LA QUALITÉ DANS L'ENSEIGNEMENT DES LANGUES ET DANS LA RECHERCHE

A PHRASEOLOGICAL APPROACH TO SPANISH SCHOLARS' RHETORICAL STRATEGIES IN ENGLISH-MEDIUM RESEARCH PUBLICATION

Oana Maria CARCIU
University of Zaragoza, Spain
ocarciu@unizar.es

Abstract:
This study particularly analyzes phraseological patterns of first person plural references as the expression of authorship in writing. Corpus evidence is obtained for two communities contrasted, that is English native scholars and their Spanish counterparts. Data on common authorship-related phraseological patterns retrieved with WSConcgram (Scott, 2008) from a section coded-corpus of research articles indicate that these phraseologies are section-specific, but that there are also differences in their configuration across cultures (i.e. Spanish and Anglo-American). Finally, these findings are discussed in view of the potential implications on Spanish scholars' rhetorical strategies who need to adjust their language to the rhetorical standards of English-medium journals (Pérez-Llantada, 2012) when engaging in global debates.

Keywords:
English-medium publication, authorship, biomedical research article, contrastive rhetoric, Spanish context.

1. Introduction

The currency of English as the common language of biomedical research publication falls within the realm of intercultural encounters occurring at a global level in scientific research. These encounters are indicative of the fact that at present scientists with different linguistic backgrounds

"*function more as members of an international community having one common language than as members of national communities, both in their writing and in their selection of background readings*"[1].

[1] Truchot 1997, p. 67.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

Therefore it is imperative that these scientists have a good command of English since this is the language which facilitates access to information and communication with fellow academics internationally[2]. These issues bear on Spanish scholars who, as Moreno argues, "*are gradually moving towards publishing their research in international journals*"[3], a trend experienced particularly in the domain of biomedicine. According to Moreno, "*until very recently Spanish scholars had little or no chance to use English for real academic purposes*"[4]. However, with the precedence of English over other languages in the internationalization process, there has been an increase in requests for

"*courses in skills relevant to publishing in English in order to enhance their chances of seeing their work accepted by international journals*"

in the Spanish academic context[5]. In addition, this trend has entailed the need to advance research as regards Spanish-English rhetorical variation in academic writing.

As a consequence, studies carried out on academic discourse have concerned themselves with aspects that facilitate Spanish scholars' access to the international discourse community, more specifically focusing on the range of prevalent academic discourse practices in journals of prestige, i.e., the Anglo-American ones[6]. These rhetorical practices pertain to communicative competence in English-medium research publication.

In order to observe and apprehend the rules of engagement that such intercultural situations compel Spanish scholars to follow when writing for publication in international journals in English, research on rhetorical strategies in academic discourse has prompted the compilation of corpora and the use of corpus tools. The availability of authentic samples of academic discourse purveys accurate descriptions, or as Partington indicates, "*knowledge of which collocations are normal in which environments*"[7]. This knowledge is assumed to be context-related in as much as language users seemingly make certain decisions about language use depending on what seems effective and appropriate language in the context of situation in accordance with academic discourse conventions.

Therefore, the current study explores the conventionalized word usage of *we* and other words that commonly go together in a corpus of biomedical research articles and thus describes the typical patterns of first person plural references, i.e., *we* phraseologies, based on corpus evidence. Such patterns are the expression of authorship in writing, namely show how the text points out to the figure of the

[2] Ammon, 2001.
[3] Moreno, 2010, p. 57.
[4] Moreno, 2010, p. 58.
[5] Moreno, 2010, p. 57.
[6] Lores, 2006; Mur, 2009; Murillo, 2011.
[7] Partington, 1998, p. 18.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

author which is deemed to be outside it on account of the objectivity requirement in scientific discourse[8]. The focus on patterns aims at gaining insight into the main thrust of routinized practice on textual patterning in association with the need of Spanish scholars publishing internationally to adjust their language to the rhetorical standards that authorship supposes in English-medium journals.

