

Pedro A. Fuertes-Olivera

Pedagogical Application of Specialized Corpora in ESP Teaching: the case of the *UVaSTECorpus*

Abstract

This article contributes to defining the concept of specialized corpora, reviews the rationale for using them instead of general corpora in teaching activities, and offers the state of art in both corpus-based and corpus-driven approaches to ESP teaching. It also explains some decisions taken regarding the compilation of the *University of Valladolid Corpus of Written Scientific and Technical English* and illustrates some uses of the corpus. In particular, it presents some tasks with concordances and defends that ESP students should be taught the niceties of lexical gender as it is a grammatical category with social and/or ideological implications.

Keywords: specialized corpora, ESP, lexical gender, classroom activities.

1. Introduction: Corpus Linguistics and the Study of the English Language

As a relatively new approach to language studies, corpus linguistics has witnessed that the number and depth of many corpus approaches to the study of the English language is constantly increasing. Since the 2000s, we have observed the development of a complementary process aiming at building both *giga-corpora* (i.e.,

billion-word corpora compiled using websites and newswire texts as data sources¹), and smaller specialized corpora designed for ESP studies.

A specialized corpus comprises representative oral and/or written texts which reflect the kind of language of a particular domain. Although there is no limit to the degree of specialization involved, specialized corpora tend to accord to a set of parameters which refer to genre, topic, size, text type, and language variety. Flowerdew (2004: 21) summarizes the parameters by which corpus linguists tend to define specialized corpora, with examples illustrating them (Table 1). She adds that although the parameters in Table 1 are presented as discrete categories, there is an overlap between some of them; for example, *contextualization* is also an aspect of genre.

Parameters	Details/Examples
Specific purpose for compilation:	To investigate particular grammatical, lexical, lexico-grammatical, discoursal or rhetorical features
Contextualization:	Setting (e.g. lecture hall) Participants (role of speaker/listener; writer/reader) Communicative purpose (e.g. promote, instruct)
Size:	
whole corpus	1-5 million words
sub-corpus or small-scale corpus	20,000 – 250,000 words
Genre:	Promotional (grant proposals, sales letters)
Type of text/discourse:	Biology textbooks, casual conversations
Subject matter/topic:	Economics, the weather
Variety of English	Learner, non-standard (e.g. Indian, Singaporean)

Source: Flowerdew (2004: 21)

Table 1: Parameters for defining corpora as specialized

Seven arguments are usually indicated for explaining why general corpora may be unsuitable for investigating specialized language. First, general-purpose corpora have been compiled for the purpose of inferring generalizations about the language as a whole, or about broad categories of texts (Flowerdew, 2004). Second, although a general corpus might contain a specialized sub-corpus, logistically it may be difficult to access such a corpus in a general corpus as the search fields have not been set up with this purpose in mind (Lee, 2001). Third, some types of discourse – for example, *occluded genres* (cf. Swales 1996) – are not easily accessible for compilation and are usually left out of general corpora because of pragmatic and economic concerns (Burnard, 2002). Four, some general corpora consist of samples of texts. Text segments do not lend themselves to top-down genre-based analyses (Swales, 2002). Five, specialized corpora favor qualitative-based analyses as their size and composition make them more manageable for qualitative studies (Flowerdew, 2005). Six, specialized corpora allow working on the semantic or discourse level of corpus

¹ For example, the *English Gigaword Corpus*, produced by the Linguistic Data consortium. See: <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2003T05>.

data through inserting tags. This kind of tagging would have to be done manually and would therefore be well-nigh impossible to carry out on very large-scale corpora running into millions of words (Biber et al. 2004). Seven, from a methodological point of view, analysis with specialized corpora favor comparison and comparison uncovers differences almost regardless of size (Sinclair, 2001). For example, comparing the presence of social titles in the Wolverhampton Corpus of Written Business English and in the British National Corpus allows researchers to draw conclusions on the presence of women in professional domains (Fuertes-Olivera, 2007a).

