
FORMULATING IDENTITY IN ACADEMIC WRITING ACROSS CULTURES: N-GRAMS IN INTRODUCTION SECTIONS¹

Oana Maria Carciu
(University of Zaragoza, Spain)

Abstract

Language and identity are claimed to be constitutive of knowledge production and thus become integral to writing. In this article, I start from the premise that a relationship between knowledge construction and academic text production is negotiated in intercultural contexts (cf. Connor 2008) under the constraints of social, cultural and personal factors (cf. Canagarajah 2002). As reported by scholarly work (Swales 2004; Bhatia 2004) this relationship is reflected at a textual level in a range of linguistic and discursive conventions that are commonly agreed for successful English-medium scholarly exchange. Further, this relationship is to be understood as constituting a frame of reference for the inquiry into language and identity-related questions. In this study I take the case of the biomedical discourse community to examine and illustrate the two different textual responses of L1 English and L2 English (Spanish) scholars publishing research in international English-medium journals. Essentially, I analyse the linguistic expression of disciplinary identity when culture (i.e. a different linguistic background) is factored into scientific discourse. Corpus results suggest that the research article is a negotiated intercultural space which promotes a shared disciplinary identity across cultures to provide a temporarily stable ground for further social action. However, results also indicate that the linguistic expression of identity throughout the different rhetorical sections of a research article does not completely erase cultural identities and that it is difficult to disengage references to these in English-medium academic writing practices.

1. Introduction

It is broadly claimed that the research article genre acts as an instrument of discourse communities in knowledge construction and dissemination practices (Atkinson 1999; Bazerman 1988; Berkenkotter & Huckin 1995; Myers 1990; Knorr-Cetina 1981; Swales 1990, 2004). Furthermore, it has been consistently argued that research article writing becomes one of the central elements of knowledge production practices in the context of an “increasing sensitivity to the connection between writing and knowledge these days” (Canagarajah 2002: 45). Currently a much discussed academic genre across a wide range of disciplines and cultures (Fløttum *et al.* 2006;

¹ This research is funded by a FPU (Formación de Profesorado Universitario) grant from the Spanish Ministry of Education: AP2008-04473 and is a contribution to the project FFI2009-09792 funded by the Spanish Ministry of Science and Innovation.

Hyland 2000), it is mainly regarded as a written specialized text that aims to inform an academic audience of peers of the research carried out in the discipline in a way that requires careful discourse planning in agreement with a community's broadly agreed set of common public goals (Swales 1990, 2004).

The focus on the relationship between research article writing and knowledge has revealed that discursive practices of the different academic tribes are contextualized and disciplinary territory distinctive, value-charged and directed at securing prestige and promotion as a result of successfully establishing claims within a community of peers (Becher & Trowler 2001). According to Canagarajah (2002: 66), “[i]f the new findings produced by the research are widely accepted, the scholar will earn more grants for research, hold patents on the products manufactured by using the findings, and finally gain more status in the academic world (as reflected in increased salary and promotions).” This claim endorses the view that research article writing does not simply describe and explain aspects of reality in a transparent objective fashion, valuing clarity, concision and rational argument above style or interpersonal factors to communicate research findings in a discipline-specific, empirically-oriented, standard textual format, i.e. Introduction-Method-Results-Discussion/Conclusions (IMRaD: see Swales 1990, 2004); the research article genre is also a rhetorical endeavour, with a strong audience- or reader-oriented dimension (cf. Bondi & Hyland 2006), most obvious in the Introduction section (Samraj 2002; Burgess 2002). Therefore, a range of rhetorical devices would be needed to support one's hypothesis against concurrent arguments and to convince an audience of peers, as illustrated by the 'create-a-research space' (CARS) rhetorical model proposed by Swales (1990: 142):

the need to re-establish in the eyes of the discourse community the significance of the research field itself; the need to 'situate' the actual research in terms of that significance; and the need to show how this niche in the wider ecosystem will be occupied and defended. It follows that the amount of rhetorical work needed to create such a space depends on existing ecological competition, on the size and importance of the niche to be established, and on various other factors such as the writer's reputation (1990: 142).

Introduction sections often set forth a statement of purpose to 'create a research space'. Therefore, this part-genre has a conventional structure of organization that can readily be compared. However it is also widely acknowledged as the part of the research article writing which often gives writers difficulty and which can be used to determine whether an article will be accepted or rejected for publication (*ibid.*). Swales associates this difficulty with the fact that “[t]he opening paragraph requires the writer to make some decisions about the amount and type of background knowledge to be included, and authoritative versus a sincere stance [...] the appropriateness of the appeal to the readership, and the directness of the approach” (*ibid.*: 137). The Introduction section thus becomes one of the most demanding ones in terms of rhetorical effort.

Furthermore, the Introduction section is a rhetorical enterprise with important implications for users of English as a foreign language. Firstly, because non-English background scholars are compelled to publish in international English-medium journals to disseminate their work and validate it in the academic tribes which matter in their profession. In this respect, Moreno (2010) indicates that Spanish speakers are

a neglected population of users of English as regards variation in non-native academic English although, as shown by a large-scale survey of Spanish scholars' training needs for research and publication purposes, there seems to be a high level of interest in publishing research articles in English in the field of medical sciences (Moreno 2012). In another recent ethnographic study from the Spanish academic context, Pérez-Llantada *et al.*'s (2011) informants point to the Introduction section as a problematic area, while Burgess (2002) has previously reported differences in the Introduction section written by Spanish scholars as compared to their native English counterparts based on a rhetorical move-analysis of linguistics research articles.

Secondly, from a different perspective, it has been argued that attitudes of academic publishing on the rhetorical processes of research article writing has implications for the identity of local scholars (see also Canagarajah 2002: 55; Bazerman 2001). On the one hand, English-only international publications seem to impose normative academic writing conventions at the expense of scholars' culture-specific rhetorical conventions (Curry & Lillis 2004). On the other hand, according to Flowerdew (2008), contrastive rhetorical analysis of academic discourse shows that there are discrepancies in the way information is organized in different languages and cultures. In other words, "vast complexities of the cultural, social, situational and contextual factors affecting a writing situation" (Connor 2008: 304) also influence scholars' self-representation. As reported by contrastive rhetoric research, identity is visible and reflects institutional policies and attitudes of academic publishing, but takes a different shape across diverse linguistic/cultural backgrounds (see Bondi 2004, 2009; Duszak 1997; Pérez-Llantada 2010, 2012; Yakhontova 2006). Since it is posited that writing is rooted in specific cultural traditions and ways of constructing knowledge (Bazerman 1988) and raises questions of identity (Berkenkotter & Huckin 1995; Ivanič 1998; Gee 1999), both culture and identity could thus be an important determinant of academic authors' rhetorical behaviour (Mauranen *et al.* 2010).

In line with this research, differences in the rhetorical behaviour of the authors across cultures and languages could be seen to produce different identity cues in writing. Features such as uncertainty avoidance, power distance, and individualism vs. collectivism may be reflected in the authors' use or avoidance of first person plural references. The textual effects thus created might range from the expression of their certainty and commitment to their claims to the appeal to group solidarity and communal opinion with a view to soliciting readers' acceptance of claims and community consensus. In addition, identity has to do with forms of engagement and disengagement with the beliefs and values of the disciplinary discourse community, and it appears to cohere with group identities in academic writing (Becher & Trowler 2001).

