



## Contrasting Specific English Corpora: Language Variation

MARÍA LUISA CARRÍO-PASTOR<sup>1</sup>  
*UNIVERSIDAD POLITÉCNICA DE VALENCIA*

### ABSTRACT

The scientific community has traditionally considered technical English as neutral and objective, able to transmit ideas and research in simple sentences and specialized vocabulary. Nevertheless, global communication and intense information delivery have produced a range of different ways of knowledge transmission. Although technical English is considered an objective way to transmit science, writers of academic papers use some words or structures with different frequency in the same genre. As a consequence of this, contrastive studies about the use of second languages have been increasingly attracting scholarly attention. In this research, we evidence that variation in language production is a reality and can be proved contrasting corpora written by native writers of English and by non-native writers of English. The objectives of this paper are first to detect language variation in a technical English corpus; second, to demonstrate that this finding evidences the parts of the sentence that are more sensitive to variation; finally, it also evidences the non-standardisation of technical English. In order to fulfil these objectives, we analysed a corpus of fifty scientific articles written by native speakers of English and fifty scientific articles written by non-native speakers of English. The occurrences were classified and counted in order to detect the most common variations. Further analysis indicated that the variations were caused by mother tongue interference in virtually all cases, although meaning was only very rarely obscured. These findings suggest that the use of certain patterns and expressions originating from L1 interference should be considered as correct as standard English.

**KEYWORDS:** Technical English, academic writing, language variation, standardisation, standard English

### RESUMEN

La comunidad científica considera al inglés técnico como un tipo de lenguaje neutral y objetivo, capaz de transmitir ideas y hallazgos en frases simples y vocabulario reconocido por los especialistas de ese campo. Sin embargo, la comunicación global y el gran tráfico de

---

<sup>1</sup> *Address for correspondence:* María Luisa Carrió Pastor. Dpto. de Lingüística Aplicada. Escuela Técnica Superior de Ingeniería del Diseño. Universidad Politécnica de Valencia. Camino de Vera, s/n, 46022 Valencia, Spain. Telephone: 963877530 (Ext.:75301). Email: lcarrio@idm.upv.es

información han producido una gran variedad de formas de transmitir el conocimiento. Aunque el inglés técnico se considera una forma objetiva de transmitir ciencia, los autores utilizan palabras y estructuras de forma diversa dentro del mismo género. Como consecuencia, los estudios contrastivos del uso de una lengua según el tipo de escritor están proliferando cada vez más. En este estudio, se pone en evidencia que la variación en la producción del lenguaje es un hecho real y que se puede demostrar al comparar corpora escrito por escritores nativos y no nativos de lengua inglesa. Los objetivos de este artículo son, primero, detectar la variación en el lenguaje técnico, segundo, demostrar que este hallazgo muestra las partes de la frase que tienen más tendencia a variar y, finalmente, evidenciar la no estandarización del inglés técnico. Para poder llevar a cabo estos objetivos, analizamos un corpus de cincuenta artículos escritos por autores ingleses y otro de cincuenta artículos escritos en inglés por autores españoles. Los casos encontrados se clasificaron y contaron para detectar las variaciones más comunes y un análisis posterior nos indicó que la influencia de la lengua materna fue la causante de la variación en la mayoría de los casos pero que el significado del texto no había cambiado. Estos hallazgos nos indican que el uso de ciertas estructuras originadas por la interferencia de la lengua materna ha de ser aceptado como correctas.

**PALABRAS CLAVE:** inglés técnico, redacción académica, variación lingüística, estandarización, inglés estándar.

## I. INTRODUCTION

Language users transmit their own perception of reality through language, using it to communicate, but also to persuade, to influence or to manipulate others. Speakers choose rhetorical strategies in discourse depending on the social, economic, political or academic position of their addressees. Furthermore, speakers change vocabulary, expressions and the disposition of sentences or paragraph elements when using a foreign language. Mother tongue influences second language speakers even when language proficiency is not a problem (Freddi, 2005; Hinkel, 2009). Socio-cultural background knowledge can lead to variation in second language use as writers can produce markedly divergent features of text. In this article, variation is referred to as the changes produced in different parts of a written text due to mother tongue influence. Smith and Wilson (1983: 182) mention a similar term that they call *register variation*, but they apply it to the variations produced depending on the context. This is not the concept of variation to be analysed in this article, as we focus on the change produced in the language performance of writers in the same linguistic production and register.