After reviewing current work with specialized corpora, the article describes some key features of the *University of Valladolid Scientific and Technical Written English Corpus (UVaSTECorpus)*. The description illustrates how a localized, in-house corpus can be compiled; it also comments on some findings that add to the amount of knowledge that the use of specialized corpora is bringing to our professional activity. In particular, it explains how concordances can be used in the classroom (for example, for enhancing terminological work), and some pedagogical applications drawn from a better understanding of the grammar of ESP (Swales 2006). For example, the research findings of this and other related studies could positively impact upon the teaching of *lexical gender* considering that there is a need to make the concept of *gender sensitivity* present in LSP activities (Bowker 2001; Fuertes-Olivera 2007a).

2. Corpus Studies of Academic and Professional English

Recent research pays attention to the ethnographic dimension and extra-linguistic content for interpretation of corpus data, to under-researched discourse genres and varieties of English, and to the changing-face of English for professional communication in the era of globalization. This has led to working with two complementary types of specialized corpora: localized, in-house corpora; and open access corpora.

2.1. Teaching and research activities with in-house localized corpora

By localized in-house corpora I understand corpora compiled by an individual or group of individuals, who work in the same institution and aim at drawing both pedagogical applications for their daily teaching activities, and empirical data for their research. These works can be grouped into four types:

- a) corpus work that reports practical comments on the pros and cons of using corpora in the classroom, paying attention to working with concordances (Gavioli, 2005), or integrating the lexical approach with data-driven corpus-

based methodology (Mudraya, 2006), or using corpus data for unveiling semantic prosodies (Nelson, 2006).

- b) corpus work that compares the language produced by native and non-native speakers (NNSs) in order to gain more precise knowledge of how NNSs understand the characteristics of the genre which allows teachers to focus on the teaching activities demanded by their NNS students. In particular, negotiations in service telephone calls (Bowles, 2006), the use of modals in forecasting (Donahue, 2006), epistemic modality (Gabrielatos, and McEnery, 2005), and the syntactic competence needed to manipulate information structure in oral presentations (Rowley-Jolivet and Carter-Thomas, 2005) has been investigated.
- c) corpus work that investigates the discursive features of a discipline or group of disciplines which provides information on how to organize the teaching tasks. Some common topics are politeness, mood and modality, personal reference, metadiscourse, hedging devices, phraseological patterning and interactive features such as *personal deictics*, *markers*, and *imprecise quantifiers* (Charles, 2006; Flowerdew & Wan, 2006; Harwood, 2005; Hyland & Tse, 2005; Webber, 2005). Flowerdew & Wan (2006: 150), for example, indicate that tax accountants use very formulaic and standardized templates, which must be incorporated into the teaching of professional writing.
- d) corpus work that focuses on the rhetorical structure of genres such as 'e-mails', 'letters to the editor', 'research articles', 'abstracts' and 'literary reviews' (Gimenez, 2006; Kanoksilapartham, 2005; Kwan, 2006; Lim, 2006; Magnet & Carnet, 2006; Pecorari, 2006; Samraj, 2005). Findings are being used to familiarize students with the variation found in academic and professional texts not just across genres but also across disciplinary boundaries.

2.2. Teaching and research activities with open access corpora

By open access corpora I refer to some corpora which have been compiled by one or more institutions with the aim of making them available to the research community which can engage into two main complimentary tasks: (i) to replicate previous analyzes in order to falsify the hypothesis being tested; (ii) to broaden the scope of research concerned with both drawing pedagogical applications and allowing the grammar of ESP to emerge. Below, I will comment on some recent findings drawn from open access corpora:

- a) The Michigan Corpus of Academic Spoken English (MICASE) is a spoken language corpus of approximately 1.9 million words of contemporary university speech recorded at the University of Michigan². Using this corpus,

² See: <http://quod.lib.umich.edu/m/micase/>

Simpson (2004), for example, identified a list of all three-, four-, and five-word formulaic expressions in MICASE and specified which expressions are typical of academic discourse differentiating between those more likely to occur in monologic academic discourse and those most used in interactive academic situations. She also found some interesting differences in the formulaic expressions used by professors versus those used by students.