Research has paid attention to identity either as a co-textual mode (i.e. personal pronouns as words in texts, sections, moves, etc.) or as a contextual one (i.e. as indicators of voice, stance, evaluation, and metadiscourse). To approach its study and show its disciplinary or cultural specificities or variations, a combination of frameworks and methodologies has been used, giving a kind of multidimensional perspective on identity (cf. Bondi 2009; Bondi & Hyland 2006 with special attention to phraseology, collocation and semantic preference; also, on cross-cultural variations Duszak 1997; Fløttum *et al.* 2006 among others). However, one of the frameworks not yet applied to

the study of identity from a cross-cultural standpoint in the Spanish-English biomedical research article Introduction section is that of phraseology (although see for instance Gledhill (2000) and Luzón Marco (2000) on grammatical collocation in medical research articles in English). Furthermore, according to Hyland (2008: 41), phraseological units constitute “an important component of fluent linguistic production”, accounting for “the needs of the interactants, irrespective of different communicative situations, in which sender and receiver or addresser and addressee have different levels of knowledge concerning the specialized domain of communication” (Schulze & Römer 2008: 269).

This paper examines the phraseological patterns of first-person plural references as markers of identity in writing, i.e. word combinations such as *we show that*, which are governed by publishing and specific textual conventions related to matters of structure and language use in academic texts, i.e. CARS rhetorical organization of the research article Introduction section and English-medium international journals’ writing conventions. The objective of this paper is to show the form, occurrence and distribution of ‘we’ n-grams in comparable sets of biomedical research article Introduction sections written by English and Spanish scholars publishing in English as a *lingua franca*, and thus identify cultural and linguistic differences in a rhetorically demanding standard communicative section.

1.1. *Theoretical framework*

As Römer (2009) suggests, the word is not the most useful unit of analysis in the search for discourse roles. As such this analysis explores word combinations instead of single words. Phraseological items are defined according to Schulze & Römer (2008: 265) as “strings of words that are highly structured, well-organized and firmly entrenched in the human being’s mind”. It could be hypothesized that an analysis of the rhetorical behaviour of the discourse community in a genre could rest on the phraseological expressions associated with communicative purposes and discourse roles specific to that genre. In this paper, ‘we’ n-grams, that is, continuous sequences of word forms that cluster around first person plural references in the biomedical research article Introductions section, are explored as indicators of writer identity, i.e. discourse roles which reflect the specific communicative purpose of the writer in a certain part of the introduction.

The theoretical framework within which ‘we’ n-grams developed in the Introduction section of biomedical research articles are explored can be summarized as follows:

- Introductions to academic articles often conform to Swales’ Create-A-Research-Space (CARS) model (1990) and therefore show the importance of the general topic and particular issue concerned, outline the gap to be filled, and make the article purpose explicit.
- The rhetorical behaviour of authors has been mapped along three lines: (1) the national culture or language background line encompassing traits such as (dis)engagement, collectivity/individuality, collaborative, in compliance with social, cultural and personal factors which prompt the redefinition of their conventional meaning and hint at “our expectation that members of different

cultures have learned different ways of expressing themselves generally and that these affect academic writing”; (2) the national science or local disciplinary culture or community line indicative of a more competitive research environment or a less competitive environment associated with the dichotomy local or national vs. international audience which may trigger the use of different rhetoric and discipline (Shaw 2003: 344-345); and (3) the personal dimension associated with responsibility but also recognition, promotion and governmental funding which is assigned to individuals who publish research in scholarly journals (see *The Uniform Requirements of the International Committee of Medical Journal Editors* 2009, http://www.icmje.org/urm_main.html).

- Personal pronouns endorse discourse roles as “signs waiting to be filled in the instance of discourse, since the deictics do not refer to any objective reality but must constantly refer to the instance of discourse that contains them.” Therefore, a set of discourse-defined semantic roles or functions which reflect the specific communicative purpose of the writer (see also Tang & John 1999; Kuo 1999) in a certain part of the introduction can be associated with personal pronouns, such as the following:

- (a) the author as describer of the research
- (b) the author as experiment conductor
- (c) the author as opinion holder
- (d) the author as cautious claim maker
- (e) the author as fully-committed claim maker.

2. Corpus and methodology

2.1. Data

For the purposes of the present study, the Biomedical and Health Sciences (BHS) component of the *Spanish-English Research Article Corpus* (SERAC²) (see Pérez-Llantada 2008) was selected. The BHS-SERAC corpus is an electronic collection of English and Spanish research articles published in English-medium and Spanish-medium journals. It is designed as a specialized corpus consisting of 270 research articles, 90 written by native English scholars (English L1), 90 by Spanish scholars in English (English L2) and another set of 90 articles in Spanish (Spanish L1). The text files were sorted into two folders: ENG texts written by native English scholars, SPENG texts written in English by Spanish scholars and published in the same international journals as the ones in the first corpus, and SP texts written by Spanish scholars and published in national journals in Spanish. Each text file has been labelled with the name of the folder and a number, e.g. ENG 1, ENG 2, etc. The third subcorpus (SP) was deemed necessary to validate the findings of the ENG and SPENG

² The three sets of texts belong to the biological & health sciences component of SERAC (The Spanish-English Research Article Corpus) compiled by the InterLAE research group at the University of Zaragoza (<http://www.interlae.com/>). I am most grateful to the research group for providing this comprehensive source for my linguistic analysis of n-grams.

comparison, following the *tertium comparationis* principle proposed by Connor & Moreno (2005). Samples of biomedical research article Introduction sections were selected, since journals in the field of medicine generally adhere to standardized conventions, i.e. the *Uniform Requirements for Manuscripts* (<http://www.icmje.org/>), as required by journals (see also Skelton 1994). This shared policy would thus secure the comparability between ENG and SPENG (see also Table 1 below for other features of the groups of texts, i.e. number of texts in the three subcorpora of Introduction sections, source journals³ & year, type of text).

Corpus	ENG	SPENG	SP
Language	English L1	English L2 by Spanish L1	Spanish L1
Journals	British Journal of Haematology, Blood, Experimental Hematology, European Journal of Cancer, Journal of Clinical Oncology, British Journal of Cancer, European Urology, BJU International, Urology	British Journal of Haematology, Blood, Experimental Hematology, European Journal of Cancer, Journal of Clinical Oncology, British Journal of Cancer, European Urology, BJU International, Urology	Anales de Medicina Interna, Medicina Clínica, Medicina Intensiva, Oncología, Cirugía Española, Archivos Españoles de Urología, Actas Urológicas Españolas
Years	2000 – 2010	2000 – 2010	2000 – 2010
Type of texts	RA Introduction section	RA Introduction section	RA Introduction section
No. of texts	90	90	90
Word count	37,987 words	38,777 words	34,197 words

Table 1. BHS-SERAC Corpus of Biomedical Research Articles Introduction section

All texts in the corpus are kept in plain text file format, and so far no metalinguistic annotational material (like word class labels or paragraph markers) has been added. This version consists of body text sections only so that targeted searches can be performed; therefore, article headers (including date of publication, email address and affiliation of the authors), footnotes, and references were cut, and any kind of information about the author (name, affiliation, research interest) was deleted as this information was not deemed necessary in the rhetorical analysis of the texts (see Cortes 2004).

2.2. Analytical steps

The main strategy for the computerized extraction of recurrent word-combinations is the analysis of n-grams (i.e. the analysis of 2-, 3-, 4-, ..., n-word sequences occurring in the corpus (cf. Biber *et al.* 1999). In this paper, the phraseological corpus engine *kFNgram* (Fletcher 2007) was used to explore the phraseological profile of the two sets of scholars' first person plural references. This program extracts contiguous lists of n-grams of different lengths (i.e. combinations of n words) from a corpus and their frequencies. Of note, although the four-word scope is "the most researched length for

³ Ten articles were extracted out of each journal, except for the SP corpus.

writing studies” (Chen & Baker 2010: 32), Simpson-Vlach & Ellis (2010: 509) have suggested in a recent study that three-word bundles constitute many important recurrent word combinations. Therefore, for the present study lists of 3-grams (such as *we show that*) have been extracted. Since the two corpora are of a relatively small size and approximately equal, the cut-off frequency was established at a raw frequency of 3, occurring in at least three texts⁴. Table 2 and Table 3 show the threshold adopted as sufficiently representative of the corpora being examined, i.e. the raw cut-off frequency and the corresponding normalized frequency are set to 3 and 80, respectively.