2. Theoretical framework

The phraseological approach to authorship is framed in this study within genre analysis[9]. This framework has been selected in view of gaining insight into the two concepts of authorship and phraseology in biomedical research articles, with particular reference to introduction sections. As a sample of academic discourse, the introduction section of biomedical research articles fits a standardized rhetorical structure instantiated as a series of moves and steps and labelled 'Create-A-Research-Space' (CARS)[10]. The CARS model shows the importance of the general topic and particular issue concerned in Move 1, outlines the gap to be filled in Move 2, and finally in Move 3 the article purpose is made explicit. The obligatory character of these three rhetorical moves underpins the claim that

"*[t]he opening paragraph requires the writer to make some decisions about the amount and type of background knowledge to be included, an authoritative versus a sincere stance [...] the appropriateness of the appeal to the readership, and the directness of the approach*"[11]

It has been shown that the role of introduction sections is promotional as well[12] a feature which could generate mentions to the authors of the text regardless of the inherent objectivity of scientific discourse. The Introduction section thus becomes one of the most demanding ones in terms of rhetorical effort.

The implications of this rhetorical structure for the conceptualization of authorship can be seen at the level of journal guidelines bearing on this aspect. Generally, biomedical publication follows the URM guidelines which delineate authorship along three lines: (1) substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; (3) final approval of the version to be published.

[8] Bazerman, 1988; Atkinson, 1999.
[9] Swales, 1990, 2004.
[10] Swales, 1990, 2004.
[11] Swales, 1990, p. 137.
[12] Bhatia, 2004.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

Here the focus is on explicit mentions to author in the text and the discourse instantiation of researcher involvement in conception and design, acquisition of data and analysis or interpretation of data. The linguistic realization of these mentions is the first person plural pronoun. However, it has been argued that the word is not the most useful unit of analysis[13]. As such this analysis relies on corpora and corpus tools to explore word combinations with first person plural reference *we* instead of a single word. Co-selected words are argued to recur in patterns associated with specific meanings distributed over several co-occurring words[14]. The main corpus linguistics assumption in the study of discourse is that

"*[a] written text (...) consists of a sequence of word-forms (word-tokens) which occur with different frequencies in particular distributions*".[15]

A particular distribution is described as co-selected words which do not combine and recombine freely and randomly, but recur in patterns. Two main implications can be observed. The first one underpins the pervasiveness of patterns as a result of the frequent occurrence of words in a pattern, while the second is related to the configuration of words in a pattern. First, patterns or the particular arrangement of words in a sequence are established on account of their frequency in the text. However, these patterns are described as frames, namely units which are not fixed but variable in their potential lexical realizations. Corpus linguistic research has lent support to the broad usage of word combinations and typical phraseological patterns as conventional units of meaning in discourse[16].

In this paper, *we* phraseologies, that is, continuous or variable sequences of word forms that cluster around first person plural references in the biomedical research article introduction section are explored as indicators of authorship which cohere with the specific communicative purpose of the writer in a certain part of the introduction. In as much as these phraseologies substantiate canonical, conventional forms which are "*evidence of norms of usage across a community of speakers*"[17], this study aims to: (1) identify and gain comparative insight into the occurrence, structure and distribution of *we* recurrent word combinations in the introduction section of the English-medium biomedical research article; (2) cross-check the use of *we* phraseologies across languages, that is in English L1 and English L2 (Spanish – English).

[13] Römer, 2009.
[14] Hunston & Francis, 2000.
[15] Stubbs 2007, p. 144.
[16] Biber, Conrad, & Cortes, 2004; Hunston & Francis, 2000.
[17] Stubbs, 2007, p. 147.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

3. Corpus data and analytical steps

For the purposes of the present study, the Biomedical and Health Sciences (BHS) component of the *Spanish-English Research Article Corpus* (SERAC[18]) 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). Each text file has been labeled with the name of the folder and a number, e.g. ENG 1, ENG 2, etc. The analysis in this paper is carried out only on the ENG and SPENG texts. 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. 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 journals[19] & 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

[18] 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.

[19] Ten articles were extracted out of each journal, except for the SP corpus.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

The automatic extraction of phraseologies relies on the operationalization of multiword combinations as concgrams. A concgram is

"*all of the permutations of constituency variation and positional variation generated by the association of two or more words*"[20].

The starting point is what the procedure terms 'origin': an iterative search-based approach is adopted to the collection of concgrams which begins with a single word as the origin to search for all the concgrams of length 2 within a specific window (see Stubbs for the window span set at 5 to the right and 5 to the left), then each of these becomes an origin of length 2 to search for all the concgrams of length 3; this is repeated with an origin of length 3 to find all concgrams of length 4, and so on. To conduct these analytical steps the WSConcGram implemented in WordSmith Tools Version 5.0[21] has been used. The first step was to generate two indexes using a span of 10 words (5 to the left and 5 to the right), at a threshold of 5 for each word in the concgram and stopping at sentence boundaries. *We* has been chosen as a single origin and the flat list has formed the basis of subsequent analysis. To identify phraseologies and their configurations in terms of positional and constituency variation, fixed vs. variable slots marked with [*], concgrams have been displayed in context through the Concord tool.