- b) The TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) Corpus contains ten different registers including service encounters, office hours, classroom lectures, textbook, syllabi brochures, and university webpages, thus providing a tremendous resource for describing the language that students encounter in a university setting and also for exploring linguistic variation in different university settings (e.g. classroom teaching vs. labs/in-class sessions or classroom teaching vs. textbooks). This 2.7 million word corpus was designed for two major purposes: (i) to study the patterns of language use found in academic registers; and (ii) to develop procedures for assuring that the language used in TOEFL exam tasks is representative of real life language uses (Biber et al. 2004).
- c) The British Academic Spoken English (BASE) Corpus and the British Academic Written English Corpus (BAWE) Corpus were developed under the leadership of Hillary Nesi³. They allow investigation on the following: frequency and range of academic lexis; the meaning and use of individual words and multi-word units; the structure of academic lectures; the pace, density and delivery styles of academic lectures; the discourse function of intonation; patterns of interaction, including turn-taking and topic selection; the interplay of visual and aural stimuli; the representation of ideas and the expression of attitudes; identifying the characteristics of proficient student writing, the similarities and differences between genres produced in different disciplines, and at different stages of university study, etc. (Nesi and Gardner 2006).
- d) The Cambridge and Nottingham Corpus of Business English (CANBEC) is a collection of spoken business English recorded in companies of all sizes, from big multinational companies to small partnerships. Formal and informal meetings, presentations, conversations on the phone, over lunch etc. were recorded, and typed into the computer for analysis by authors and editors. McCarthy & Handford (2004: 187), for example, summarize some pedagogical implications regarding spoken business English (SBE):
 - (i) a good deal of the linguistic content of SBE is shared with casual conversation. Hence, a comprehensive SBE pedagogy would prioritize awareness of areas such as personal deixis, face-protection and indirectness;

³ <http://www.rdg.ac.uk/slals/base/>

- (ii) business English materials must also focus on abstract states and qualities, such as politeness. The evidence shows that mitigating face-threats is vital;
- (iii) skill in hedging and use of purposive vagueness must be stressed;
- (iv) close observation of the achievement of speech acts such as requests and directives while maintaining comity in SBE contexts is a useful awareness-raising exercise;
- (v) although many users of SBE will be using it as a *lingua franca* in non-native business contexts, successful business exchanges still rest on good interpersonal relations. Getting things done can be facilitated by a greater awareness of what the linguistic resources have to offer;
- (vi) as more spoken business corpora become available, data-driven learning using concordances and open access to corpus files will enable business users of English to access resources aligned to their own situations and linguistic goals.

3. Compiling your In-house corpus: The *UVaSTECorpus*

Although we are working for converting the *University of Valladolid Scientific and Technical Written English Corpus (UVaSTECorpus)* in an open access corpus, accessible to the research community through the Internet, at the moment it is a localized, in-house corpus. Below, I will briefly explain some decisions which were taken before and during the process of compilation of such a corpus.

This corpus stands at our original target of 3,000,000 words of scientific and technical English (my emphasis). It was designed by Fuertes Olivera (2007b), collected at the University of Valladolid by José María Rodrigues and Pedro A. Fuertes-Olivera (April 2005, April 2006), and enriched with information on text type, genre, author(s), geographical variety, subject matter, year of production, purpose of the text, and origin of the documents.

Assuming that a corpus should be compiled falling back on non-probability sampling techniques involving *judgment* and *convenience* (Meyer 2002:44), we decided that the Corpus would be randomly selected from English collections of scripted speech and edited writing comprising scientific and technical informational and instructional genres, adequate for investigating particular grammatical, lexical, lexico-grammatical, discoursal or rhetorical features of scientific and technical English.

The documents are restricted to the last 25 years, and are the product of three types of users: native researchers; non-native researchers using English as a *lingua*

*franca*⁴; and official bodies in the UK and the USA producing different types of reports and/or instructional documents on 'science and technology'. Hence, the corpus is subdivided into two main sub-corpora of documents: informational learned 'research' texts (70.5% of the corpus) and informational reportage and instructional administrative documents (29.5% of the corpus).

