Web-based Annotation Interface for Derivational Morphology

Lukáš Kyjánek

kyjanek@ufal.mff.cuni.cz

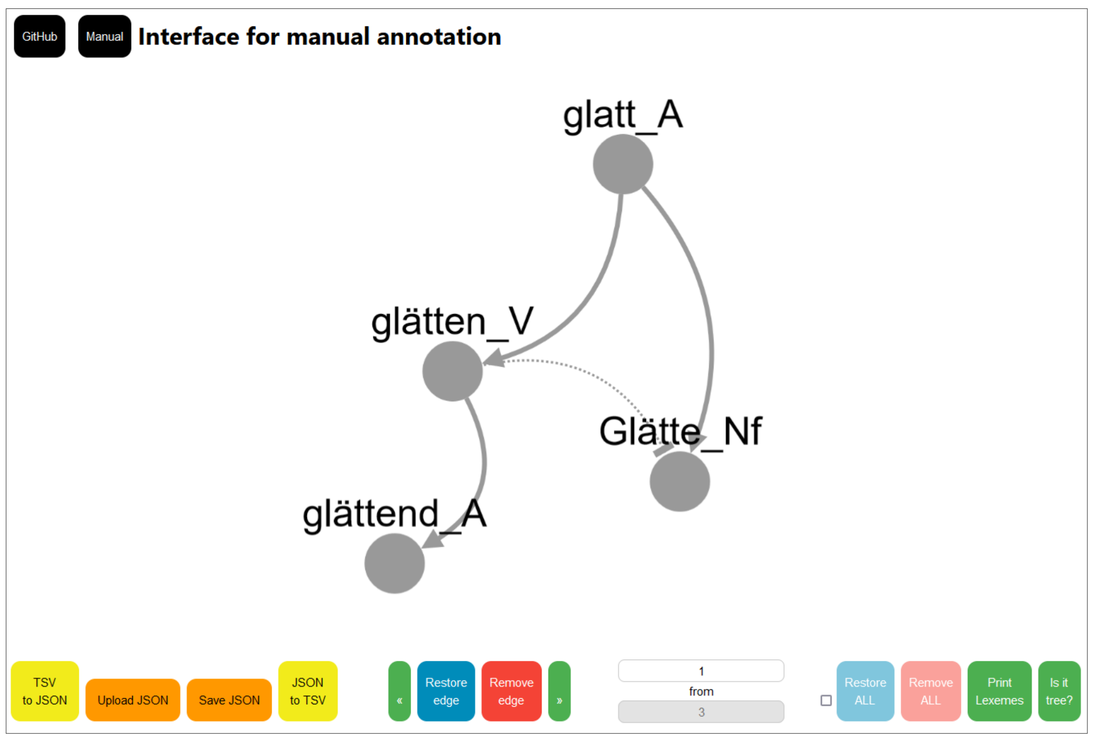
Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Prague, Czechia

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`, `CCS3`, 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.

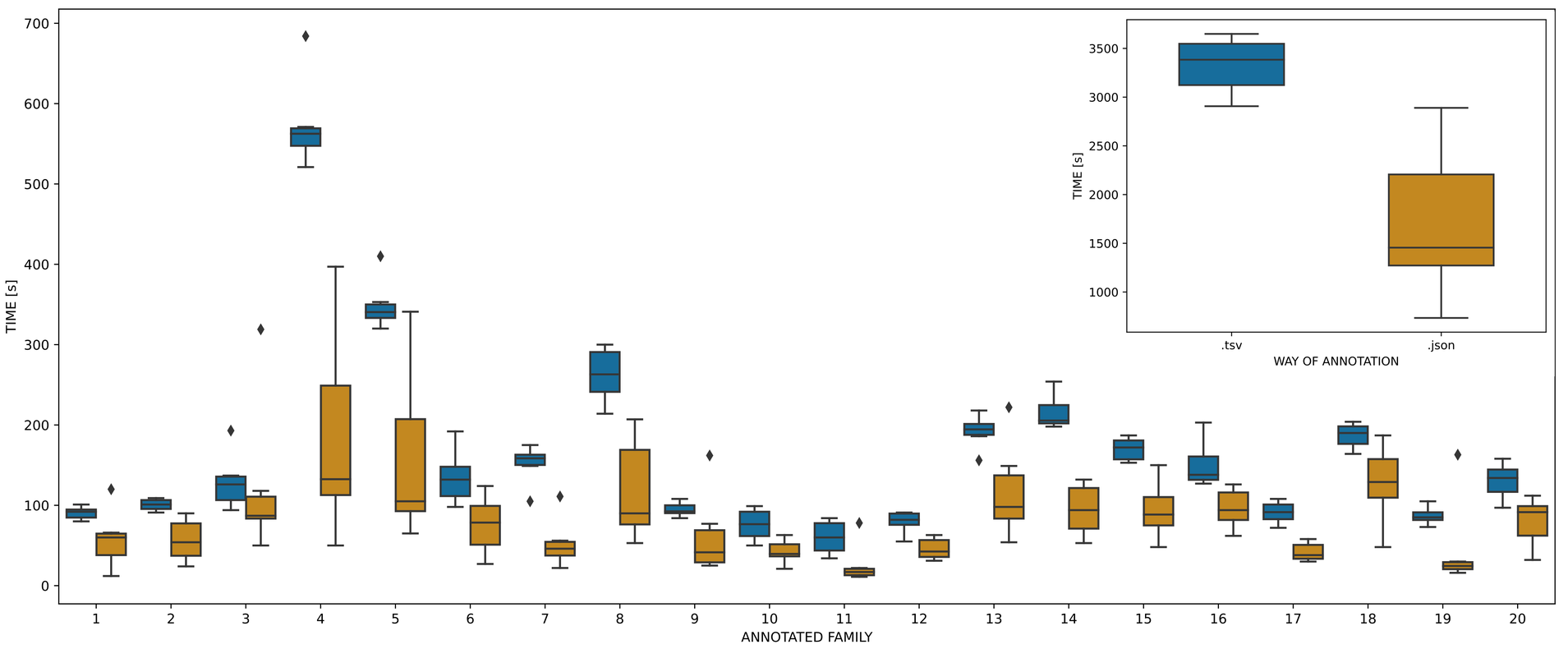


Human Validation

- 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.



ARTICLE

Presented at the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 10–15 Jul 2022.

Supported by grants No. GA19-14534S of the Czech Science Foundation and No. START/HUM/010 of Grant schemes at the Charles University (reg. No. CZ.02.2.69/0.0/0.0/19_073/0016935), and LINDAT/CLARIAH-CZ project of the Ministry of Education (LM2015071, LM2018101). It was using language resources developed, stored, and distributed by the LINDAT/CLARIAH-CZ project.