Lexis
Journal in English Lexicology

4 | 2010
Corpus Linguistics and the Lexicon

Introduction

Henri Béjoint, Diana Lewis and François Maniez

Electronic version
URL: http://journals.openedition.org/lexis/551
DOI: 10.4000/lexis.551
ISSN: 1951-6215

Publisher
Université Jean Moulin - Lyon 3

Electronic reference
Henri Béjoint, Diana Lewis and François Maniez, « Introduction », *Lexis* [Online], 4 | 2010, Online since 14 April 2010, connection on 24 September 2020. URL : http://journals.openedition.org/lexis/551 ;
DOI : https://doi.org/10.4000/lexis.551

Introduction

Over the past two decades, monolingual and bilingual lexicography, as well as lexical semantics, have been transformed by the use of electronic corpora, and particularly by the study of co-occurrence phenomena. Large electronic corpora have opened up many new avenues of research, and this fourth issue of *Lexis* brings together papers representing a range of approaches to corpus-based lexical studies.

The first three articles are concerned with the exploitation of corpora for pattern extraction of various kinds, involving large sets of lexemes. A major problem in word-sense disambiguation, in computational linguistics, is its dependence on semantic typing. The different sets of arguments that correspond to the different senses of a verb, for instance, do not constitute neat semantic types. This is the starting-point for Jezek and Hanks' article on "What lexical sets tell us about conceptual categories". **Elisabetta Jezek** and **Patrick Hanks** perform "corpus pattern analysis" to build statistically-based, syntagmatic lexical sets; this allows a prototype-style approach to type membership. Their paper illustrates the procedure with a number of English verbs. "Using parallel text for the extraction of German multiword expressions", by **Fabienne Fritzinger**, proposes a procedure for the automatic extraction of German multiword expressions of the type 'verb + PP' (for example *ins Leben rufen*) from parallel texts. The method exploits the fact that such opaque combinations are translated as wholes. The results show that the procedure works reasonably well in terms of precision, but not so well in terms of recall. "'Inheritance' relations in corpora: their extraction and implications for dictionaries", by **Ekaterina Lapshinova-Koltunski**, describes a semi-automatic method for the classification of German predicates extracted from corpora. She uses the subcategorisation properties of morphologically related predicates (verbs, deverbal nouns and multi-word expressions containing a nominalisation). The classification cannot be fully automatic because inheritance is not totally predictable, but the method saves time and effort in lexicon and dictionary creation.

There follow two articles on the use of corpora in the analysis of the lexicons of specialist genres. **Viviana Gaballo** studies the specialized field of stone processing and quarrying with a view to provide a lexicographic reference work for this field, for which very few resources are available. Using English and Italian comparable corpora containing a variety of texts from brochures to technical specifications, she shows how a number of term candidates can be identified and analysed in their contexts to formulate hypotheses for translation equivalents. **Amélie Josselin** studies the description of terms in general language dictionaries in the field of volcanology. Following a general description of the use of corpora in lexicography, she draws examples taken from a corpus of scientific reviews for the general public and studies the treatment of French and English terms that are related to this relatively unexplored field. She argues convincingly for the use of non-specialized corpora for improving coverage of specialized fields in the macrostructure of monolingual as well as bilingual dictionaries.

The final paper in the issue exemplifies the kind of fine-grained lexical semantic analysis that has been facilitated by the advent of electronic corpora. In her quantitative lexical semantic study of the English verb *grasp*, **Marie Nordlund** demonstrates how the type of frequency and distribution information provided by corpus analysis can complement the definitions traditionally found in dictionaries. *Grasp* is a fine illustration of a polysemous verb, resulting from concrete meanings based in the physical world having come to express more abstract concepts. In her register-diverse BNC data, Nordlund finds that use of non-

physical *grasp* is more frequent than use of physical *grasp*, and that the use of *grasp* in the sense of "mental grasping" is often associated with negation or uncertainty.

The contributions reflect the wide and growing range of applications of corpora to lexical studies. The increasing availability via corpus analysis of the kinds of information on lexical patterning referred to in these papers allows new insights into the interplay between lexis and syntagmatic relations, and is changing the way we conceive of lexical meaning itself.

Henri Béjoint
Diana Lewis
François Maniez