The former comprise texts produced by researchers in different fields published in peer-reviewed international journals. This sub-corpus groups publications either as 'research' or as 'review', depending on the purpose of the paper. If the paper presents new findings or methodologies we classified it as "research". If it summarizes the state of the art in a particular topic we grouped it as "review". Considering the amount of pages usually devoted in journals to the two categories, we set a 3:1 selection criterion which resulted in 71.5% of the sub-corpus for 'research articles' and 28.5% for 'review articles'.⁵ All the texts were taken from leading journals indexed in the ISI Science Citation Index covering Science and Technology domains, and covering the most important domains dealing with science and technology. For example, in the genre 'research article' included in the sub-corpus 'research', the following domains and number of words per domain were compiled: (i) applied mathematics and statistics (237,911 words), (ii) physics, X-Rays and radiology (138,615 words), (iii) chemistry and food processing (174,921 words), (iv) engineering (154,784 words), (v) materials, composites, and minerals (180,574 words), (vi) environment (206,437 words), (vii) computing (147,399 words), (viii) biotechnology (137,221 words), and (ix) information systems (151,382 words) (Figure 1).

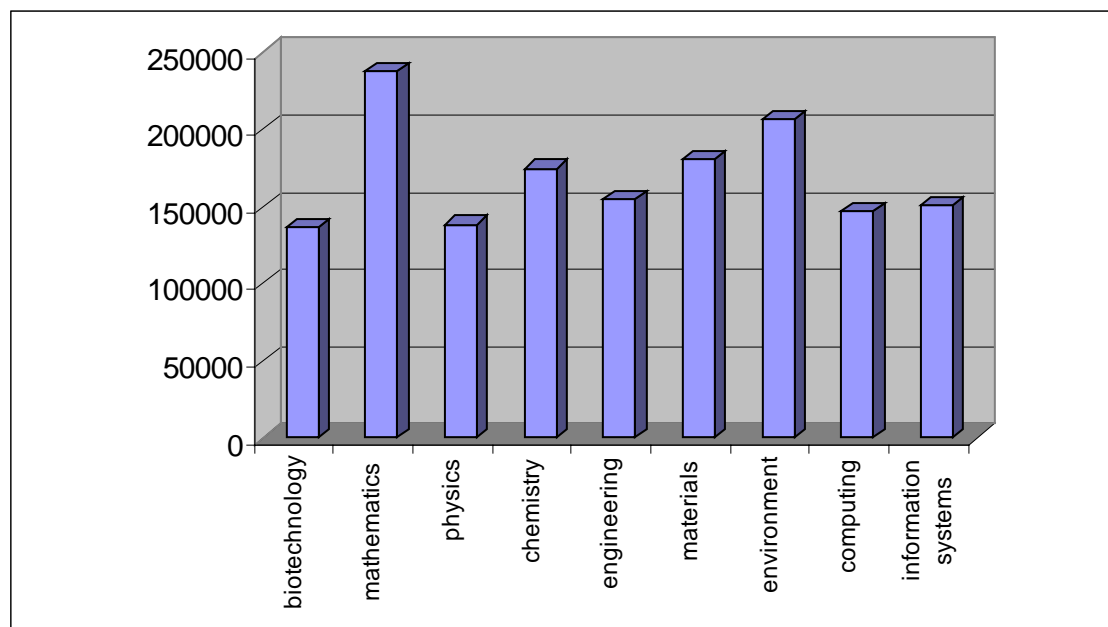


Figure 1: Number of words per domain of the research articles included in the sub-corpus 'Research'

⁴ Around 80% or more of scientific and technical research is published in English (cf. Crystal 1997).

⁵ including 'profiles', 'PhD summaries, letters, discussions, notes, etc.