4. Data analysis

This section reports on *we* phraseologies in ENG and SPENG introduction sections, i.e., their systematization and configuration in terms of position of plural references *we* and constituents variation in the phraseological unit. In addition, these phraseologies are compared across cultures, namely English L1 and English L2. The description of *we* phraseologies is presented in association with an analysis of the typical meanings which corroborate the expression of authorship in writing, namely researcher involvement in conception and design, acquisition of data and analysis or interpretation of data.

On the basis of the flat list displayed in WSConcGram, each WE concgram has been analysed in context through the Concord tool, which is a valuable tool in order to check for possible concgrams overlap, as has been previously observed in the case of n-grams[22]. As a result, the retrieved WE concgrams were checked manually for overlapping instances. The procedure for dealing with overlapping concgrams involved merging examples such as *we analysed the expression of* and *confirm we the* and *results we the expression* and *these results we the* in order to obtain one phrase unit, namely *confirm these results we analysed the expression*

[20] Cheng et al. 2006, p. 414.
[21] Scott, 2008.
[22] Chen & Baker, 2010.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

of. Findings point to several methodological implications of the use of the concgram procedure. First, it seems that there is never overlap with the concgram procedure when the origin is the same (*we*), but there is overlap, as in the above-mentioned example where alternative origins can be considered for the concgram, e.g., *confirm*, *results*, *these*. This observation entails that the concgram is considered here as an alternative procedure to identify phrase units better than other techniques, i.e., n-grams extraction of contiguous words which correspond to lexical bundles, due to the fact that they record both positional and constituency variation.

In addition, once the concgrams have been extracted, it is possible to see that some concgrams are self-standing multi-word sequences which can be found as either sentence-initial or premodified phrase units, e.g., *we recently reported that*, *here we show that*. However, other concgrams do combine together, e.g., the concgrams *this we the* and *we studied the* form a unique phraseological unit, i.e., *this we studied the*. This phenomenon may be explained by the fact that some words which combine recurrently enough to form a concgram might be in a premodifying position with respect to another concgram.

These observations provide a record of three modes in which the concgrams can be found, i.e., self-standing sentence-initial (C), self-standing premodified (PC), and combined concgrams (CC). This systematization of concgrams into the resulting types of phraseological units allows comparison across ENG and SPENG (differences and similarities for each mode across the two groups of texts are shown in Figure 1). As can be seen, variability can be reported in the case of self-standing sentence initial concgrams in as much as there seems to be an overuse in English L1 texts as compared to English L2 (Spanish) texts. The reverse trend can be observed for combine concgrams, which result in a lower score of *we* phraseological units for SPENG scholars. In turn, premodified concgrams seem to be quantitatively similar in ENG and SPENG.

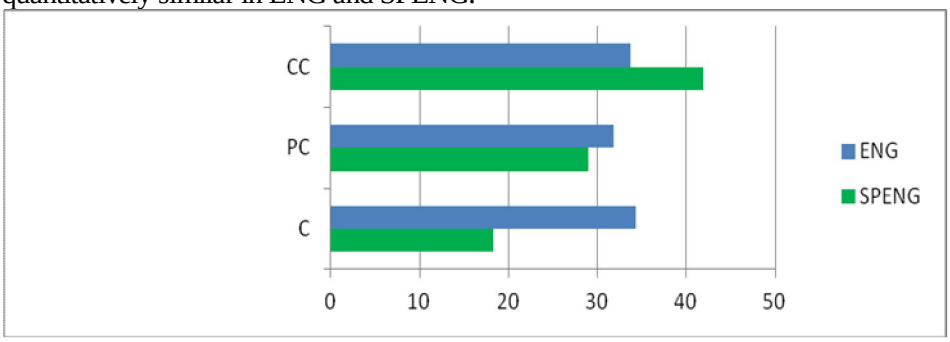


Figure 1. Sentence-initial (C), premodified (PC) and combined concgrams (CC) in ENG and SPENG

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

The concgrams extracted account for the positional configuration of words which combine to form a phraseological unit. The codification of the first person plural pronoun *we* as N permits observation of the preferred position of this element in the phraseological unit in view of comparison across the two groups of texts. In Figure 2, as the patterns show, the plural reference *we* most is more frequent in initial position in a phraseological unit, which is explained by its grammatical function as a subject. However, differences can be identified as regards the positional configurations displayed in Figure 2 below, which amounts to higher frequencies in ENG as compared to SPENG.