Language changes or variations are caused because communication can be performed in various manners and styles. Writers, depending on their language proficiency amongst other factors, decide to use specific or general terms and complex linguistic strategies or simple texts. Language change can be clearly observed when we contrast texts of the same genre but performed by writers with different social, cultural or economic backgrounds.

Nevertheless, it should be considered that the internal structure of the genre within a particular professional or academic context restricts the form of the linguistic resources and the functional values they assume in discourse. The stereotyped guidelines of a specific genre are used by non-native speakers (NNS) of a language to express in a standardized way, although there are occasions when language variation occurs. As an example, Spanish NNS of English tend to use their mother tongue linguistic models; therefore, they may be prone to copy these structures when they communicate in a second language (Carrió Pastor, 2002; 2005; 2007). Nevertheless, language experts recommend avoiding variation and following standard English rules and structures. In addition, language manuals do not reflect language variation and second language writers are recommended to follow standard rules.

All languages have standardized rules to write genres, as for example, technical or scientific English. This has been the result of a long lasting effort to avoid variation and language change, which has nowadays been demolished by communication technology use. Both the Internet and the World Wide Web have a strong influence on the pace of language change if compared with the last century and are causing language to change quicker than before. As Duszak (1997: 9) points out, “Recent insights into academic writing have shown considerable variation in text characteristics across fields, languages and cultures. [...] Among the most notable differences are field-and culture-bound disparities in global organization schemata of texts.” Text variation should not change their interpretation; otherwise, the main aim of language, i.e. communication, could not be performed.

In particular, technical language has peculiar features associated with technical thinking, such as short sentences, domain specific vocabulary and simple and direct language structures (Dudley-Evans and St. John, 1998; Alcaraz Varó, 2000; Duque García, 2000). Technical writing differs from other genres in being very formal and direct and consequently, rhetorical expressions, metaphors, colloquial expressions, etc. are avoided. As Duszak (1997: 2) notes referring to academic English: “All this contributed to the image of a dehumanised language of science, and likewise to the image of a dehumanised writer [...] uniformity of academic writing styles was taken for granted and was accounted for in terms of objectivised research standards.” In the same way as academic English, technical writing has specific characteristics that differentiate it from other genres, as Alcaraz Varó (2000: 138-9) states. High semantic density, impersonal forms and specialized expressions highlight objectivity, the results of the research and specialized conclusions. These ‘rules’ help second language writers to understand and use specific language appropriately as they are key to native-like fluency; although at the same time these rules constrain natural communication. Eggins & Martin (2000: 336) suggest further characteristics: the use of standard syntax without abbreviations; no reference to the author of the text; the topic is considered the most relevant aspect; frequent use of incrustations; i.e. putting several

subordinated sentences together and long complex noun phrases; reduced and highly specialized vocabulary with action words consisting of nouns and rare adverb use.

This paper focuses on language variation, particularly on the use of corpus analysis to identify the most sensitive parts of the sentence when used in a different way by speakers from diverse linguistic backgrounds.

## 2. CORPUS ANALYSIS

Linguistic research with scientific rigour and objective results should be based on real data and not on intuition. Corpus analysis allows us to investigate language use as it provides real information about the most frequent language structures and rhetoric strategies. Its only concern is the usage patterns of the empirical data and what that reveals to us about language behaviour. Corpus linguistics is a research area which can be described as a study of examples of *real life* language via a corpus, interpreted as a body of text representative of a particular variety of language (McEnery and Wilson, 2001; Mudraya, 2006). Huizhong (1985: 93) justified language corpus use in this way:

Corpus linguistics is able to provide a better model for the description of the English language, which because of the very large amount of data involved cannot be studied directly by human observations. In language study the sampling of linguistic data is indispensable.

