**Word proximity and dependencies in parliamentary discourse in Finnish parliament**

In digital humanities, so-called distant reading often relies on methods based on statistics on proximity of words used in the given textual context. Prior to the data analysis, textual data is typically lemmatized and stop words are removed, which from linguistic perspective erases a layer of textual meaning as well as the actual informational content from the sentences (see e.g. Lambrecht 1996). Even if the content words can provide an overall comprehension of discourse entities mentioned in a given text, the majority of research questions in the fields of humanities and social sciences are concerned with *how* the entities are spoken about as well as the explicated *relations* between the entities, both typically expressed grammatically: with function words, inflection and word order.

This study sets out to evaluate the extent to which a computational analysis based on grammatical relations (as in syntactic dependencies) - instead of word proximity - can capture central features of temporal relations expressed in parliamentary debate discourse and elaborate the methodological approaches to parliamentary data and political temporality. Parliamentary discourse is nothing but straightforward; discourse entities and processes MPs refer to during parliamentary debates are typically abstract and expressed with complex noun phrases or infinitive constructions. Due to this, the rhetorical wordings under scrutiny can comprise the core of the expression as well as stand as a more distant frame or a modifier for what actually is stated. We report three cases where a noun referring to time is used in different syntactic positions: head of a noun phrase, modifier of a noun phrase and as one of the main arguments in the clause (subject, object). We focus on plenary sessions of Finnish parliament, a showcase for highly inflectional language with flexible word order.

The data set consists of the official records of Finnish parliamentary debates from 1980 to 2022 (see Andrushchenko et al. 2021) and is dependency-parsed with the Finnish neural parser (Turku NLP, Kanerva et al. 2018). Universal Dependencies provide a language independent framework thus it also enables systematic comparisons between languages used in different parliaments. The analyses are obtained from a machine-learned parser in a standardized syntax tree format. They are then sent to a rule-based pattern matching tool, which finds subtrees and sequences of trees satisfying relevant conditions such as "sentence where the word 'future' is used as subject or object", or "sequence of sentences in the past tense".

**References**

Andrushchenko, M., Sandberg, K., Turunen, R., Marjanen, J., Hatavara, M., Kurunmäki, J., Nummenmaa, T., Hyvärinen, M., Teräs, K., Peltonen, J. & Nummenmaa, J. (2021) Using parsed and annotated corpora to analyze parliamentarians' talk in Finland. *Journal of the Association for Information Science and Technology (JASIST*) https://doi.org/10.1002/asi.24500

Kanerva, J., Ginter, F., Miekka, N., Leino, A., Salakoski, T. (2018). Turku neural parser pipeline: An end-to-end system for the conll 2018 shared task. Proceedings of the conll 2018 shared task: Multilingual parsing from raw text to universal dependencies. Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/K18-2013

Lambrecht, K. (1996) *Information Structure and Sentence Form : Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press