The latter consists of official documents produced by government bodies and professional societies. These were grouped into two main categories: informational reports; instructional documents. Informational reports are the work of national and/or international committees (for example, the National Science and Technology Committee) and aim at reviewing national and/or international policies on science and technology (for example, *national priorities in science and technology policy*), or informing on research priorities, or other science and technology issues (for example, *the science of climate change: adapt, mitigate or ignore?*). Instructional documents are also produced by official bodies aiming at giving instructions on different aspects: 'talks' (typically delivered by politicians at the opening and/or closing sessions of Science and Technology-related Conferences instructing on future course of action); administrative documents such as bulletins and guidelines (to universities, professional societies and research bodies commenting on, say, allocation of funds for Science and Technology and instructing on national priorities, etc.). Considering that reports are usually extracted by mass media, discussed in Parliaments and constantly referred to by politicians, researchers, commentators, etc., we selected them according to a 5:1 criterion: reports comprise around 81% of the sub-corpus and instructional documents amount to a figure just below 19% of the sub-corpus) (Table 2).

Sub-corpus	Genre	Number of words	% of sub-corpus	% of corpus
Sub-corpus: 'Research'	Learned Informational:	2,137,171	100	70.5
	Research Articles	1,529,595	71.5	50.4
	Review Articles	607,576	28.5	20.1
	Sub-total	2,137,171	100	70.5
Sub-corpus: 'Official'	Informational (Reportage):	727,876	81.3	24
	Report	727,876	81.3	24
	Instructional:	167,078	18.7	5.5
	Talks	76,745	8.6	2.5
	Bulletin	51,792	5.8	1.7
	Guidelines	29,721	3.3	1.0
	News	8,820	1.0	0.3
	Sub-total	894,954	100	29.5
TOTAL CORPUS		3,032,125		100

Table 2. Documents in the *UVaSTECorpus*

4. Working with the UVaSTECorpus

In terms of methodology, corpus linguists share some general criteria about what corpus-based and corpus-driven analyses should be (Gries 2006): (i) the analysis must be carried out in a corpus of naturally-occurring language which is machine-readable so that the retrieval of the research information is computerized; (ii) the corpus must be balanced and/or representative of the modality/subject matter/variety/genre, etc. the study is aimed at; (iii) the analysis must be systematic and exhaustive; (iv) the analysis must not only account for categorical either-or phenomena, but also use statistical data to explain, say, the middle ground between what is possible and what is not regarding particular grammatical, lexical, lexico-grammatical, discoursal or rhetorical features; (v) both quantitative and qualitative data must be analyzed; (vi) the level of granularity of the study depends on its objective. With these methodological criteria in mind, the *UVaSTECorpus* is used in some ESP teaching activities:

- Students are informed on the importance of the “idiom principle” (Sinclair 1991): many words have a tendency to occur together or in each other vicinity, “unusuality” (Partington 1998) and creativity go hand in hand, and *collocations*, *colligations*, *semantic preferences*, and *semantic prosodies* (Sinclair 1996) are part and parcel of the English language. For example, *bridge gaps*, found in both the *WebCorpus*⁶ and the *UVaSTECorpus*, derives from *bridge the gap*:

“NSF co-sponsors these events and invites researchers to give academic talks on selected topics in order to *bridge gaps* between research and policy.”

(*UVaSTECorpus*: Official)

- Students are explained that words form conceptualizations of reality which define the culture of a discourse community. For example, the noun *bucket* in computing refers to a storage area containing data for an application (example 2), whereas the expression “bucket chemistry”, also found in the corpus, refers to traditional experiments in chemistry (example 3):

“The naïve comparison algorithm is On.29, but this is easily improved by hashing lines into B buckets, and then comparing only lines in the same bucket.”

(*UVaSTECorpus*: Research)

“Branches around the country run *bucket chemistry* competitions for secondary schools. The competition introduces the idea of producing chemicals in large amounts to children.”

(*UVaSTECorpus*: Official)

⁶ In the WebCorp (<http://www.webcorp.org.uk/cgi-bin/webcorp2.nm>), *bridge gaps* occurs 70 times whereas *bridge the gap* occurs 48 times.

