

Classifying Discourse-Givenness (as Defined in Coreference Resources)

Julia Ritz

Determining whether a noun phrase is discourse-given (i.e. referring to something already mentioned, c.f. Prince, 1981, 1992) or discourse-new is an important step in information structure and discourse analysis. The goal of my work is to automatically accomplish this task. As coreference annotation subsumes discourse-givenness, I am using existing coreference-annotated corpora for training a classifier. The talk will comprise (i) a discussion of existing coreference corpora for English (incl. e.g. van Deemter & Kibble's (2000) comments), and (ii) a presentation of classification experiments and results (incl. Ritz, 2010). It will conclude with an outlook on the applicability of such classifiers in corpus linguistics and Natural Language Processing.

Prince, Ellen F. 1981. "Toward a taxonomy of given/new information." In P. Cole, (ed.), *Radical Pragmatics*. NY: Academic Press, pp. 223-255.

Prince, Ellen F. 1992. "The ZPG letter: Subjects, definiteness, and information-status." In S. Thompson and W. Mann, (eds.), *Discourse Description: Diverse Analyses of a Fund-Raising Text*. John Benjamins, pp. 295-325.

van Deemter, Kees; Kibble, Rodger, 1999. "On coreferring: coreference in MUC and related annotation schemes". *Computational Linguistics* 26 (4). MIT Press.

Ritz, 2010. "Using tf-idf-related measures for determining the anaphoricity of noun phrases." In M. Pinkal, I. Rehbein, S. Schulte im Walde, A. Storrer, (eds.), *Proceedings of KONVENS 2010*, p. 85-92, Saarbrücken, September 2010.