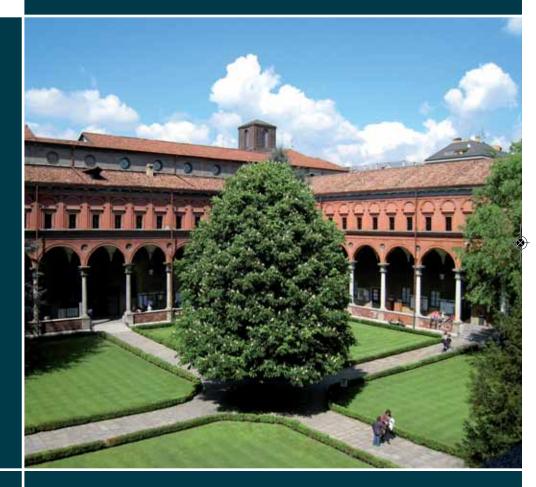# Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories (TLT8)

4-5 December 2009, Milan, Italy

Since its first edition in 2002, the International Workshop on Treebanks and Linguistic Theories (TLT) has provided a yearly forum for discussion of methods and tools for the design, creation and exploitation of treebanks and the linguistic theories acting as their background. Treebanks are syntactically annotated corpora that have become essential resources for data-driven approaches to both natural language processing and linguistic research. Indeed, while treebank data are frequently exploited for tasks in computational linguistics such as grammar induction and the training of NLP tools, in linguistic research they can also be used in order to refine and improve pre-corpus linguistic theories. Furthermore, large-scale data annotation allows for empirically evaluating the accuracy of a grammar and revising it on the basis of evidence.

Recently, many treebank projects for less-resourced languages have begun. The increasing spread of such treebanks benefits from the exploitation of tools and methods developed over the years for many similar projects for other languages. The language-independent status of these methods and tools has indeed allowed their re-usability (or easy adaptation) to many different languages. This has eased and sped up the process of creation and dissemination of treebanks for less-resourced languages.

These proceedings consist of 19 selected papers on various issues related to both long-standing and new treebanks.

Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories (TLT8), Milan, Italy

edited by

Marco Passarotti
Adam Przepiórkowski
Savina Raynaud
Frank Van Eynde