MORPHOLOGY MEETS COMPUTATIONAL LINGUISTICS

NICOLA GRANDI FABIO MONTERMINI FABIO TAMBURINI

The studies collected in this special issue of *Lingue e Linguaggio* aim to sketch a picture of the interfaces between morphology, computational linguistics, and corpus linguistics. Some of these studies are the outcome of the Galileo project *Lessico e regole di formazione di parola: un approccio tipologico e computazionale*, supported by the Università Italo-Francese and the French Ministry of Foreign Affairs, and were presented at the workshop *Morphology Meets Computational Linguistics*, held at the University of Bologna on October 7-8, 2010. The papers published in this issue have been selected through a double-blind peer review process.

A general trend of present-day linguistic research is the emergence of interfaces between different domains and approaches. In particular, morphology has proven to have much to gain from the interaction with Natural Language Processing in terms of tools for the automatic extraction of data and the implementation of theoretical hypotheses. Conversely, NLP can benefit from an observation of data that is less impressionistic and based on a solid theoretical background.

Since it is primarily linked with the lexicon, morphology is one of the domains of linguistic analysis that has benefited the most from technological innovation in the last decades. Automatized lexicographic resources first, and, more recently, the huge amount of data the Net has given access to has had a strong influence on research both on inflectional and derivational morphology. In some cases, this has not only led to more accurate analyses of data, but to real changes in theoretical perspectives. In this respect, morphologists cannot avoid having systematic interaction with specialists of NLP in order to develop tools for the construction, the annotation, and the analysis of large-scale corpora. All the contributions contained in this issue exemplify at least one of these interfaces, and each of the papers presents an original look on the interplay of morphological issues and computational techniques.

In the first contribution, Basilio Calderone and Chiara Celata propose a computational simulation of an emergent model of paradigmatic organization based on French verbal data. Their analysis leads to a reassessment of certain properties of verb inflection, such as the repartition

of lexemes into inflectional classes and the repartition between regular and irregular forms.

In the same vein, and using similar techniques, Marcello Ferro, Claudia Marzi and Vito Pirrelli present a computational model supporting the idea that stored information and on-line computation in morphological competence are the outcomes of a unique set of principles of memory self-organization.

The main topic of the contribution by Nicola Grandi, Fabio Montermini and Fabio Tamburini is the annotation of large corpora with morphological information. In particular, the authors propose to encode morphological information into a Derivational Graph in order to overcome the traditional linear representation of morphologically complex words and successfully handle apparently "odd" data extracted by corpora.

Nabil Hathout's contribution presents an automatic tool (*Morphonette*) designed to group French complex words into paradigmatic sets (families and series) based on a large lexicon of written forms. This resource is grounded on a measure of morphological proximity and on formal analogy. The paradigmatic organization of the lexicon is represented by means of "filaments" which account for the different morphological properties of individual words.

The paper by Aurélie Guerrero presents an example of how the observation of large corpora of data can have an impact on theoretical modelization. In particular, the author proposes an analysis of Catalan verb inflection within a Word-and-Paradigm model in which morphological relations emerge as generalizations on the existing lexicon.

The paper by Malvina Nissim and Andrea Zaninello is based on a large-scale corpus analysis as well. In particular, the authors consider Italian multiword expressions and propose measures for the degree of internal fixedness of this type of complex unit (a kind of linguistic object which has drawn the attention of morphologists only in recent years).

A similar phenomenon is considered in the last paper of this issue. In it, Valentina Efrati and Francesca Masini present Cop-It, which aims to create an online resource endowed with a database of Italian word combinations (i.e. lexical expressions of different nature that are characterized by varying degrees of fixity and familiarity) extracted from corpora and an interface for simple and advanced searches with a dedicated query language.

The authors wish to thank all the anonymous reviewers for their valuable help in the review process.

Bologna and Toulouse, October 2011