Finite–state methods and mathematics of language. Introduction to the special issue

Marco Kuhlmann[1] *and Christian Wurm*[2]
[1] Department of Computer and Information Science
Linköping University, Sweden
[2] Department of Computational Linguistics
University of Düsseldorf, Germany

There is a long and fertile interaction between research on finite-state methods and the mathematics of language: many central results in mathematical linguistics are based on finite-state models such as automata and grammars, and mathematical linguistics in turn has laid the foundations for the application of finite-state methods in natural language processing (NLP). One important outcome of the cross-semination between the two fields is the characterisation of adequate classes of string languages and tree languages for linguistic modelling.

Our intention with this special issue is to highlight current work in the intersection between mathematics of language and finite-state methods, as presented at two premier conferences in the respective fields: the 12th International Conference on Finite-State Methods and Natural Language Processing (FSMNLP), which was held 22–24 June 2015 in Düsseldorf, Germany; and the 14th Meeting on Mathematics of Language (MOL), which was held 25–26 July 2015 in Chicago, USA. To this end we invited the authors of the two conferences to submit revised and extended versions of their contributions, which were then subjected to an entirely new peer-review process – something that would have been impossible without the dedication and thorough work of our reviewers, to whom we owe our sincere gratitude.

At the end of the peer-review process, we selected four submissions for publication in this special issue:

"Chomsky–Schützenberger parsing for weighted multiple context-free languages" by Tobias Denkinger generalises the well-known char-

acterisation of context-free languages as the homomorphic image of the intersection of a Dyck language and a regular language to an expressive class of weighted languages, and then uses this characterisation to derive a parsing algorithm.

"Relative clauses as a benchmark for Minimalist parsing" by Thomas Graf, James Monette, and Chong Zhang presents a careful and comprehensive evaluation of a large number of complexity metrics that have been proposed to relate parsing difficulty to memory usage. The results show that only a handful of these metrics can explain observed contrasts in human sentence processing.

"Rewrite rule grammars with multitape automata" by Mans Hulden addresses the following problem: relation composition is one of the most frequently used methods in finite-state approaches; in particular, it allows to construct complex transformations out of simpler ones via intermediate steps, which then are discarded. This discarding is not desirable in some applications, such as the reconstruction of old languages. However, if one does not discard intermediate steps, then relations become more than binary, which is a problem for existing program libraries. The article addresses this problem both from a theoretical and practical point of view, by encoding arbitrary tuples as simple strings, hence relations as languages.

"A probabilistic model of Ancient Egyptian writing" by Mark-Jan Nederhof and Fahrurrozi Rahman provides a formal model for the transliteration of hieroglyphic writing. Ancient Egyptian writing is particularly complex, because the same hieroglyph can have many different functions: it can have (among other) a semantic content, a phonological content, or just be used to specify the semantic or phonological content of some other hieroglyph (both redundantly or not). The authors approach this extremely complex system by introducing "sign functions", which go beyond the power of finite-state machines and lay the foundation for "machine transliteration" of Ancient Egyptian writing.