Corpus	Set raw frequency threshold	Corresponding normalized frequency (per million words)
ENG	3	79
SPENG	3	77.4
SP	3	87.8

Table 2. Raw and corresponding normalized frequency thresholds adopted

Corpus	Set raw frequency threshold	Corresponding normalized frequency (per million words)
ENG	80	3
SPENG	80	3.1
SP	80	2.7

Table 3. Normalized and corresponding raw frequency thresholds for comparison

Frequencies of both tokens (in parenthesis) and types are provided. Overlapping word sequences were checked manually through concordance to avoid inflating quantitative results (cf. Chen & Baker 2010, Ädel & Erman 2012). Following Ädel & Erman (2012: 84), overlapping grams such as *here we report* and *we report the* merged into *here + we report the*, and the word is put in parenthesis when the sequence does not occur with it in all instances. In addition to producing lists of repeated word combinations, *kfNgram* also identifies patterns in the extracted n-gram lists and groups n-grams that differ by only one word in the same position together, e.g. *we investigated the*, *we report the*. Such groups of n-grams are called phrase-frames (henceforth “p-frames”) and contain a wildcard character (*) that replaces any one word. The p-frame *we * the* thus summarizes the 3-grams *we investigated the*, *we report the*. *kfNgram* lists how many variants are found for each of the p-frames. A p-frame analysis hence provides insights into pattern variability to inquire to what extent John Sinclair’s Idiom Principle (Sinclair 1991) is at work, i.e. how fixed language units are or how much they allow for variation across culture and/or languages.

For the analysis a contrastive L1/L2 English comparable corpus of research articles will be used following Connor & Moreno’s (2005) comparability criteria for corpus

⁴ Following Chen & Baker (2010), in this study all the frequencies of 3-*we*-grams are raw frequencies.

studies. To illustrate differences across cultures, a reference corpus of research articles written in Spanish is used.

3. Data analysis

This section reports on the refined lists of 3-*we*-grams in ENG, SPENG and SP groups of texts. In addition, the realizations of these sequences are compared across cultures and languages and the structural variation in the corpus is presented, followed by the functional classification of *we* 3-grams found in this study, i.e. discourse roles.

3.1. *We* 3-grams in Introduction sections

This part of the analysis consisted of finding both common and different 3-*we*-grams together with their frequency of occurrence. Attention has recently been drawn to the need to distinguish between type and token frequencies in comparisons across corpora, since there might be differences and/or similarities which account either for a “narrow range of bundles but have very high frequencies of them, while another [corpus] might have the opposite pattern” (Chen & Baker 2010: 33). Therefore, the first step in the analysis was to check for 3-grams either overlapping or embedded in larger sequences (see also Biber *et al.* 1999). As shown in Table 4, the refinement prevented the retrieval of inflated quantitative data that resulted in the exclusion of six types from both ENG and SPENG, accounting for 24 tokens from the first and 26 tokens from the latter. As can be seen (Table 4), ENG seems to be similar to SPENG with regard to both occurrence (tokens) and types. These frequencies are in sharp contrast with the number of the tokens (3) and types (1) found in the Spanish L1 corpus for the Spanish *we*-sequence counterparts, i.e. queries of 3-grams containing *nosotros* and/or **mos*.

Corpus	Before refinement		After refinement	
	No. of 3- <i>we</i> -grams (types)	No. of 3- <i>we</i> -grams (tokens)	No. of 3- <i>we</i> -grams (types)	No. of 3- <i>we</i> -grams (tokens)
ENG	22	93	16	69
SPENG	18	86	12	60
SP	1	3	1	3

Table 4. Number of bundles before and after the exclusion of overlapping and embedded 3-grams *we*-sequences

Although, as Ädel & Erman (2010: 85) have pointed out, in the lexical bundles tradition only simple descriptive statistics have been used, more recent studies (Simpson-Vlach & Ellis 2010; Chen & Baker 2010) have used statistics in their studies. Therefore, the frequency difference across ENG and SPENG was tested for statistical significance, using the log-likelihood statistic⁵. No statistical significance was found

⁵To check statistical significance, Paul Rayson’s online calculator was used (<http://ucrel.lancs.ac.uk/llwizard.html>), accessed August 2012.

when comparing type frequency (LL= 1.33). However, in the case of tokens the log-likelihood value points to a significant overuse at the level of $p < 0.05$ in ENG as compared with SPENG (LL= 5.21, $p < 0.05$).

3.2. *Non-shared and shared we 3-grams*

Table 5 provides a list of *we* 3-grams divided into those that are found in L1 and L2 English writers groups. Table 6 shows only those grams which are shared across subcorpora, indicating in brackets first the frequency for the L1 English group, and second the L2 English group. For the most part, as Ädel & Erman (2012: 85) argue, it should also be taken into account that criteria, i.e. frequency and dispersion, influence what is included in the list, but it does not necessarily mean that if an n-gram is present only in the list from the L1 English writers it is not used by their L2 counterparts and vice versa.

Corpus	ENG	SPENG
We 3-grams	. we also (9)	#. we (10)
	. we previously (4)	and we have (3)
	. we therefore (3)	in addition we (4)
	et al ##### we (5)	in + this paper we (3)
	(to test) + this hypothesis we (3)	in + this report we (3)
	we have used (3)	we have recently (3)
	we hypothesized that (7)	we investigated the (4)
	we recently reported (3)	
	(here) + we report the (5)	
	we set out (+to) (3)	
	we sought to (4)	
	we studied the (3)	
Type	11	7
Token	52	30

Table 5. Distribution of *we* 3-grams in ENG and SPENG, including raw frequency in brackets

Corpus	ENG & SPENG	SP
We 3-grams	. here we (4; 4)	en este artículo presentamos (3)
	. we have (4; 7)	
	in the + present study we (4; 8)	
	in + this study we (3; 8)	
	we show that (6; 3)	
Type	5	1
Token	21; 30	3

Table 6. Shared *we* 3-grams in ENG and SPENG

As can be seen in Table 5, overall there are 52 *we* 3-grams in ENG and 30 in SPENG, the second group representing around 57% of the tokens in the first subcorpus. Thus, it can be said that L1 English scholars generally use more *we* 3-gram sequences than their Spanish counterparts in their writing. With regard to the n-grams absent from the non-native list, the sequence containing both the research noun, i.e. *hypothesis* and the research verb, i.e. *hypothesize* and the grams *we sought to*, *we set out to* offer a complex picture, in that native English scholars seem to show more variation as far as the possibilities for structuring *we* sequences are concerned. Furthermore, another difference is established in that non-native English scholars seem to draw on the research noun *report* in the sequence *in this report* as an alternative to the noun commonly used in this structure, i.e. *study*, whereas in the native list it appears in the active voice construction *we report the*. In contrast, the use of the sequence *and we have* might point to a cultural pattern specific to the Spanish language which favors the use of long sentences, a claim supported by the use of the coordinating conjunction *and* in this gram. Perhaps it is also important to consider that the *we* 3-gram which has the highest number of occurrences in the native list, i.e. *we also* (9), is a sequence which joins the human agent *we* and an adverb related to signalling the connection between specific information and the authors' point to ensure coherence in the development of arguments. Since academic discourse has the reputation of foregrounding the information being conveyed rather than the human agent, it is a pattern which should not be ignored. If this is weighted against the most frequent sequence in the non-native list which emphasizes the use of sentence-initial *we* after citation markers, i.e. *#. we* (10), it could be inferred that these sequences would help to build a promotional space in the Introduction section (cf. Bhatia 1993, 2004).