- Students are guided to read concordances. In this respect, it is useful for them to know “that they do not have to read the entire lines, but simply scroll the search word, looking, for instance, for adjectives and verbs preceding it” (Gavioli 2005: 74). Following Gavioli (2005: 75-76), we may ask students to carry out the following task in connection with *gender*:
 1. Read the concordance vertically, following the search word column and the word to the left of it. Underline the words to the left of, say, *Ms* in the text.
 2. Do these words give you an idea of what *Ms* is? Discuss your idea with your colleagues and teacher.
 3. The following is the information on *Ms* provided in the Cobuild Dictionary ; Is it helpful in clarifying the meaning of this word?:
 - i. *Ms* is used, especially in written English, before a woman’s name when you are speaking to her or referring to her.
 - ii. If you use *Ms*, you are not specifying if the woman is married or not.
 4. On the basis of the dictionary information, look at the concordance again and identify examples illustrating information i. or ii. next to the lines you selected.
 5. Look at the example extracted from the dictionary: does it illustrate information i. or ii. (or both) clearly?:

“I wrote to Ms Walters and gave my opinion.”
 6. Select some examples in the concordance of *Ms* which you think illustrate information i. or ii. or both clearly. You can widen the text of each line to get more context.
 7. Which information is more frequent in the concordance of *Ms*, i. or ii.? Why, do you think?
- Students are explained the meaning of grammatical terms. Biber et al. (1999), and Biber (2006) use the term *grammatical term* as a general covert term for anything that recurs in texts that can be given a linguistic description, and explain them in terms of linguistic theories. It is claimed that much of the variation among features is highly systematic: speakers of a language make choices in morphology, lexicon, and grammar depending on a number of linguistic and non-linguistic factors. Important components of the situational context include the purpose of communication, the physical mode, the production circumstances, and various demographic characteristics of speaker/writer. One of these is *gender*, which can be analyzed with respect to two main language ‘varieties’: **dialects**, or varieties associated with different groups of speakers (it was a recurrent issue in

the 1980s; see Coates, 1998), and **registers**, or situationally defined varieties, such as the language that is used to discuss specialized fields of knowledge (LSP). This line of research started in the 2000s and is mostly concerned with terminological work (Bowker, 2001; Fuertes-Olivera et al. 2003), cognitive processes (Velasco-Sacristán and Fuertes-Olivera 2006), and grammatical patterns (Fuertes-Olivera 2007a). The grammatical term *lexical gender*, for example, refers to the existence of lexical units such as *mother*, *son*, etc. carrying the semantic property [female] or [male], respectively. In English, such nouns may be described as female-specific or male-specific, in contrast to nouns such as *worker*, which is considered to be gender-indefinite or gender-neutral.

One issue subject to much debate is the use of “generic *man*”, as many English grammars, mainly prescriptive ones, usually state that masculine terms such as “generic *man*” can often be used as duals to refer to both women and men. Advocates of ‘non-sexist English’, however, oppose this belief and regard it as an example of an underlying belief system which results in the invisibility of feminine/female expressions. They claim that uses of ‘generic *man*’ must be avoided, and that the concept of *gender sensitivity*, which indicates that some ways of using a language are functionally, aesthetically or morally preferable to others (Cameron 1996), must be incorporated into LSP research and teaching activities. For example, they defend the use of *person* with general reference (Table 3).

	N	%
<i>Person</i>	93	3.1
<i>Man</i>	26	0.8

Table 3. Occurrences of ‘generic *man*’ and ‘*person*’ in the UVaSTECorpus. % per 100.000 words

Table 3 shows that advocates of non-sexist English are influencing the use of lexical gender in English, as *person* is almost four times more frequent than *man*. Students therefore must be taught accordingly. This means making *gender sensitivity* an overt and deliberate part of ESP teaching, and identifying potential areas of conflict in multicultural settings. As a result, several significant improvements can be made regarding the use of gender neutral language in ESP teaching and learning. For example, it seems reasonable to teach our students that until the 1970s it was usual to distinguish between *gender* (grammatical) and *sex* (extralinguistic). Since then, *sex* refers to biological attributes and *gender* to the social construction of sex, and this usage has been incorporated into linguistics.

