

Rhetorical structure as a factor of referential choice: in search of an optimal measurement

Zalmanov D.A.

Lomonosov Moscow State University

dm.zalmanov@gmail.com

One of the factors affecting the choice of referring expressions in discourse is the proximity of the current mention of a given referent to its previous mention(s). It can be measured in terms of discourse units that immediately follow one another (linear distance, LinD), as well as in terms of hierarchical, or rhetorical structure of discourse (rhetorical distance, RhetD). RhetD takes structural complexity of discourse into account, and cannot be derived from plain linear discourse structure. In certain instances, RhetD differs from LinD significantly, and may provide a good explanation for failed pronominalization at small LinD, and vice versa. However, it is unclear how RhetD should be calculated for its effects to be captured with sufficient precision. I apply machine learning techniques to test various calculation options against data from the WSJ MoRA 2015 corpus (Kibrik et al. 2016). The corpus uses rhetorical structure annotation from the RST Discourse Treebank (Carlson et al. 2002) without alterations to the original annotation conventions. Kibrik and Krasavina (2005) suggest the following possibilities for adjustment of the current RhetD calculation rules:

1. Symmetrical structures, i.e. structures with a coordinate nucleus-nucleus relation. Contribution of such structures to overall rhetorical complexity is unclear, and may be counted either the same way as for asymmetrical structures or differently.

2. Different types of rhetorical relations allow for a differentiated weighing, depending on the level of assumed processing simplicity.

Adjustments to calculation rules yield a higher efficiency of RhetD as a factor for predicting the type of referring expressions.

Carlson, L., Marcu, D., and Okurowski, M. (2002). *RST Discourse Treebank*. Philadelphia, PA: Linguistic Data Consortium.

Kibrik, A. A., and Krasavina, O. N. (2005). A corpus study of referential choice: the role of rhetorical structure. *Papers from the Annual International Conference "Dialogue": Computational Linguistics and Intellectual Technologies*, eds I. M. Kobozeva, A. S. Narin'jani, and V. P. Selegey (Moscow: Nauka), 561-569.

Kibrik AA, Khudyakova MV, Dobrov GB, Linnik A and Zalmanov DA (2016) Referential Choice: Predictability and Its Limits. *Front. Psychol.* 7:1429. doi: 10.3389/fpsyg.2016.01429

Keywords: generation of referring expressions, rhetorical structure, mathematical modeling.