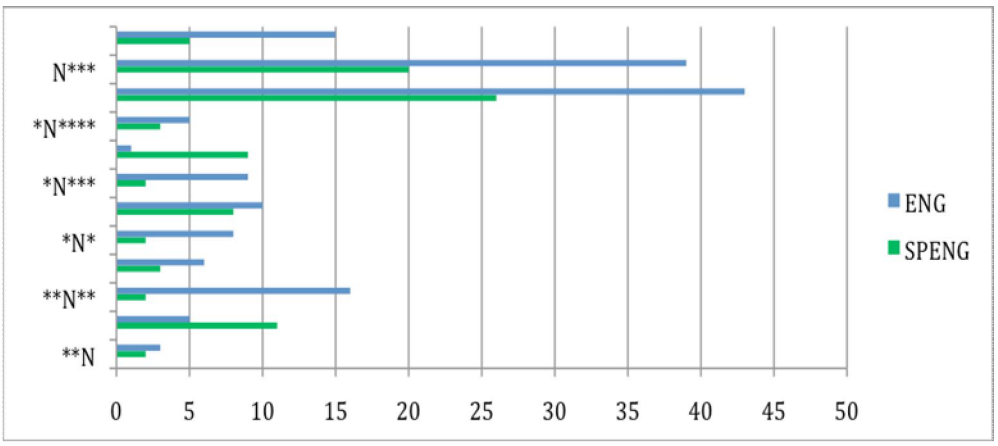


Figure 2. Sentence-initial (C), premodified (PC) and combined concgrams (CC) in ENG and SPENG

With regard to the configuration of fixed, e.g., 123 and variable slots (marked with *), e.g., 12*3, more variety, not only quantity, can` be seen to emerge within *we* phraseological units retrieved from the texts of English L1 scholars. Therefore, constituency configuration confirms the general trend observed from the systematization of concgrams and positional configuration which give preliminary evidence of both differences and similarities between ENG and SPENG texts.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

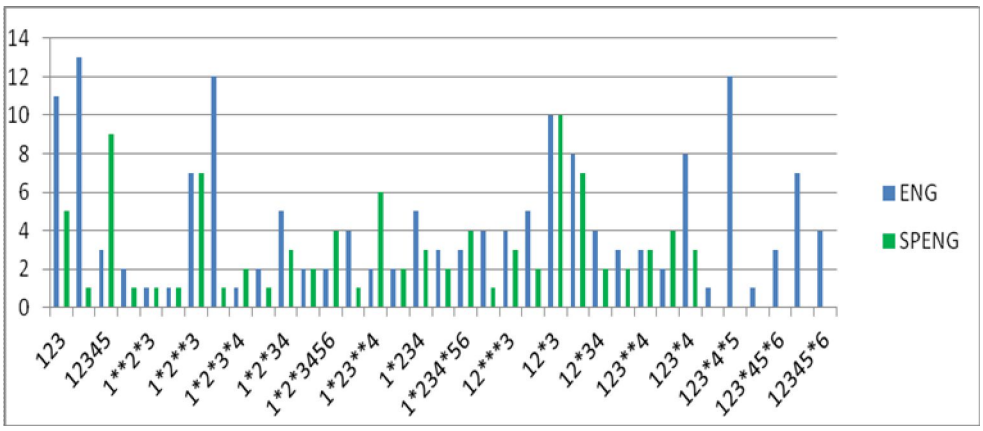


Figure 3. Constituency configuration in ENG and SPENG

At discourse level, *we* phraseological units automatically extracted function as discourse frames[23], i.e., *we (show) that*, *we (report) the*, thus of high importance in framing arguments as well as to reflect the specific communicative purpose of the writer in a certain part of the introduction. The textual effect of *we* phraseologies may consist in signalling writers' intentions as in the following examples which instantiate the announcing present research 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

(1) Here we show *that dietary PUFAs themselves are not strong stimulators of CaP invasion but require adipocyte processing. (ENG 54)*

(2) 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 (SPENG 6)*

For instance, in example (3) below, the phrase *we recently reported* is part of the rhetorical move 1, establishing a territory. As can be seen, in order to place their research within the field scholars opt for the visible phrase rather than using an impersonal construction. Example (4) corroborates that non-native writers also place their research visibly (*we, the most important)* within the field.