A corpus, following Huizhog, should include the highest number of entries in order to obtain reliable results, although lately some researchers point out that corpus size is not so important, depending on the research goals (Krausse, 2005). Corpus researchers agree that there are three basic requirements for obtaining a reliable corpus: first, the samples should be obtained from similar texts; second, the samples should be representative of the whole corpus, and third, the texts should be useful for the research purposes. The nature of a corpus is determined by its purpose and it is vital to interpret what it is meant to represent.

Corpora are used to support theories and ideas, providing examples that support knowledge as Hornero, Luzón and Murillo (2006) point out. Nowadays, more and more researchers have accepted corpus analysis as a way of justifying their research, using percentages and frequencies to analyse language use. The importance of corpora analysis and its application to applied linguistics is beyond doubt, as recent studies can confirm (Holmes, 1994; Stubbs, 1994; Kourilova, 1996; Ceirano and Rodriguez, 1997; Biber, Conrad and Reppen, 1998; de Monnik, 1998; Martí Guinovart, 1999; Oostdijk, 2000; Meyer, 2002; Hornero, Luzón and Murillo, 2006; Lee and Swales, 2006).

One of the most well-known approaches of corpus linguistics is the lexicographic analysis of texts, lead by Sinclair with the COBUILD project in the University of Birmingham (Sinclair, 1991; Carter, 1998: 167; McCarthy, 2001: 125). They have designed

software to classify and search lexical units in order to extract information about language use and English collocations. From this pioneering work, many other tagging projects have been developed. They can be divided in monitor corpus (attempts to be a representative cross-section of the spoken or written language to be studied) and sample corpus (does not pretend to be representative of the whole spoken or written forms of language). The most well-known are ICAME (International Computer Archive of Modern and Medieval English); The Oxford Text Archive; The Cambridge International Corpus; The British National Corpus; Linguistic Resources on the Internet; IT Centers for English Linguistics Corpus; the Corpus of IULA, etc. Nowadays, multimodal corpora are compiled to provide further information to written corpora, as a consequence, context is becoming more and more important to analyse communication patterns. The Internet and the World Wide Web contain multimodal texts that combine photos, videos, text, images, etc. Information is transmitted through diverse modalities; hence multimodal corpora reflect the use of content and context.

Researchers gain access to the most suitable corpus and concordance program in order to determine the usage patterns of the empirical data and what that reveals to us about language behaviour. Two of the most popular computer programs are *MonoConc Pro* (Barlow, 1998) and *WordSmith Tools* (Scott, 1998), which process corpora in order to count occurrences and calculate frequencies.

Through frequency analysis we can observe certain structures in specific genres and this can determine language *rules* adapted to real use. Language registers, the varieties of language which are used for different situations, range from the general to the highly specific. Corpus analysis reveals that language often behaves differently according to their register, each with some unique patterns and rules.

The advantages of the use of corpora for the systematic study of authentic examples of language are evident, but some researchers advise to pay special attention to data interpretation and corpus design, as Carter (1998: 233) comments:

Computer corpora allow access to detailed and quantifiable syntactic, semantic and pragmatic information about the behaviour of lexical items. There is little doubt that such corpora offer invaluable data for vocabulary materials development. But there are obvious dangers in using such data without carefully interpreting it as data and without careful assessment of the kinds of pedagogic criteria which might inform its use.

A well designed corpus can support our generalizations, but if the figures are interpreted erroneously, all our research is not acceptable. If corpus linguistics is viewed as a methodology, it becomes increasingly important the way corpora are created so that those analysing them can be sure that the results of their analysis will be valid.

In this paper, corpus analysis is going to be used to demonstrate language variation in technical English research articles. The objectives of this paper are first to detect language variation in noun phrases and verb phrases produced by Spanish writers using English as their second language (L2) in a technical English corpus; second, to demonstrate that this finding evidences the parts of the sentence that are more sensitive to variation; finally, it also evidences the non-standardisation of technical English.