With reference to shared bundles, around 20% of the bundles are shared by the two groups. However, there are differences in the frequencies of the majority of these shared grams across subcorpora (as shown in Table 6). As such, non-natives use almost twice as often the sequences *.we have*, *present study we*, *this study we* as compared to the other group, whereas native English scholars use *we show the* more than their Spanish counterparts. With regard to the sequence *. we have*, its use may be considered as typical of academic discourse in that it implies the continuing validity of earlier findings or practices (Biber *et al.* 1999: 465). It is also important to note that the presence of the *we* 3-gram *en + este artículo presentamos* [in+ this article/paper/report we present] in the SP corpus points not to a cultural pattern, but to a genre-related use of these shared grams in the Introduction section, i.e. *here we*, *present study we*, *this study we*, *we show the*, as they have been shown to be important devices in the CARS model inasmuch as they introduce the purpose of the article (Swales 1990, 2004). Furthermore, it is unclear whether the fact that non-natives overuse this type of metadiscourse (such as *in+this study we*), but also that they use different wordings from the native writers (e.g. *in+this paper/report/study we*) is related to a cultural pattern.

3.3. *P-frames in Introduction sections*

Sinclair (1991) argued that multi-word sequences can be discontinuous, with high frequency function words as fixed elements co-occurring with variable lexical slots.

Here I turn from n-grams to p-frames to inquire into pattern variability: the patterns that occur most commonly in the corpus are identified (with the floor set at three, meaning that only items that occur three or more times are included). The different ways in which those patterns are variable or fixed are also determined, contrasting the overall patterns of use in ENG and SPENG. No instance of patterns for *we* 3-grams has been recorded in the SP subcorpus.

Three type patterns can be distinguished (see Table 7), namely:

- a) the pattern in which the first and second positions are fixed, while the third slot is variable (12*), e.g. *this study we/was, .we also/have/report/recently/therefore, we report the/on*; as can be seen from these examples, within this type pattern it can be observed that there are cases in which *we* occupies the variable slot (illustrate impersonal vs. personal constructions choices) and cases in which *we* is fixed with other items in the variable slot (personal constructions);
- b) the pattern in which the first and the third positions are fixed, while the second is variable (1*3), e.g. *we show/hypothesized that, this study/report/paper* (personal constructions);
- c) and the pattern in which the first slot is variable, with fixed second and third positions (*23), e.g. *present/this study we, ./study/and we have* (personal constructions).

Pattern Type	Categories	ENG	Variable slot (*)	SPENG	Variable slot (*)
(1) Pattern type 12*	(1a) we in variable slot (3)	<i>al ##### *</i>	and (17), the (13), in (7), although (7), we (6), a (4), which (3), these (3), it (3), but (3), as (3), however (3)	<i>in addition *</i>	to (10), we (6), in (3), the (3)
			<i>present study * was</i> (7), we (4)		<i>present study * was</i> (9), we (8)
	(1b) we fix in 2 (ENG) or 1 (SPENG)	<i>. we *</i>	also (9), have (5), report (4), therefore (3), recently (3), previously (3), hypothesized (3)	<i>we report *</i>	the (3), on (3)
(2) Pattern type 1*3	(2a) we fix in 1	<i>we * that</i>	show (6), hypothesize (6)	<i>we * that</i>	show (4), demonstrate(d) (6)
		<i>we * the</i>	report (5), studied (3), evaluated (3)	<i>we * the</i>	report (3), investigated (4), analyzed (3)
(3) Pattern type *23	(2b) we fix in 3	<i>this*we</i>	study (3), hypothesis (3)	<i>this*we</i>	study (8), report (3), paper (3)
	(3a) we fix in 3	<i>* study we</i>	present (4), this (3)	<i>* study we</i>	present (8), this (8)
	(2b) we fix in 2	<i>* we report</i>	. (4), here (3)	<i>* we have</i>	. (8), study (5), and (3)

Figure 1. Pattern types and variable slots (*) in ENG and SPENG

Figure 1 shows the proportions for the two cases found for pattern type 12*, that is, the case in which the third variable slot is filled by a word different from the personal pronoun (impersonal construction) or in which the third variable slot is the personal pronoun *we* (personal construction). These proportions indicate that the use of personal constructions is more common in the non-native corpus (42%) than in the native corpus (13%). However, if we compare impersonal and personal constructions, it seems that the latter represent more than half of the total occurrences of the pattern 12* in SPENG, whereas the highest percentage is found in ENG. The variable slot in this

type pattern for impersonal constructions (see Table 7) is either a function word or the copular verb *be*.



Table 7. Distribution of impersonal and personal constructions, the percentage of all sequences patterned 12* in ENG vs. SPENG

Whereas the first pattern type subcategory (1a) can be considered a variable one as regards the position of *we* in the 3-gram, the other subtypes show a fixed position of *we* in all possible fix slots of the 3-gram. As far as the variable slot (*) is concerned, this is filled either by function words or lexical items, i.e. research verbs and nouns such as *study*, *report*, *hypothesize/hypothesis*, *analyse*, *investigate*, *evaluate*, *demonstrate*. Contrasting across corpora, the pattern type (1b) records a wide range of options in the variable slot in ENG, thus illustrating perhaps the most striking difference between native scholars and Spanish scholars publishing in English (Table 7).

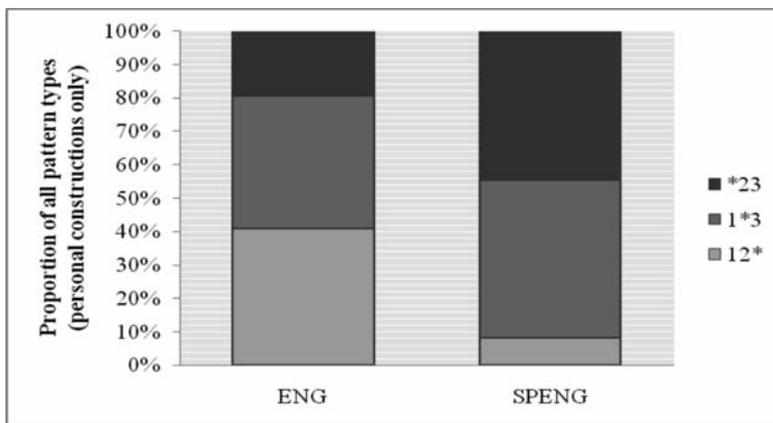


Figure 2. Proportional distribution of all pattern types (personal constructions only) in ENG and SPENG

In relation to the distribution of the three patterns for personal constructions across the native and non-native subcorpora, the following tendencies can be observed (see Figure 2). SPENG writers use the pattern type 12* less than their native counterparts. Conversely, the pattern *23 is used more in the non-native corpus than in the native one. Finally, the pattern 1*3 offers a more balanced picture of its distribution across ENG and SPENG. This is not unsurprising since the *we* 3-grams in this pattern are discourse frames (Biber & Barbieri 2007), i.e. *we (show) that, we (report) the*, thus of high importance in framing arguments and, consequently, of a more fixed or even formulaic nature (formulaic frames).

