

Treebank querying with GrETEL 3: bigger, faster, stronger

**Liesbeth Augustinus, Bram Vanroy and Vincent Vandeghinste –
University of Leuven**

We describe the new version of GrETEL (<http://gretel.ccl.kuleuven.be/gretel3>), an online tool which allows users to query treebanks by means of a natural language example (example-based search) or via a formal query (XPath search).

The new release comprises an update to the interface and considerable improvements in the back-end search mechanism.

The update of the front-end is based on user suggestions. In addition to an overall design update, major changes include a more intuitive query builder in the example-based search mode and a visualizer for syntax trees that is compatible with all modern browsers. Moreover, the results are presented to the user as soon as they are found, so users can browse the matching sentences before the treebank search is completed. We will demonstrate that those changes considerably improve the query procedure.

The update of the back-end mainly includes optimizing the search algorithm for querying the (very) large SoNaR treebank. Querying this 500-million word treebank was already made possible in the previous version of GrETEL, but due to the complex search mechanism this often resulted in long query times or even a timeout before the search completed. The improved version of the search algorithm results in faster query times and more accurate search results, which greatly enhances the usability of the SoNaR treebank for linguistic research.

Keywords: treebank querying – corpus linguistics – search tool