5. Conclusion

This article comments on some pedagogical applications of small specialized corpora in ESP teaching activities. After reviewing current work with both in-house and open access corpora, the article illustrates some issues regarding the compilation of the *University of Valladolid Corpus of Written Scientific and Technical English* and

explains its use in teaching and learning activities. It also adds to the amount of knowledge that the use of corpus data is bringing to our professional activity. In particular it explains why ESP students should be taught the niceties of lexical gender as this grammatical category also has social and/or ideological implications.

References

- Biber, D. (2006). *University Language. A Corpus-based Study of Spoken and Written Registers*. Amsterdam / Philadelphia: John Benjamins.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, D., S. Conrad, R. Reppen, P. Byrd, M. Helt, V. Clark, V. Cortes, E. Csomay, & A. Urzua (2004). *Representing Language Use in the University. Analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus*. Princeton, NJ: ETS.
- Bowker, L. (2001). "Terminology and gender sensitivity: a corpus-based study of the LSP of infertility". *Language in Society* 30: 569-610.
- Bowles, H. (2006). Bridging the gap between conversation analysis and ESP – an applied study of the opening sequences of NS and NNS service telephone calls. *English for Specific Purposes* 25, 332-357.
- Burnard, L. (2002). "A retrospective look at the British National Corpus". In B. Kettemann & G. Marko (eds), *Language and Computers: Studies in Practical Linguistics*, 51-70. Amsterdam: Rodopi.
- Cameron, D. (1996). "The language gender interface: challenging co-operation". In V. L. Bergvall, J. M. Bing & A. F. Freed (eds.), *Rethinking Language and Gender Research: Theory and Practice*, 31-53. London: Longman.
- Charles, M. (2006). "Phraseological patterns in reporting clauses used in citation: a corpus-based study of two disciplines." *English for Specific Purposes* 25: 310-331.
- Coates, J. (ed.) (1998). *Language and Gender. A Reader*. Oxford: Blackwell.
- Crystal, D. (1997). *English as a Global Language*. Cambridge: Cambridge University Press.
- Donohue, J. P. (2006). How to support a one-handed economist: The role of modalisation in economic forecasting. *English for Specific Purposes* 25, 200-216.
- Flowerdew, J. and A. Wan (2006): "Genre analysis of tax computation letters: How and why tax accountants write the way they do". *English for Specific Purposes* 25: 133-153.
- Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Connor & T. A. Upton (Eds.), *Discourse in the Professions. Perspectives from Corpus Linguistics*, 11-33. Amsterdam/Philadelphia: John Benjamins
- Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies. *English for Specific Purposes* 24: 321-332.
- Fuertes-Olivera, P. A. (2007a). "A corpus-based view of lexical gender in written business English". *English for Specific Purposes* 26: 219-234.

Fuertes Olivera, P. A. (2007b). "El lenguaje de la ciencia y la tecnología". In Enrique Alcaraz Varó et al. (eds.), *Las lenguas profesionales y académicas*, 205-217. Barcelona: Ariel.

Fuertes Olivera, P. A., M. Velasco Sacristán, M. & E. Samaniego Fernández (2003). "Gender sensitivity in specialized communication: a preliminary corpus-based study of the LSP of Economics. *Ibérica* 6: 65-87.

Gabrielatos, C. & T. McEnery (2005). "Epistemic modality in MA dissertations". In P. A. Fuertes Olivera (ed.), *Lengua y sociedad. Investigaciones recientes en lingüística aplicada*, 311-331. Valladolid: Universidad de Valladolid.

Gavioli, L. (2005). *Exploring Corpora for ESP Learning*. Amsterdam/Philadelphia: John Benjamins.

Gimenez, J. (2006). Embedded business emails: Meeting new demands in international business communication. *English for Specific Purposes* 25: 154-172.

Gries, S. Th. (2006). "Introduction". In S. Th. Gries & A. Stefanowitsch (eds.), *Corpora in Cognitive Linguistics. Corpora-based Approaches to Syntax and Lexis*, pp. 1-17. Berlin & New York: Mouton de Gruyter.

