



Recent and Applied Corpus-based Studies

Corpus-based studies are gaining momentum in current linguistic studies. Just a quarter of a century ago it was rather obsolete to extract conclusions on data previously arranged, systematized and analyzed. In the first part of the 21st century, this is precisely the most frequent method of analysis in both historical and synchronic linguistics. There seems to be sound reasons to proceed in this way: since samples of language in digital format are nowadays easily accessible and computers allow for a quick processing of huge amounts of data, the change of the research paradigm is shifting from theoretically based constructs to data based ones. Moreover, theoretical constructs have proved to be rather precarious, if we consider their instability along the history of ideas, theories or theoretical proposals advanced once and again by different authors. Language is after all something we can easily ‘grasp’ in so far as it is a formal system subject of quantification in some way in the field of lexis, morphology, syntax, and the supporting sound or graphical systems. From a scientific point of view, the analysis of language or linguistic items should not be subordinated to prefabricated theories, or representative of real language use. This is precisely what corpora claim and what corpora facilitate in linguistic studies.

We include here papers related to four areas of linguistic research: (i) the complex area of word sense analysis and computational management; (ii) vocabulary as a key issue in language teaching materials and language acquisition and learning; (iii) the function and meaning of various linguistic elements or pragmatic resources as they appear in discourse; and (iv) a more incipient area of research by means of computational tools, literary texts, traditionally reluctant to objective and formal analysis, let alone the kind of analysis based on the results extracted through computational tools.

In **Section I**, **Stuart and Botella** design a methodology for ‘the reticular representation of knowledge of a specific discourse community’. With the co-occurrence matrices resulting from their analysis, they generate semantic networks of subject and discipline knowledge and they conclude that this methodology ‘may be viable to extract and to represent the intellectual capital of an academic institution’. **Almela** analyses the issue of WSD. He describes the complexity underlying lexical meaning and the difficulties this fact poses for WSD. After analysing some samples of WSD, he advances some ideas on how to solve the problem of fine-grained WSD departing from the concept of ‘lexical constellation’ (as described by Cantos & Sánchez, 2001).

Section II includes five contributions centred on lexis and vocabulary. The articles by Criado, Alcaraz and Olmos analyse and discuss the question of vocabulary in ELT from different perspectives and points of view. **Criado** analyses the kind and amount of words present in a sample textbook and compares the results against the expectations based on frequency of usage. **Alcaraz** claims the need to take functionality as an additional criterion, together with frequency of usage, for deciding on what to include in syllabi for ELT. **Olmos’** perspective centres on what students of a particular educational institution (upper secondary educational level) in Spain have really learned against the expectations derived from the vocabulary ranges defined by Nation.

In the three additional contributions within this section, **Rea** offers a detailed analysis of a specific term in the field of telecommunication engineering English. She aims at illustrating a model of analysis in order to determine the specialized character of a lexical unit (*Wireless*). The information based on the variables of frequency, distribution and keyness, are combined with the data extracted from the exploration of the surrounding co-text, in order to describe the syntagmatic relations established. **Oncins** studies the case of a specific borrowing through the analysis of the English word ‘dramatic/dramatically’ against the Spanish ‘dramático/dramáticamente’ in order to illustrate how new meanings enter other languages with the addition of new senses to the same form. Data extracted from large current corpora of English and Spanish support his analysis.

Section III includes six articles on the function and meaning of various linguistic elements or pragmatic resources as they appear in discourse. **Orts and Almela** describe the abundant loans of terms from the field of economy entering Spanish discourse at present. The need for precision and accurateness –they claim– is at the basis of this kind of borrowing. They base the analysis on a corpus of economic news items. **Lavid, Arús and Moratón** advocate a new methodology for contrastive studies, based on bidirectional translations and the use of comparable texts. Samples extracted from corpora serve the authors for supporting their

proposal. **Gil-Salom and Soler-Monreal** analyse the presence of face-redressive politeness strategies in engineering research articles. Data are extracted from a corpus with discussion sections in the fields of computing, telecommunications, nanotechnology and robotics. **Carbonell-Olivares** undertakes a corpus-based analysis of the meaning and function of *although*. She contributes empirical evidence for certain claims made in the literature, and complements such descriptions with some novel findings. **Salazar and Verdaguer** presents a corpus-based analysis of a selection of polysemous lexical verbs used to express modality in student argumentative writing. A careful study of concordances allows the authors to determine the use of the verbs analysed in three corpora. The results reveal that the non-native writers have a limited grasp of the full range of meanings of lexical verbs such as *feel*. **Carrió** aims at detecting language variation in a technical English corpus and at demonstrating that some parts-of-speech are more sensible to variation. Her findings suggest that the variations are 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.

Finally, **Section IV** includes an article by **Keshabyan** aiming at carrying out a structural and lexical analysis of two contrasting plays –Shakespeare’s *Hamlet* and Sumarokov’s *Gamlet*– in a specific linguistic domain. In this contribution, the author attempts to gain some insight into two essential content words: *vengeance* and *hono(u)r*, their derivatives and related words, through quantitative analysis of these words and qualitative analysis of their collocates and concordances. The author analyses and compares the ways Shakespeare and Sumarokov perceive the concepts of *vengeance* and *hono(u)r*.