de estudios ingleses*, 40, 77-94.
- Prince, E.F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223-256). Academic Press.
- Rollison, P. & Mendikoetxea, A. (2010). Learner corpora and second language acquisition: Introducing WriCLE. In J. L. Bueno-Alonso, D. González-Álvarez, U. Kirsten-Torrado, A.E. Martínez-Insua, J. Pérez-Guerra, E. Rama-Martínez & R. Rodríguez-Vázquez (eds.), *Analizar datos>Describir variación/Analysing data>Describing variation* (pp. 1-12). Universidade de Vigo.
- Sinclair, J. (1991). Preface. In M. Ghadessy, A. Henry & R.L. Roseberry (Eds.), *Small corpus studies and ELT. Theory and practice* (pp. vii-xvi). John Benjamins.
- Thompson, G. (1991). Corpus, comparison, culture. Doing the same things differently in different cultures. In M. Ghadessy, A. Henry & R.L. Roseberry (Eds.), *Small corpus studies and ELT. Theory and practice* (pp. 311-334). John Benjamins.
- Ventola, E. & Mauranen, A. (1991). Non-native writing and native revising of scientific articles. In E. Ventola (Ed.), *Functional and Systemic Linguistics. Approaches and uses* (pp. 457-492). Mouton de Gruyter.
- Whittaker, R. (1995). Theme, processes and the realization of meanings in academic articles. In M. Ghadessy (Ed.), *Thematic development in English texts* (pp. 105-128). Pinter.

Ana Elina Martínez-Insua holds a PhD in English Linguistics from the University of Santiago de Compostela (2003) and is currently working as tenured Senior Lecturer at the University of Vigo. Her current research interest is the study of (learner) language and the process of language teaching/learning from a Systemic Functional perspective. She has published in national and international journals (*International Journal of Applied Linguistics*, *English Language and Linguistics*, *Journal of Pragmatics*, *English Studies*, *Atlantis*, *Miscelanea*) and in collective volumes in John Benjamins, Mouton de Gruyter, Peter Lang, Cambridge Scholars, Cambridge University Press, and Novus Press.