### 3. THE STUDY

The computer-readable corpus has 100 technical articles and consists of 50 articles written by Spanish NNS of English and 50 articles written by NS of English. The latter corpus was collected according to criteria such as availability and prestige. They were the most well known journals in the given areas of study and were available on line. The length of each article ranged from 1,354 to 2,492 words. Those research articles, whose main authors did not seem to be native speakers of English, as judged by name and institutional affiliation, were disregarded. The former corpus was unpublished articles written in English by researchers whose institutional affiliation was Universidad Politécnica de Valencia. Language assessors, who were Spanish, revised the articles in order to avoid errors or mistakes and to preserve mother tongue influence.

In addition, papers with extensive mathematical procedures and/or statistical treatment were eliminated. Abstracts, titles, footnotes, graphs along with their legends, comments, tables, acknowledgements and bibliographic references were excluded from the corpus.

Once the research corpus was compiled, all the variations were located and counted, and percentages and frequencies were calculated in the corpus. The corpus was processed using the Wordsmith Tools suite of programmes (Scott, 1998), which enable the user to identify and compute recurrent patterns in a bunch of texts. The implication is that in looking for recurring patterns, notions such as frequency and probability tell us that if something happens frequently, then it is significant because of regularity, and therefore future behaviour can be predicted.

Afterwards, we classified them grouping the findings into samples of NWs' and NNWs' speech. We divided our findings into noun phrases (NP) and verbal combinations which were classified in tables in order to compare occurrences and frequency. The results were analysed and we calculated the p-value in order to carry out a quantitative analysis, which relies on counts of specific linguistic features as they occur in text. Statistically significant results are those with a p-value below 0.05, this means that there is a 5 per cent or lower probability that the result was gained by chance. After classifying, counting

occurrences and calculating percentages, we contrasted our results to find the corresponding conclusions to our research.

#### 4. RESULTS AND DISCUSSION

The statistical data of the corpora gathered can be observed in Table 1:

SENTENCE DATA	OCCURRENCES	OCCURRENCES
	NNS (%)	NS (%)
Total words	184,357 (47.11%)	206,907 (52.89%)
Word list	10,590 (45.43%)	12,716 (54.57%)
Sentence number	9,017 (50.00%)	9,017 (50.00%)
Word average	20.44 (46.11%)	22.94 (53.89%)
Paragraph number	1,145 (55.51%)	916 (44.49%)
Paragraph word number	161.29 (41.58%)	225.88 (58.12%)

Table 1. Data from technical English articles.

As shown in Table 1, we analysed an equal number of sentences in the corpus of NS of English and in the corpus of Spanish NNS of English. As the research presented in this article focuses on noun phrases and verb phrases, it was significant to choose the same number of units to be contrasted.

Noun phrases are important in technical English as complex noun phrases are commonly used to transmit information in a compact way. This cluster is not used in Spanish as the nouns are linked by prepositions, making more explicit element relationship. NS occurrences and NNS occurrences were contrasted in order to observe if complex combinations were used in the same way by writers with different linguistic background. We also considered the occurrences of noun phrases followed by the preposition *of* and the occurrences of the article to obtain results that could show mother tongue influence. The results found and the p-values are given in Table 2:

NOUN PHRASE COMBINATIONS	OCCURRENCES NNS (%)	OCCURRENCES NS (%)	P-value
N3	679 (53.61%)	590 (46.49%)	P = 0.14
A+ N2	906 (49.81%)	913 (50.19%)	P = 0.04
A2+ N	313 (46.58%)	359 (53.42%)	P = 0.00
N4	52 (63.41%)	30 (36.59%)	P = 0.03

A+ N3	126 (60.29%)	83 (39.71%)	P = 0.01
A2+ N2	53 (45.69%)	63 (54.31%)	P = 0.19
A3+ N	8 (44.44%)	10 (55.56%)	P = 0.53
N5	3 (60.00%)	2 (40.00%)	P = 0.70
A+ N4	12 (80.00%)	3 (20.00%)	P = 0.02
A2+ N3	7 (50.00%)	7 (50.00%)	P = 0.89
A3+ N2	1 (33.33%)	2 (66.67%)	P = 0.52
A4+ N	0 (0.00%)	1 (100.00%)	-
N6	0	0	-
<i>Total NP</i>	2839 (51.69%)	2653 (48.31%)	-
N+ 'OF'	4341 (46.41%)	5013 (53.59%)	-
Articles Total	21626 (48.33%)	23113 (51.67%)	-
A	3965 (46.23%)	4611 (53.77%)	P = 0.00
AN	841 (48.50%)	893 (51.50%)	P = 0.89
THE	16820(48.85%)	17609(51.15%)	P = 0.00

Table 2. Noun phrase variations.

