

Preface

This is the sixth workshop on computational phonology (and now morphology!) held by SIGPHON, the ACL Special Interest Group in Computational Phonology. We are happy this year to be holding the workshop with the cooperation of SIGNLL, the ACL Special Interest Group in Natural Language Learning.

As the name of our workshop suggests, the topic this year includes not only phonology, but also morphology—and in fact all but one of the papers are about morphology. The boundary between these two disciplines has always been fuzzy, and indeed many of the interesting problems in phonology would not arise were it not for morphology.

The traditional way to build a computationally interpretable morphological grammar and lexicon (if one can speak of a tradition in such a young discipline) is to do so by hand. While this is still the dominant way to build morphological grammars, there has been an increasing interest in the last few years in machine learning techniques.

The papers in this workshop explore some of the ways machine learning could be brought to bear on the task of building grammars for morphology and phonology. Given the traditional definition of a morpheme as the smallest unit with meaning, linguists may be surprised that none of the morphology learning methods described here uses bilingual text to infer semantics. There are several reasons for that, but some of the approaches described here do use word-level co-occurrence as a substitute for richer forms of semantic representation.

It is my pleasure to thank our invited speaker, David Yarowsky (of Johns Hopkins University), for his presentation, “Bootstrapping morphological analyzers from monolingual and bilingual data—building multimodal bridges”. In addition, I thank Mikhail Belkin and John Goldsmith (both of The University of Chicago) and Adam Albright and Bruce Hayes (of UCLA) for accepting our invitation to present their work. Without ACL’s generous support, it would not have been possible to hold this workshop.

Finally, I want to thank the other members of the program committee for their advice, and for reviewing the papers on a tight schedule: Antal van den Bosch (Tilburg University, and our representative from SIGNLL), Jason Eisner (Johns Hopkins University), Steven Bird (University of Pennsylvania), Lauri Karttunen (Parc Inc.), and John A. Goldsmith (University of Chicago).

Mike Maxwell
Linguistic Data Consortium, University of Pennsylvania
Chair, Program Committee

Program Committee

Mike Maxwell (chair) Linguistic Data Consortium, University of Pennsylvania

Steven Bird, University of Pennsylvania

Antal van den Bosch, Tilburg University

Jason Eisner, Johns Hopkins University

John A. Goldsmith, University of Chicago

Lauri Karttunen, Parc Inc.