

Diachronic Treebanks

Convenors:

Hanne Eckhoff (University of Tromsø, Norway),

Silvia Luraghi (Università di Pavia, Italy),

Marco Passarotti (Università Cattolica del Sacro Cuore, Milan, Italy)

The workshop aims at bringing together researchers interested in historical linguistics, who combine a solid linguistic background with an interest for the exploitation of electronic resources, and in particular of syntactically parsed corpora, in research on language change. We welcome proposals addressing diachronic issues under any type of approach and methodology, provided that they highlight the contribution of empirical evidence retrieved from treebanks in achieving meaningful results.

Over the last two decades, treebanks have become an increasingly useful instrument for data-driven study of linguistic structures at various levels. The proliferation of treebanks has led to a very large number of resources available for different languages, which can support comparative research of various issues cross-linguistically. In recent years, a growing number of treebanks has also become available for ancient languages and for different historical stages of the same language: the York-Toronto-Helsinki corpus (<http://www-users.york.ac.uk/~lang22/YcoeHome1.htm>) and the Penn Corpora of Historical English (<https://www.ling.upenn.edu/hist-corpora/>) for English, Tromsø Old Russian and OCS Treebank (<https://nestor.uit.no>) and RRuDi (<https://www.slawistik.hu-berlin.de/de/member/meyerrol/subjekte/rrudi>) for Russian, PROIEL (<http://www.hf.uio.no/ifikk/english/research/projects/proiel/>) for various ancient Indo-European languages and recently extended to host treebanks for medieval stages of Romance and Germanic languages, Perseus Latin and Ancient Greek Dependency Treebanks (<http://nlp.perseus.tufts.edu/syntax/treebank/>) for Latin and Ancient Greek, the Index Thomisticus Treebank (<http://itreebank.marginalia.it>) for Latin, and several others. This allows data extraction aimed to assessing the scope and the effects of diachronic developments, managing a large amount of data and retrieving information whose relevance can then be evaluated through statistical methods. Possible issues that can be tackled through diachronic treebanks are potentially numerous and of different nature, and include increasing or decreasing productivity of syntactic or morphological constructions, and, most interesting, interrelationships between different changes that have previously been considered unrelated or whose interrelation is otherwise hard to prove.

The basic research questions concern the possible contribution of historical treebanks to historical linguistics research: how can data extraction from treebanks bring new insights on language change? What sort of micro- or macro-changes can be fruitfully studied with the help of syntactically parsed historical corpora? What features should treebanks have in order to bring new advances in historical linguistics?

Possible topics include (but are not limited to):

- historical developments of constructions as evidenced by data extracted from diachronic treebanks;
- suitability of different types of treebanks (constituent-based vs. dependency-based) for research on specific diachronic changes;
- correlations between developments in different areas of a language's grammar;
- similarities and differences between parallel developments of similar changes in different languages;
- how evidence from already known and documented diachronic change can give input for annotation;
- how semantic and/or pragmatic information can be supplied in order to better understand the rationale of changes highlighted by data extracted from treebanks;
- specific issues raised by the development of diachronic treebanks;
- methods and tools to build and access diachronic treebanks;
- issues in data selection for representativeness purposes;
- issues pertaining to scarce and non-standardized data