The occurrences related to the use of verb tenses, modal verbs and passive voice were also considered relevant in this study as potential indicators of language variation. Table 3 shows the general category results found in the corpora analysed and their p-value:

VERB PHRASES	OCCURRENCES NNS (%)	OCCURRENC ES NS (%)	P-value
Present simple	3034 (47.71%)	3324 (52.29%)	P = 0.01
Present continuous	34 (58.62%)	24 (41.38%)	P = 0.14
Past simple	5145 (48.98%)	5359 (51.02%)	P = 0.93
Past continuous	5 (35.71%)	9 (64.29%)	P = 0.32
Present perfect	40 (42.55%)	54 (57.45%)	P = 0.21
Past perfect	1 (11.11%)	8 (88.89%)	P = 0.02
Future (will)	424 (60.65%)	275 (39.35%)	P = 0.00
Total verb tenses	8683 (48.95%)	9053 (52.83%)	-
Modal verbs	1769 (54.16%)	1497 (45.84%)	-
Passive voice	248 (43.43%)	323 (56.57%)	-

Table 3. Verb phrase variation.



It was considered relevant to offer more detailed results in the sub-category of modal verbs, as modality is transmitted in a different way in Spanish and in English. We excluded modal verbs with less than 10 occurrences and those which were not used as modal verbs, i.e. *will*. The occurrences and p-value can be seen in Table 4:

MODAL VERBS	OCCURRENCES NNS (%)	OCCURRENCES NS (%)	P-value
CAN/ BE ABLE	877 (59.82%) 78 (76.47%)	589 (40.18%) 24 (23.53%)	P = 0.00 P = 0.00
COULD	166 (48.82%)	174 (51.18%)	P = 0.03
MAY	181 (39.69%)	275 (60.31%)	P = 0.00
MIGHT	13 (24.07%)	41 (75.93%)	P = 0.00
MUST	213 (62.64%)	127 (37.36%)	P = 0.00
NEED	90 (38.96%)	141 (61.04%)	P = 0.00
SHOULD	151 (54.51%)	126 (45.49%)	P = 0.90
Total	1769 (54.16%)	1497 (45.84%)	-

Table 4. Modal verb variation.

The results of corpus analysis clearly show that there are certain parts of the sentence more sensitive than others to variation when used by native speakers of English and by non native speakers of English. In the first place, the results of the analysis of complex noun phrases demonstrated that the longer the complex noun phrases, the more variation we could find. Native speakers of English used less complex noun phrases formed by four or five elements than Spanish non native speakers of English. These results were quite surprising since complex noun phrases are more difficult for NNS as there is not a similar structure in Spanish. An overuse of recommended structures in technical English could be the cause of the results found in our corpora. Furthermore, the use of noun phrases followed by the preposition *of* was more common among NS of English. The fact that NNS use more complex noun phrases and less noun phrases followed by the preposition 'of' than NS evidences variation in the use of complex noun phrases in technical English. In the second place, it was also noticed that NS also used more articles than NNS, a result that could be associated to mother tongue influence.

With regard to verb tenses, most of them were used with the same frequency in both corpora. Nevertheless, special attention should be paid to the results of *will* used to express future tense. It was more used by NNS of English than by NS of English, as a result of the different way to express future in English and in Spanish. In Spanish, the simple future form expresses certainty; however, the future tense formed with *will* conveys uncertainty in

English. The overuse of the future tense formed with *will* by Spanish writers is caused by mother tongue influence; although this form means certainty in Spanish when translated literally, it is not frequently used in technical English. Turning now to passive voice, the variation in its use is also due to the influence of Spanish. In the corpus analysed, English writers used passive voice more than Spanish writers. A possible cause of this could be the scarce use of passive voice in Spanish as an impersonal expression, as other rhetorical strategies are preferred. On the contrary, English uses the passive voice in scientific English to express objective results and conclusions.