3.4. Discourse roles

As Rounds (1987: 15) suggests, personal pronouns endorse discourse roles as “signs waiting to be filled in the instance of discourse, since the deictics do not refer to any objective reality but must constantly refer to the instance of discourse that contains them”. Furthermore, based on interviews, Harwood (2007: 27) accounts for seven textual effects of the pronouns *I* and *we* which

are said to help (i) make the readership feel included and involved in the writers’ argument; (ii) make the text more accessible; (iii) convey a tentative tone and hedge writers’ claims; (iv) explicate the writers’ logic or method regarding their arguments or procedures; (v) signal writers’ intentions and arguments; (vi) indicate the contribution and newsworthiness of the research; and (vii) allow the writer to inject a personal tenor into the text.

The textual effects of the personal pronouns may be summarized as a set of discourse-defined semantic roles or functions which reflect the specific communicative purpose of the writer (see also Tang & John 1999; Kuo 1999; Martín Martín 2003) in a certain part of the introduction. The following discourse roles are considered here:

The author as describer of the research role, guiding the reader, underpins the engagement of the writer with readers. The textual effect of *we* 3-grams may consist in signalling writers’ intentions, as in the following examples:

- announcing present research (as in 1), or announcing principal findings (examples 2 and 3) in Move 3, the last rhetorical move of the Introduction section, that is, the “occupying the niche” move, using, for instance, *we* 3-grams such as *in the + present study we, in + this study we, .here we, we show that, we report the*.

(1) *In this study, we estimate the breast cancer risks for women with a strong family history of breast cancer, but tested negative for a mutation in BRCA1 or BRCA2 (ENG 60)*

(2) *Here we show that dietary PUFAs themselves are not strong stimulators of CaP invasion but require adipocyte processing. We also show that AA induces invasion itself and induces differentiation of BM mesenchymal stem cells (MSC) into adipocytes, which are themselves potent inducers of invasion (ENG 54)*

(3) *In the present study we show that this type of non-apoptotic Fas signalling during*

the process of T cell blast generation is needed for the induction of Bim expression and the sensitization of these cells to death by cytokine deprivation (SPENG 6)

- outlining purposes (Move 3); here it can be seen how writers use adverbials to develop arguments, e.g. *in addition we*, *. we also* (see examples 4 and 5):

(4) *We also assessed the effects of FLT3 inhibition on proliferative and antiapoptotic signaling to enable greater understanding of the inter-patient variations in signaling patterns that appear to influence the onset of cytotoxicity. (ENG 16)*

(5) *(...) the objective of this study was to expand our understanding of the antiaromatase properties of melatonin and to assess whether the promoters that drive aromatase expression are regulated by melatonin and to evaluate (...). In addition, we studied whether the effects of melatonin on aromatase expression are related to the effects of this hormone on intracellular cAMP concentration (SPENG 42)*

- or establishing a niche (Move 2); in the case of the sequence *. we therefore* it might be said that, in addition to its linking role, it is used to connect the writers' claim to supporting facts, and thus it could be read as an attempt to make the reader feel included and involved in the writers' claims, through the emphasis of the conclusions that the writer expects the reader to draw (as shown in 6). Perhaps it is worth pointing out that although this expression did not occur in SPENG, it does not mean that it is not used at all by the L2 scholars, but that it did not meet the criteria previously established in the retrieval of these sequences in the current study (see example 7):

(6) *A range of tools is available for the genetic manipulation of the zebrafish, as are extensive genomics resources, including a draft sequence of the entire genome. We therefore selected this species as an ideal organism for the generation of a simplified, genetically tractable model using fluorescent neutrophils to track the inflammatory response (ENG 13)*

(7) *Moreover, a longer confirmatory study for this strategy in poor prognosis DLBCL has not been reported. Therefore, we performed a prospective study with DA-EPOCH plus rituximab (DA-EPOCH-R) in newly diagnosed intermediate-high and high-risk DLBCL. (SPENG 29).*

Inasmuch as the majority of *we* sequences in the above-mentioned examples might represent fixed, formulaic discourse frames, no cultural-related patterns seem to be recorded in SPENG Introduction sections concerning this role.

The author as experiment conductor role is related to explaining research or reference to the main research procedure. The textual effect of *we* 3-grams for this role is that of foregrounding the writer as agent in situations where it is necessary to detail writers' logic or method regarding their arguments or procedures, as contributors in the development of arguments, e.g. *we studied*, *we analyzed*, *we demonstrate(ed)*, *we investigated*, *we evaluated*, *we used*, *we hypothesized*, *(to test) this hypothesis we*. These are alternatives to impersonal expressions traditionally used in academic writing.

(8) *To test this hypothesis, we developed and externally validated a contemporary nomogram predicting the probability of SVI in a split-sample cohort. Moreover, we used our external validation cohort to perform a head-to-head comparison of predictive accuracy estimates of our nomogram with that of Koh's and Partin's tools [16,19] (ENG 61)*

(9) *[...] we analysed the role of CB2 receptor in the androgen-resistant prostate cell line, PC-3, which represents the androgen-refractory phase of advanced prostate cancer. We used the anandamide analogue, R(+) Methanandamide (MET), for comparison with previous results, and a potent and selective CB2 receptor agonist, JWH-015 (JWH), as well as CB2 antagonists and RNA silencing to show the role of CB2 in PC-3 cells (SPENG 43).*

The author as opinion holder role might be interpreted as allowing the writer to inject a personal tenor into the text, the opposite of what has been done previously. Only one *we* 3-gram could be said to fit into this category, namely *we set out (to)* inasmuch as it marks the source of knowledge, and it was found in the native subcorpus. Since this role is related to tenor, it might be expected that such expressions would not occur in fixed continuous sequences. In example 10 this sequence performs the rhetorical transition from establishing the niche (Move 2) to occupying it (Move 3).

(10) *Previous studies examining gene regulation by COX-2 in CRC cells have focused on long time points and have used relatively high doses of NSAIDs (Zhang and DuBois, 2001). With this in mind, we set out to explore early changes in gene expression in CRC cells resulting from low-dose treatment with a selective COX-2 inhibitor, to improve our understanding of the early signalling events downstream of prostaglandin production. [...] The aim of this study was to explore the relationship between COX-2 and DRAK2 as a potential downstream regulator of cell survival in CRC (ENG 57).*

The author as cautious claim maker role puts forth a textual effect in which communality can be said to be strong inasmuch as there is a concern to convey a tentative tone and hedge writers' claims, e.g. *we sought to* in ENG. As can be seen in example 11, it could be used to establish the niche and mitigate the potential negative effect of the gap indicated previously. Again, although this sequence or another similar one was not observed in SPENG, as example 12 below shows, non-native scholars are also cautious when making claims.

(11) *To the best of our knowledge, no comparative study has evaluated the impact of LRN and LPN on long-term renal function. We sought to investigate the long-term effect of LPN and LRN on sCr in patients with two normal kidneys on imaging and normal preoperative sCr (ENG 78)*

(12) *On the other hand, we should consider that the cellular type (tumoural cell/stromal cell) expressing these factors might be of biological importance in breast cancer (see Decock et al, 2005, for review) (SPENG 46).*

The author as fully-committed claim maker role reveals the author as a committed community member who indicates the contribution and newsworthiness of the research. Consider for example the use of *we* 3-grams such as *we (have) recently* with an emphasis on the temporal relationship between two events; furthermore, the use of

the present perfect can be noted in the sequence *we have*, an aspect which has been suggested as typical of academic discourse in that it implies the continuing validity of earlier findings or practices (Biber *et al.* 1999: 465). For instance, in example 13 below, the phrase *we recently reported* is part of the rhetorical move 1, establishing a territory by placing their research within the field opting for the visible phrase rather than using an impersonal construction. On the other hand, example 14 shows how non-native writers also place their research visibly (*we, the most important*) within the field.

