Web-based Annotation Interface for Derivational Morphology

Lukáš Kyjánek
kyjanek@ufal.mff.cuni.cz
Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Prague, Czechia

Summary

- Manual annotation of not only derivational morphology is easier and faster, and thus cheaper if the data is visualised.
- A simple open-source visual interface has been developed for manual annotation of tree-shaped data structures.
- Annotators validated the interface on derivational morphology.

Interface for Manual Annotation

- The interface is created as web-based and responsive to prevent problems with installation and work with various devices.
- It is programmed by using simple techniques: HTML5, CSS3, and JavaScript (jQuery, CytoScape.js, and Notify.js).
- It works on a client's side and processes JSON input/output.
- The interface screen consists of top and bottom button bars and a central canvas for data visualisation and annotation.
- It can delete/restore edges, search for an instance by typing its ID or characters into the textbox, and check treeness.
- It converts the input/output from/to traditional TSV format.

Human Validation

- Two methods of manual annotation were compared: text processor (TSV) versus visual interface (JSON) on a sample of 20 derivational sub-families from Czech.
- The experiment involved 12 human annotators (university students of other than linguistic studies); half of them used text processor and the others used the interface.
- Result: The annotation process is faster if the interface is used instead of the traditional annotating in text processor.

Conclusion & Future Work

- The robustness of the interface was tested also on annotating syntactic data from Universal Dependencies. It worked well but implementation of new functions would be useful.
- Takeaway Messages
  - Developing the interface in collaboration with annotators is crucial as they provide necessary feedback to developers.
  - Although the interface is created by relatively basic techniques, it can save notable annotators’ time and effort.