Furthermore, when observing the occurrences of modal verbs found in the corpora, we can notice that Spanish NNS of English used *can*, *be able* and *must* more frequently than English speakers. On the contrary, the latter used more frequently *may* and *might* than the former. Modal verbs related to certainty or uncertainty are used in a different way by both groups of writers. Spanish writers prefer to express findings assertively, whereas English writers prefer to be extra cautious and hesitant to assure the other's positive response. *May* and *might* are predominantly employed as markers of logical possibility and doubt usually expected in formal academic prose (Biber et al., 2002), whereas *can* and *must* are used as markers of obligation and certainty.

In sum, the results obtained confirmed that there are several parts of the sentence more sensitive to variation. Mother tongue influence is the main cause of the language variation found in the corpora analysed. It has been proved that writers transmit their own language models when writing in a second language; however it should also be considered the consequences of language variation on discourse understanding.

## 5. CONCLUSION

Technical writing in English recommends language economy and a direct way to transmit ideas. These standard characteristics should be followed by non native writers of English if they wish to publish scientific articles in international journals, since the refereeing committees are mainly British or American. The articles are revised and corrected by the referees following the standard guidelines of American or British English. As a consequence, most of the articles published in international journals are written in standard English.

Non native writers of English suffer the serious difficulties to publish in an international journal as they have to revise their articles several times or even look for a professional language assessor. These adaptations to the international style of journals leaded by American and British committees reinforce the English language standardization in technical writing. As a consequence, these editorial guidelines are blocking language variation and the emergence of language evolution in technical English. Language change

has been potentiated with the use of Internet to communicate, whereas technical English publications follow the same 'rules' established in the last century.

Summing up, the language variation detected after the analysis of the corpora does not interfere in the understanding of the texts. Furthermore, we wish to vindicate the acceptance of language variation as a consequence of the use of English as a lingua franca. Variation should be considered to enrich the language used by international writers and thus, it should be accepted once determined its neutral interference in communication. Cultural differences are not relevant to interpret a specialized text, as international language should integrate different ways to express the same concepts.

Further studies should be developed in order to incorporate variations to English as a lingua franca. These variations should be identified in a multimodal international corpus showing the changes in language produced by second language speakers. This corpus would allow linguists to incorporate mother tongue variations when updating manuals of the English language.

## 6. REFERENCES

- Alcaraz Varó, E. (2000). *El inglés profesional y académico*. Madrid: Alianza.
- Barlow, M. (1998). "MonoConc concordance programs for text analysis". [Http://www.ruf.rice.edu/barlow/mono.html](http://www.ruf.rice.edu/barlow/mono.html) (03-06-08).
- Biber, D.; Conrad, S. and Reppen, R. (1998). *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber et al. (2002). *Longman student grammar of spoken and written English*. Harlow: Longman.
- Carrió Pastor, M. L. (2002). "Use of phrasal verbs by native and non-native writers of technical articles." *Linguistic Insights*, 2: 233-246.
- Carrió Pastor, M. L. (2005). *Contrastive analysis of scientific-technical discourse: common writing errors and variations in the use of English as a non-native language*. Ann Arbor: UMI.
- Carrió Pastor, M. L. (2007). "La variación en el uso de las expresiones de certeza en los textos científicos". *BELLS*, 16.
- Carter, R. (1998). *Linguistic Choice across Genres: Variation in Spoken and Written English*. Amsterdam: John Benjamins.
- Ceirano, V. & Rodriguez P. G. (1997). "Análisis del discurso asistido por computadora. Nuestra experiencia con el NUD IST". [Http://www.geocities.com/Athens/Forum/5917/analista.html](http://www.geocities.com/Athens/Forum/5917/analista.html) (06-04-08)
- Dudley-Evans, A. and St. John, M. J. (1998). *Developments in English for Specific Purposes: A Multidisciplinary Approach*. Cambridge: Cambridge University Press.