(13) *Despite advances in therapy, there exists a growing recognition of potential long-term health problems related to therapies for childhood cancer. We recently reported that by 30 years after a cancer diagnosis, 73% of survivors suffer from a chronic health condition, with 42% of these individuals having a severe or life-threatening disease or death owing to a chronic condition 2 (ENG 49)*

(14) *We have recently shown that the methylation of cytosine nucleotides in ALL cells may be the most important way of inactivating cancer-related genes in this disease (SPENG 1).*

It could be claimed that, inasmuch as the agent is marked for the rhetorical purpose of establishing a research niche, this fact points to and supports the previously observed characteristic of the Introduction section as a promotional space (Bhatia 1993, 2004).

4. Conclusion

In this paper I set out to explore the phraseological profile of *we* first-person pronoun references in native and non-native biomedical research article Introduction sections and address the question of whether culture-specific features can be explored through phraseological items in written academic English. I compared frequencies and lists of *we* 3-grams and 3-p-frames taken from the corpus (each capturing native and non-native productions) in order to see in what ways nativeness affects language patterning and displays cultural patterns with regard to *we* sequences.

The section-based frequencies show that there are differences in that English L2 scholars show a greater degree of formulaicity. The 3-gram analyses show that there is an overlap between the ENG and SPENG-based lists, but that there are also differences among the two sets of scholars which cannot be explained only by the use of personal references in Spanish L1. The latter group shows a high degree of impersonality, seemingly almost avoiding the use of n-grams with *we* references. Whereas similarities between English L1 and English L2 can be explained by the disciplinary and genre-specific phraseology, differences can be interpreted as a sign of the L1 vs L2 language user status. The 3-p-frame analysis supported the findings from the n-gram explorations and also point towards specific genre strategies that the L1 writers use such as the use of fixed expressions to point out the characteristic communicative purposes of Introduction sections.

As can be seen in the corpus used in this study, it is particularly common in Introduction sections to use *we* 3-grams as part of textual sentence stems, that is, main

clauses which present assertions and observations indirectly and emphasize the interpersonal dimension. Some of these framing sequences have specific semantic content (expressing attitude, e.g. *we sought to*), but others are discursive formulae related to specific communicative purposes in the Create-A-Research-Space (CARS) rhetorical structure of the Introduction section (see also Bondi 2010). Examples include: *in the + present study we*, *in + this study we*, *.here we*, *we show that*, *we report the*, *in addition we*, *. we also*, *. we therefore*, *we (have) recently*.

The *we* 3-gram found to be similar in the three subcorpora, ENG, SPENG and SP, namely *in+this study we* may indicate a high awareness of genre-related phraseological items commonly used in the Introduction section as discourse structuring devices on the part of Spanish scholars publishing in English and Spanish, especially for the rhetorical purpose of announcing present research to occupy the niche in Move 3 (cf. Bondi 2010).

Overall, the set of discourse roles explored here for *we* 3-grams can be said to support the claim that it is important to establish an interpersonal dimension in discourse. The textual effects did not differ for the native and non-native scholars in the cases where instances have been found in both subcorpora. This finding perhaps adds to the hypothesis that the shared *we* 3-grams can be seen as formulaic discourse frames, as suggested by their use in the L2 English non-native corpus. However, as the contrast of the discourse roles across cultures shows, and perhaps as a limitation of the approach adopted in this study for the analysis of *we* sequences, other *we* phrases associated with the stance dimension (for instance *the author as cautious claim maker*) cannot be captured with the continuous sequences methodology, and it might be the case that cultural patterns could be revealed in these instances.

These items, together with their function and use, would probably be worth focusing on in academic writing instruction for L1 and L2 English scholars. Their significance consists in shaping a consistent disciplinary discourse identity, not necessarily subjective but where writers are visible as agents in the organization of discourse as shown by the discourse-structuring phraseological items, but also involved in establishing a genre-specific interpersonal relationship with their peers.

Canagarajah (2002: 291) emphasizes that knowledge brokering should become a multidirectional rather than unidirectional process with regard to content and linguistic and rhetorical forms. In addition, a different linguistic background, when factored into scientific discourse, brings to the fore social, cultural and personal factors that interfere with the expression of disciplinary identity and draw the use of rhetorical strategies when communicating research to a multilingual discourse community. First, social and personal factors interfere in the context of publication in English. Thus academics in Spain are under pressure to contribute to the advancement of knowledge using the current lingua franca of academia, i.e. English (Curry & Lillis 2004; Pérez-Llantada *et al.* 2011; Tardy 2004); in particular, “Spain is giving priority to publication in high impact international journals” (Pérez-Llantada *et al.* 2011: 22). As such, research on phraseologies in Introduction sections is needed to reveal rhetorical strategies of non-native scholars in internationally published research papers. Second, from the standpoint of culture, phraseological items suggest that the linguistic expression of identity is different in the two cultures, i.e. Spanish and English. Spanish

scholars publishing in English-medium journals could thus be claimed to be in search of a global identity, instead of bringing to the fore their national identity, as they work in intercultural settings. Similar results have been shown in studies contrasting English and Spanish academic writing (Burgess 2002; Lorés Sanz 2006; Mur Dueñas 2009; Murillo 2011; Pérez-Llantada 2012) or English and other languages (Bondi 2004, 2009; Dahl 2004; Duszak 1997; Shaw 2003; Vassileva 2001; Yakhontova 2006). As a result, English can be argued to be “developing into an autonomous variety” which “is far removed from its speakers’ linguacultural norms and identities” (Bondi 2004: 58).

An important implication for cross-cultural research is that there are similarities of n-grams among scholars engaging in similar practices regardless of their national affiliation. At this point, the control corpus allows us to turn from a focus on the national membership of scholars to what they do, to their participation in practices.

In conclusion, the introduction of a research article is a negotiated intercultural space which promotes a shared disciplinary identity across cultures to provide a temporarily stable ground for further social action. However, the linguistic expression of identity throughout the research article does not completely erase cultural identities and it is difficult to disengage references to it in international academic writing practices (see also Gotti 2009). These findings seem to indicate that when we deal with research article Introduction sections, biomedical writers move beyond the subjective vs. objective and L1 vs. L2 English distinctions to signal disciplinary and genre-related roles and identities. In this context, communicative performance seems a more important aspect to consider than nativeness/mother language status (cf. Bondi 2010; Römer 2009; Swales 2004). Non-native authors seem to write in a style which they consider to be similar to that of English L1 scholars.

References

- Ädel A. & B. Erman 2012. Recurrent word combinations in academic writing by native and non-native speakers of English: a lexical bundles approach. *English for Specific Purposes* 31/2: 81-92.
- Atkinson D. 1999. *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London, 1675-1975*. Mahwah, N.J./London: L. Erlbaum Associates.
- Bazerman C. 1988. *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison, Wisconsin: University of Wisconsin Press.
- Bazerman C. 2001. Distanced and refined selves: educational tensions in writing. In M. Hewings (ed.), *Academic Writing in Context: Implications and Applications. Papers in Honour of Tony Dudley-Evans*. Birmingham: Birmingham University Press: 23-30.
- Becher T. & P. Trowler 2001. *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines* (2nd ed.). Buckingham: The Society for Research into Higher Education & Open University Press.
- Berkenkotter C. & T. N. Huckin 1995. *Genre Knowledge in Disciplinary Communication: Cognition, Culture, Power*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Bhatia V. K. 1993. *Analyzing Genre: Language Use in Professional Settings*. London/New York: Longman.

