Towards the automatic detection of syntactic differences

Kroon, M.S.

Propositions

accompanying the dissertation

Towards the Automatic Detection of Syntactic Differences

by

Martin Kroon

1. Computational algorithms can be successfully deployed to semi-automatically generate hypotheses on syntactic differences between languages. *(This dissertation.)*

2. The Levenshtein distance on POS sequences, the graph-edit distance on dependency parses and the sentence-length ratio can be used as indicators to filter out syntactically incomparable sentence pairs from a parallel text corpus with reasonable success. However, the task remains hard, because it is inherently subjective and difficult to define. *(Chapter 2.)*

3. The Minimum Description Length principle is a good framework for summarizing syntactic patterns in language, and can be used to generate hypotheses on syntactic differences between languages based on the distributional differences in parallel text. *(Chapter 3.)*

4. Bitext word alignment can be used successfully to gain insights into grammatical and syntactic phenomena in an unannotated target language through the semi-automatic analysis of annotations of the source language. *(Chapter 4.)*

5. The workings of any tool developed for the purposes of a Natural Language Processing (NLP) task, its model and its output must be transparent and interpretable for the human researcher, who is invaluable to the process.

6. The environmental impact of computational research, including NLP, and the training of complex models used therein often goes unnoticed, and is higher than many realize; less environmentally impactful algorithms with lower complexity are therefore preferable and should first be explored or even exhausted.

7. Bigger data is not always the answer for improving the performance of NLP algorithms or models, and we should instead strive to design algorithms that process the data in a more efficient and smarter manner.

8. All of science begins and ends with language. The study of language itself is therefore as important as any other field of science.

9. In the current climate of fake news and conspiracy theories, we scientists and scholars have more than ever a duty to inform people openly, honestly, completely and correctly, and publish our papers and tools where they will be accessible for everyone.