- Duque García, M. M. (2000). *Manual de estilo. El arte de escribir en inglés científico-técnico*. Madrid: Paraninfo.
- Duszak, A. (1997). "Cross-cultural academic communication" in A. Duszak [ed.] *Culture and styles of academic writing*. New York: Mouton de Gruyter: 11- 40.
- Eggins, S, & Martin, J. R. (2000). "Géneros y registros del discurso" in T. A. Van Dijk [ed.] *El discurso como estructura y proceso*. Barcelona: Gedisa Editorial: 335- 371.
- Freddi, M. (2005). "Arguing linguistics: corpus investigation of one functional variety of academic discourse." *Journal of English for Academic Purposes*, 4: 5-26.
- Hinkel, E. (2009). "The effects of essay topics on modal verb uses in L1 and L2 academic writing." *Journal of Pragmatics*, 41: 667-683.
- Holmes, J. (1994). "Inferring language change from computer corpora: Some methodological problems". *ICAME Journal. Computers in English Linguistics*, 18: 27- 40.
- Hornero, A. M.; Luzón, M. J. and Murillo, S. [Eds.] (2006). *Corpus Linguistics. Applications for the Study of English*. Bern: Peter Lang.
- Huizhong, Y. (1985). "The use of computers in English teaching and research in China" in R. Quirk & H. G. Widdowson [eds.] *English in the World. Teaching and Learning the Language and Literatures*. Cambridge: Cambridge University Press: 86- 104.
- Kourilova, M. (1996). "Interactive functions of language in peer reviews of medical papers written by non-native users of English". *UNESCO-ALSED-LSP Newsletter*, 19- 1: 4- 21.
- Krausse, S. (2005), "Testing the validity of small corpus information." *ESP World*, 1-9.
- Lee, D. and Swales, J. (2006). "A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora." *English for Specific Purposes*, 25: 56-75.
- Leech, G. (1997). "A brief user's guide to the grammatical tagging of the British National Corpus". [Http://www.hcu.ox.ac.uk/BCN/what/gramtag.html](http://www.hcu.ox.ac.uk/BCN/what/gramtag.html) (03-06-02)
- Martí Guinovart, M. A. (1999). "Panorama de la lingüística computacional en Europa". *Revista Española de Lingüística Aplicada. Volumen Monográfico: Panorama de la Investigación en Lingüística Aplicada*, 11- 24.
- Meyer, C. F. (2002). *English corpus linguistics: an introduction*. Cambridge: Cambridge University Press.
- McCarthy, M. (2001). *Issues in Applied Linguistics*. Cambridge: Cambridge University Press.
- McEnery, T. and Wilson, A. (2001). *Corpus linguistics: an introduction*. Edinburgh: Edinburgh university Press.
- Mönnink, I. de (1997). "Using corpus and experimental data: a multimethod approach". [Http://iris1.let.kun.nl/literature/demonnink.1997.2.html](http://iris1.let.kun.nl/literature/demonnink.1997.2.html) (13-03-98, 12:26)
- Morel, J. (1997). *El Corpus de l'IULA: etiquetaris*. Barcelona: Universitat Pompeu Fabra.
- Mudraya, O. (2006). "Engineering English: a lexical frequency instructional model." *English for Specific Purposes*, 25: 235-256.

- Oostdijk, N. (2000). "Corpus-based English linguistics at a cross-roads". *English Studies*, 81- 2: 127- 141.
- Scott, M. (1998). *WordSmith Tools 3.0*. Oxford: Oxford University Press. <http://www.liv.ac.uk/ms2928/wordsmith.htm>.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Smith, N. & Wilson, D. (1983). *La lingüística moderna*. Barcelona: Editorial Anagrama.
- Stubbs, M. (1995). *Texts and Corpus Analysis: Computer- Assisted Studies of Language and Culture*. Cambridge: Blackwell.