POSTagging and Semantic Dictionary Creation for Hittite Cuneiform

Timo Homburg
timo.homburg@hs-mainz.de
Mainz University of Applied Sciences, Germany

Presentation Topic and State Of The Art

On our poster we want to present ongoing work to create an automatic natural language processing tool for Hittite cuneiform. Hittite cuneiform texts are to this day manually transcribed by the respective experts and then published in a transliteration format (commonly ATF). Pictures of the original cuneiform tablet may be provided and more rarely cuneiform representations in Unicode are present. Due to recent advancements in the field (such as Cuneify) an automatic translation of many Hittite cuneiform transliterations to their respective cuneiform representation is possible.

Research Contributions

We build upon this work by creating tools that aim to automatically translate Hittite cuneiform texts to English from either a Unicode cuneiform representation or their transliteration representation.

POSTagger

We have created a morphological analyzer to detect nouns, verbs, several kinds of pronouns, their respective declinations and appendices as well as structural particles. On a sample set of annotated Hittite texts from different epochs in cuneiform and transliteration representation we have evaluated the morphological analyzer, its advantages, problems and possible solutions and intend to present the results as well as some POSTagging examples in section one of our poster.

Dictionary Creation

Dictionaries for Hittite cuneiform exist in often non-machine readable formats and without a connection to Semantic Web concepts. We intend to change this situation by parsing digitally available nonsemantic dictionaries and using matching algorithms to find concepts of the English translations of such dictionaries in the Semantic Web e.g. DBPedia or Wikidata. Dictionaries of this kind are stored using the Lexical Model for Ontologies (Lemon). In addition to freely available dictionaries we intend to use expert resources developed by the academy of sciences in Mainz/Germany to verify and extend our generated dictionaries. We intend to present the dictionary creation process, statistics about the content of generated dictionaries and their impact in section two of our poster.

Machine Translation

Using the newly created dictionaries as well as the POSTagging information we intend to test several automated machine translation approaches of which we will outline the process and possible approaches in poster section three.

Contributions for the Communities

With our approaches we intend to contribute to the archaeological community in Germany by analysing Hittite cuneiform tablets. Together with work from the University of Heidelberg on image recognition of cuneiform tablets, we want to focus on creating a natural language processing pipeline from scanning cuneiform tablets to an available translation in English.