Harwood, N. (2005). 'I hoped to counteract the memory problem, but I made no impact whatsoever': discussing methods in computing science using *I*. *English for Specific Purposes* 24: 243-267.

Hyland, K. & P. Tse (2005). Hooking the reader: a corpus study of evaluative *that* in abstracts. *English for Specific Purposes* 24: 123-139.

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes* 24: 269-292.

Kwan, B. S.C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes* 25: 30-55.

Lee, D. (2001). "Genres, registers, text types, domains and styles: clarifying the concepts and navigating a path through the BNC jungle." *Language Learning and Technology* 5 (3): 37-72.

Lim, J. & M. Hwa (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes* 25: 282-309.

Magnet, A. & D. Carnet (2006). Letters to the editor: Still vigorous after all these years? A presentation of the discursive and linguistic features of the genre. *English for Specific Purposes* 25: 173-199

McCarthy, M. & M. Handford (2004). "Invisible to us": A preliminary corpus-based study of spoken business English. In U. Connor & T. A. Upton (Eds.), *Discourse in the Professions. Perspectives from Corpus Linguistics*, 167-201. Amsterdam/Philadelphia: John Benjamins.

Meyer, C., F. (2002). *English Corpus Linguistics. An Introduction*. Cambridge: Cambridge University Press.

Mudraya, O. (2006). Engineering English: a lexical frequency instructional model. *English for Specific Purposes* 25: 235-256.

Nelson, M. (2006). Semantic associations in Business English: A corpus-based analysis. *English for Specific Purposes* 25: 217-23.

Nesi, H. and S. Gardner (2006) " [Variation in Disciplinary Culture: University Tutors' Views on Assessed Writing Tasks](http://www2.warwick.ac.uk/fac/soc/celte/research/bawe/papers/journals/)". In: Kiely, R., Clibbon, G., Rea-Dickins, P. & Woodfield, H (eds), *Language, Culture and Identity in Applied Linguistics* London: Equinox Publishing. Available: <http://www2.warwick.ac.uk/fac/soc/celte/research/bawe/papers/journals/> (18 January 2007).

- Partington, A. (1998). *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. Amsterdam/Philadelphia: John Benjamins.
- Pecorari, D. (2006). Visible and occluded citation features in postgraduate second-language writing. *English for Specific Purposes* 25: 4-29.
- Rowley-Jolivet, E. & S. Carter-Thomas (2005): "Genre awareness and rhetorical appropriacy: Manipulation of information structure by NS and NNS scientists in the international conference setting". *English for Specific Purposes* 24: 41-64.
- Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes* 24: 141-156.
- Simpson, R. C. (2004). Stylistic features of academic speech: The role of formulaic expressions. In U. Connor & T. A. Upton (Eds.). *Discourse in the Professions. Perspectives from Corpus Linguistics*, 37-64. Amsterdam/Philadelphia: John Benjamins
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. (1996). "The search for units of meaning". *Textus* 9 (1): 75-106.
- Sinclair, J. M. (ed.) (2001). *Collins Cobuild English Dictionary*. London: Harper Collins. 3rd ed.
- Swales, J. M. (1996). "Occluded genres in the academy". In E. Ventola & A. Mauranen (eds), *Academic Writing*, 45-58. Amsterdam/Philadelphia: John Benjamins.
- Swales, J. M. (2006). Corpus linguistics and English for Academic Purposes. In Elisabet Arnó Macià, Antonia Soler Cervera & Carmen Rueda Ramos (Eds.), *Information Technology in Languages for Specific Purposes. Issues and Prospects*, 19-33. New York: Springer
- Velasco-Sacristán, M. & P. A. Fuertes-Olivera (2006). "Towards a critical cognitive-pragmatic approach to gender metaphors in Advertising English". *Journal of Pragmatics* 38: 1982-2002.
- Webber, P. (2005). Interactive features in medical conference monologue. *English for Specific Purposes* 24: 157-181.