[23] Biber & Barbieri, 2007.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

(3) *Despite advances in therapy, there exists a growing recognition of potential long-term health problems (...).* We recently reported *that by 30 years after a cancer diagnosis, 73% of survivors suffer from a chronic health condition (ENG 49).*

(4) 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[24].

Another textual effect is related to explaining research or reference to the main research procedure. In this case the writer is foregrounded 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.

(5) *Moreover,* we used *our external validation cohort to perform a head-to-head comparison of predictive accuracy estimates of our nomogram (ENG 61)*

(6) We used the *anandamide analogue, R(+) Methanandamide (MET), for comparison with previous results (SPENG 43).*

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.

5. Discussion

Preliminary evidence substantiated by the analysis of *we* phraseologies point to important implication for cross-cultural research. As can be seen, there are similarities of *we* phraseologies among scholars engaging in similar practices regardless of their national affiliation. However, there are also differences which cannot be explained simply by contrasting samples of English L1 and English L2. At this point, the use of a control corpus, i.e., samples of Spanish L1 texts, would allow us to increase the reliability of these findings, which is a limitation of the present study that could be addressed in future research. With regard to authorship,

[24] Bhatia, 2004.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

the textual effects of *we* phraseologies entail the claim that they are resources the writer draws on to display knowledge of the conventions in the discipline, professional knowledge, discourse as professional practice, professional space.

Overall, the phraseological approach endorses the postulate that words co-occur in patterns in discourse, associated with a specific communicative function. From this standpoint it can be surmised that the introduction section of an English-medium research article is a negotiated intercultural space which promotes a shared disciplinary identity across cultures. However, as the case of Spanish scholars reveals, the linguistic expression of authorship does not completely erase cultural-specific traits and it is difficult to disengage references to them in international academic writing practices[25]. These findings seem to indicate that, in research article introduction sections, biomedical writers move beyond the subjective vs. objective and L1 vs. L2 English distinctions to signal disciplinary and genre-related rhetorical strategies. In this context, Spanish scholars seem to write in a style which they consider to be similar to that of English L1 scholars. On the other hand, the coexistence of similar and different features in English L2 texts require further research on the adjustments needed to publish in international journals in English and descriptions of diglossic contexts and hybrid texts[26].

Bibliography

AMMON, U., 2001, *The Dominance of English as a Language of Science. Effects on Other Languages and Language Communities*, Berlin: Mouton de Gruyter.

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.

BHATIA, V. K., 2004, *Worlds of Written Discourse: A Genre-based View*. London/New York: Continuum.

BIBER, D.; Barbieri, F., 2007, "Lexical bundles in university spoken and written registers". *English for Specific Purposes*, 26/3: 263-86.

BIBER, D.; CONRAD, S.; CORTES, V., 2004, "If you look at…: Lexical bundles in university teaching and textbooks". *Applied Linguistics*, 25/3: 371-405.

CHENG, W; GREAVES, C; SINCLAIR, J; WARREN, M., 2009, "Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams". *Applied Linguistics*, 30/2: 236-252.

[25] Gotti, 2009.
[26] Pérez-Llantada, 2012.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)

CHEN, Y. H.; BAKER, P., 2010, "Lexical bundles in L1 and L2 academic writing". *Language Learning and Technology,* 14/2: 30-49.

GOTTI, M. (ed.) 2009. *Commonality and Individuality in Academic Discourse*. Bern: Peter Lang.

HUNSTON, S.; FRANCIS, G., 2000, *Pattern grammar: A corpus-driven approach to the lexical grammar of English.* Amsterdam: John Benjamins.

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.

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.

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.

PARTINGTON, A., 1998, *Patterns and meanings: Using corpora for English language teaching and research*. Amsterdam: Benjamins.

PÉREZ-LLANTADA, C., 2012, *Scientific Discourse and the Rhetoric of Globalization: The Impact of Culture and Language*. London/New York: Continuum.

RÖMER, U., 2009, English in academia: does nativeness matter? *Anglistik: International Journal of English Studies,* 20/2: 89-100.

SCOTT, M., 2008, *WordSmith Tools Version 5.0*. Lexical Analysis Software, Liverpool.

STUBBS, M., 2007, "On texts, corpora and models of language". In Hoey, M. et al. (ed.). *Text, Discourse and Corpora: Theory and Analysis*. London: Continuum: 127-161.

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.

TRUCHOT, C., 1997, "The spread of English: from France to a more general perspective". *World Englishes*, 16: 65-76.

BDD-A3980 © 2013 Editura Muzeul Literaturii Române
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-14 13:09:58 UTC)