- Bhatia V. K. 2004. *Worlds of Written Discourse: A Genre-based View*. London/New York: Continuum.
- Biber D. & F. Barbieri 2007. Lexical bundles in university spoken and written register. *English for Specific Purposes* 26: 263-286.
- Biber D., S. Johansson, G. Leech, S. Conrad, E. Finegan & R. Quirk. 1999. *Longman Grammar of Spoken and Written English*. London/New York: Longman.
- Bondi M. 2004. If you think this sounds complicated, you are correct: awareness of cultural difference in specialized discourse. In M. Gotti & C. Candlin (eds.), *Intercultural Aspects of Specialized Discourse*. Bern: Peter Lang: 53-78.
- Bondi M. 2009. Writing economic history: the narrator and the arguer. In M. Gotti (ed.), *Commonality and Individuality in Academic Discourse*. Bern: Peter Lang: 163-191.
- Bondi M. 2010. Metadiscursive practices in introductions. Phraseology and semantic sequences across genres. *Nordic Journal of English Studies* 9/2: 99-123.
- Bondi M. & K. Hyland (eds.) 2006. *Academic Discourse across Disciplines*. New York: Peter Lang.
- Burgess S. 2002. Packed houses and intimate gatherings: audience and rhetorical structure. In J. Flowerdew (ed.), *Academic Discourse*. London: Pearson: 196-215.
- Canagarajah A. S. 2002. *A Geopolitics of Academic Writing*. Pittsburgh, London: University of Pittsburgh Press: Eurospan.
- Chen Y. H. & P. Baker. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology* 14/2: 30-49.
- Connor U. 2008. Mapping multidimensional aspects of research: reaching to intercultural rhetoric. In U. Connor, E. Nagelhout & W. Rozycki, *Contrastive Rhetoric: Reaching to Intercultural Rhetoric*. Amsterdam/Philadelphia: John Benjamins: 299-315.
- Connor U. & A. I. Moreno 2005. Tertium Comparationis: a vital component in contrastive research methodology. In P. Bruthiaux, D. Atkinson, W. G. Egginton, W. Grabe & V. Ramanathan (eds.), *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan*. Clevedon, England: Multilingual Matters: 153-164.
- Cortes V. 2004. Lexical bundles in published and student disciplinary writing: examples from history and biology. *English for Specific Purposes* 23 /4: 397-423.
- Curry M. J. & T. Lillis 2004. Multilingual scholars and the imperative to publish in English: negotiating interests, demands, and rewards. *TESOL Quarterly* 38/4: 663-688.
- Dahl T. 2004. Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics* 36: 1807-1825.
- Duszak A. (ed.) 1997. *Culture and Styles of Academic Discourse*. Berlin: Mouton de Gruyter.
- Fletcher W. H. 2007. *kfNgram*. Available at <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>. Accessed 1 December 2011.
- Fløttum K., T. Dahl & T. Kinn 2006. *Academic Voices – Across Languages and Disciplines*. Amsterdam/Philadelphia: John Benjamins.
- Flowerdew J. 2008. Scholarly writers who use English as an additional language: what can Goffman's "Stigma" tell us? *Journal of English for Academic Purposes* 7/2: 77-86.
- Gee J. P. 1999. *An Introduction to Discourse Analysis: Theory and Method*. London/New York: Routledge.
- Gledhill C. 2000. *Collocations in Science Writing*. Tübingen: Gunter Narr Verlag.
- Gotti M. (ed.) 2009. *Commonality and Individuality in Academic Discourse*. Bern: Peter Lang.
- Harwood N. 2007. Political scientists on the functions of personal pronouns in their writing:

- An interview-based study of 'I' and 'we'. *Text & Talk: An Interdisciplinary Journal of Language, Discourse Communication Studies* 27/1: 27-54.
- Hyland K. 2000. *Disciplinary Discourses: Social Interactions in Academic Writing*. Singapore: Longman.
- Hyland K. 2008. Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics* 18/1: 41-62.
- Ivanič R. 1998. *Writing and Identity: The Discoursal Construction of Identity in Academic Writing*. Amsterdam: John Benjamins.
- Knorr-Cetina K. D. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon.
- Kuo C.-H, 1999. The use of personal pronouns: role relationships in scientific journal articles. *English for Specific Purposes* 18/2: 121-138.
- Lorés Sanz R. 2006. 'I will argue that': first person pronouns as metadiscoursal devices in research article abstracts in English and Spanish. *ESP Across Cultures* 3: 23-40.
- Luzón Marco M. J. 2000. Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes* 19/1: 63-86.
- Martín Martín P. 2003. Personal attribution in English and Spanish scientific texts. *Bells* 12. Available at <http://www.publicacions.ub.es/revistes/bells12/PDF/art09.pdf>. Accessed 15 November 2010.
- Mauranen A., C. Pérez-Llantada & J. M. Swales 2010. Academic Englishes. A standardised knowledge? In A. Kirkpatrick (ed.), *The Routledge Handbook of World Englishes*. London/New York: Routledge: 634-653.
- Moreno A. I. 2010. Researching into English for research publication purposes from an applied intercultural perspective. In M. F. Ruiz-Garrido, J. C. Palmer-Silveira, & I. Fortanet-Gómez (eds.), *English for Professional and Academic Purposes* 22. Amsterdam: Rodopi: 59-73.
- Moreno A. I. 2012. The communication problems of Spanish researchers to get research articles published in Applied Linguistics English-medium journals: proposal and analysis. Available at <https://buleria.unileon.es/handle/10612/1756>. Accessed 6 June 2012.
- Mur Dueñas P. 2009. Logical markers in L1 (Spanish and English) and L2 (English) Business research articles. *English Text Construction* 2/2: 246-264.
- Murillo S. 2011. The use of reformulation markers in Business Management research articles. An intercultural analysis. *International Journal of Corpus Linguistics* 17/1: 64-90.
- Myers G. 1990. *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. Madison, Wisconsin: University of Wisconsin Press.
- Pérez-Llantada C. 2008. Humans vs. machines. A multi-perspective model for ESP discourse analysis in intercultural rhetoric research. *ESP Across Cultures* 5: 91-104.
- Pérez-Llantada C. 2010. The 'dialectics of change' as a facet of globalisation: epistemic modality in academic writing. In M. F. Ruiz-Garrido, J. C. Palmer-Silveira, & I. Fortanet-Gómez (eds.), *English for Professional and Academic Purposes* 22. Amsterdam: Rodopi: 25-41.
- Pérez-Llantada C. 2012. *Scientific Discourse and the Rhetoric of Globalization: The Impact of Culture and Language*. London/New York: Continuum.
- Pérez-Llantada C., R. Plo & G. Ferguson 2011. "You don't say what you know, only what you

- can”: the perceptions and practices of senior Spanish academics regarding research dissemination in English. *English for Specific Purposes* 30/1: 18-30.
- Rayson P. *LL Calculator*. Available at <http://ucrel.lancs.ac.uk/llwizard.html>. Accessed 1 August 2012.
- Römer U. 2009. English in academia: does nativeness matter? *Anglistik: International Journal of English Studies* 20/2: 89-100.
- Rounds P. L. 1987. Multifunctional personal pronoun use in an educational setting. *English for Specific Purposes* 6/1: 13-29.
- Samraj B. 2002. Introductions in research articles: Variations across disciplines. *English for Specific Purposes* 21/1: 1-17.
- Schulze R. & U. Römer 2008. Introduction: patterns, meaningful units and specialized discourses. *International Journal of Corpus Linguistics* 13/3: 265-270.
- Shaw P. 2003. Evaluation and promotion across languages. *Journal of English for Academic Purposes* 2: 343-357.
- Simpson-Vlach R. & N. C. Ellis. 2010. An academic formulas list: new methods in phraseology research. *Applied Linguistics* 31/4: 487-512.
- Sinclair J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Skelton J. 1994. Analysis of the structure of original research papers: an aid to writing original papers for publication. *British Journal of General Practice* 44: 455-459.
- Swales J. M. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Swales J. M. 2004. *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press.
- Tang R. & S. John 1999. The ‘I’ in identity: exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes* 18: S23-S39.
- Tardy C. 2004. The role of English in scientific communication: lingua franca or tyrannosaurus rex? *Journal of English for Academic Purposes* 3: 247-269.
- The Uniform Requirements of the International Committee of Medical Journal Editors 2009. Available at http://www.icmje.org/urm_main.html. Accessed 1 November 2010.
- Vassileva I. 2001. Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes* 20/1: 83-102.
- Yakhontova T. 2006. Cultural and disciplinary variation in academic discourse: the issue of influencing factors. *Journal of English for Academic Purposes* 5/2: 153-167.

