

Testing trade-offs between gender and number indexing and other cues to A and P arguments *A corpus-based perspective*

All languages have some formal and semantic cues that help the addressee understand “who did what to whom”, such as word order, case marking, agreement (indexing) and semantic and pragmatic properties of the referents. The trade-off relationship between word order and case marking has been well studied (Sapir, 1921; Sinnemäki, 2008; Koplenig *et al.*, 2017; Levshina, 2021), but what role indexing plays is less known. This study aims at developing a new methodological perspective to investigate trade-offs between gender and number indexing and other cues to A and P arguments using data from corpora.

The body of research on trade-offs is characterized by two assumptions that are seldom explicitly tested. There is often an implicit assumption that language users try to communicate efficiently and that complexity in one area of language demands simplicity in another. Kemp & Regier (2012) find a trade-off between cognitive and communicative costs in lexical systems of kinship words or color terms and Koplenig *et al.* (2017) demonstrate trade-offs between information conveyed by word order and word structure in Bible translations. Conversely, Levshina (2021) finds that in investigating the trade-off between cues for A and P arguments, there is little evidence for efficient language use and argues that the interaction between cues can best be explained by sociolinguistic factors. Similarly, Sinnemäki (2008) tests the trade-off theory on head/dependent marking and word order correlations and finds that, while all correlations were indeed negative, only two were statistically significant: the correlations between word order and dependent marking, and between word order and morphological marking (the sum of head and dependent marking).

The research goals are twofold. The first one is to develop a quantitative corpus-based measure of the contribution of indexing (agreement) to disambiguation of A and P. The second is to check if there are trade-offs (correlations) between this measure and other corpus-based variables which represent case marking, word order rigidity, semantic tightness and the position of the verb in a clause.

Nominal transitive clauses with gender and number indexing were obtained from the SUD treebanks, a collection of syntactically annotated corpora based on Universal Dependencies, for a sample of 30 cross-linguistically diverse languages (Gerdes *et al.*, 2018). We test pairwise correlations between agreement and the other cues with the help of mixed models which include phylogenetic information and geographical distances between the sample languages. The expected results are a matrix of correlations between the variables based on these mixed models.

References

- Gerdes, Kim, Guillaume, Bruno, Kahane, Sylvain, & Perrier, Guy. 2018 (Nov.). SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. *In: Lynn, Teresa, & Schuster, Sebastian (eds), Universal Dependencies Workshop 2018.*
- Kemp, Charles, & Regier, Terry. 2012. Kinship Categories Across Languages Reflect General Communicative Principles. *Science (New York, N.Y.)*, **336**(05), 1049–54.
- Koplenig, Alexander, Meyer, Peter, Wolfer, Sascha, & Müller-Spitzer, Carolin. 2017. The statistical trade-off between word order and word structure – Large-scale evidence for the principle of least effort. *PLoS ONE*, **12**(03), e0173614.
- Levshina, Natalia. 2021. Cross-Linguistic Trade-Offs and Causal Relationships Between Cues to Grammatical Subject and Object, and the Problem of Efficiency-Related Explanations. *Frontiers in Psychology*, **12**.
- Sapir, E. 1921. *Language: An Introduction to the Study of Speech*. A Harvest book HB7. Harcourt, Brace.
- Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. *Pages 67–88 of: Miestamo, M., Sinnemäki, K., & Karlsson, F. (eds), Language Complexity: Typology, contact, change*. Studies in language companion series. John Benjamins.