Appendix A

List of articles by native English scholars (judged by affiliation) from the **ENG subcorpus**, and by Spanish scholars publishing in the same journals in English from the **SPENG subcorpus**, used as data for the examples cited in the paper.

ENG Subcorpus

ENG 60 *Genetics and Genomics* KA Metcalfe, A Finch, A Poll, D Horsman, C Kim-Sing, J Scott, R Royer, P Sun and SA Narod (2009) "Breast cancer risks in women with a family history of breast or ovarian cancer who have tested negative for a BRCA1 or BRCA2 mutation". *British Journal of Cancer* 100(2): 421-425. doi:10.1038/sj.bjc.6604830.

ENG 54 *Molecular Diagnostics* MD Brown, C Hart, E Gazi, P Gardner, N Lockyer and N Clarke (2010) "Influence of omega-6 PUFA arachidonic acid and bone marrow adipocytes on metastatic spread from prostate cancer". *British Journal of Cancer* 102(2): 403-413. doi: 10.1038/sj.bjc.6605481.

ENG 16 *Neoplasia* Knapper, Steven, Mills, Kenneth I., Gilkes, Amanda F., Austin, Steve J., Walsh, Val and Burnett, Alan K. (2006) "The effects of lestaurtinib (CEP701) and PKC412 on primary AML blasts: the induction of cytotoxicity varies with dependence on FLT3 signaling in both FLT3-mutated and wild-type cases". *Blood* 108:3494-3503. DOI 10.1182/blood-2006-04-015487.

ENG 13 *Plenary paper* Renshaw, Stephen A., Loynes, Catherine A., Trushell, Daniel M.I., Elworthy, Stone, Ingham, Philip W. and Whyte, Moira K.B. (2006) "A transgenic zebrafish model of neutrophilic inflammation". *Blood* 108:3976-3978. DOI 10.1182/blood-2006-05-024075.

ENG 61 *Prostate Cancer* Andrea Gallina, Felix K.-H. Chun, Alberto Briganti, Shahrokh F. Shariat, Francesco Montorsi, Andrea Salonia, Andreas Erbersdobler, Patrizio Rigatti, Luc Valiquette, Hartwig Huland, Markus Graefen, Pierre I. Karakiewicz (2007) "Development and Split-Sample Validation of a Nomogram Predicting the Probability of Seminal Vesicle Invasion at Radical Prostatectomy". *European Urology* 53: 98-105. doi:10.1016/j.eururo.2007.01.060.

ENG 57 *Molecular Diagnostics* GA Doherty, SM Byrne, SC Austin, GM Scully, DM Sadlier, TG Neilan, EW Kay, FE Murray and DJ Fitzgerald (2009) "Regulation of the apoptosis-inducing kinase DRAK2 by cyclooxygenase-2 in colorectal cancer". *British Journal of Cancer* 101 (101): 483-491. doi:10.1038/sj.bjc.6605144.

ENG 78 *Adult Urology* Kevin C. Zorn, Edward M. Gong, Marcelo A. Orvieto, Ofer N. Gofrit, Albert A. Mikhail, Lambda P. Msezane, and Arie L. Shalhav (2007) "Comparison of Laparoscopic Radical and Partial Nephrectomy: Effects on Long-Term Serum Creatinine". *Urology* 69: 1035-1040. doi:10.1016/j.urology.2007.01.092.

ENG 49 *Original report* Edward G. Garmey, Qi Liu, Charles A. Sklar, Lillian R. Meacham, Ann C. Mertens, Marilyn A. Stovall, Yutaka Yasui, Leslie L. Robison, and Kevin C. Oeffinger (2008) "Longitudinal Changes in Obesity and Body Mass Index Among Adult Survivors of Childhood Acute Lymphoblastic Leukemia: A Report From the Childhood Cancer Survivor Study". *Journal of clinical Oncology* 26(28): 4639-4645. DOI: 10.1200/JCO.2008.16.3527.

SPENG Subcorpus

SPENG 6 *Immunobiology* Alberto Bosque, Juan Ignacio Aguiló, M^a Ángeles Alava, Estela Paz-Artal, Javier Naval, Luis M. Allende and Alberto Anel (2006) "The Induction of Bim Expression in Human T Cell Blasts is Dependent on Non-Apoptotic Fas/Cd95 Signalling". *Blood* 1: 1-34. DOI 10.1182/blood-2006-05-022319.

SPENG 42 *Molecular Diagnostics* C Martínez-Campa, A González, MD Mediavilla, C Alonso-González, V Alvarez-García, EJ Sánchez-Barceló and S Cos (2009) "Melatonin inhibits aromatase promoter expression by regulating cyclooxygenases expression and activity in breast cancer cells". *British Journal of Cancer* 101(9): 1613-1619. doi:10.1038/sj.bjc.6605336.

SPENG 29 Julio García-Suárez, Helena Bañas, Ignacio Arribas, Dunia De Miguel, Teresa Pascual and Carmen Burgaleta (2006) “Dose-adjusted EPOCH plus rituximab is an effective regimen in patients with poor-prognostic untreated diffuse large B-cell lymphoma: results from a prospective observational study”. *British Journal of Haematology* 136: 276-285. doi:10.1111/j.1365-2141.2006.06438.x.

SPENG 43 *Translational Therapeutics* N Olea-Herrero, D Vara, S Malagarie-Cazenave and I Díaz-Laviada (2009) “Inhibition of human tumour prostate PC-3 cell growth by cannabinoids R(+)-Methanandamide and JWH-015: Involvement of CB₂”. *British Journal of Cancer* 101(6): 940–950. doi:10.1038/sj.bjc.6605248.

SPENG 46 *Molecular Diagnostics* LO González, I Pidal, S Junquera, MD Corte, J Vázquez, JC Rodríguez, ML Lamelas, AM Merino, JL García-Muñiz and FJ Vizoso (2007) “Overexpression of matrix metalloproteinases and their inhibitors in mononuclear inflammatory cells in breast cancer correlates with metastasis-relapse”. *British Journal of Cancer* 97(7): 957-963. doi:10.1038/sj.bjc.6603963.

SPENG 1 José Román-Gómez, Lucia Cordeu, Xabier Agirre, Antonio Jiménez-Velasco, Edurne San José-Eneriz, Leire Garate, María José Calasanz, Anabel Heiniger, Antonio Torres, Felipe Prosper (2006) “Epigenetic Regulation of Wnt Signaling Pathway in Acute Lymphoblastic Leukemia”. *Blood* 1: 1-29. DOI 10.1182/blood-2006-09-047043.