
Predicting processing cost of anaphora resolution

Olga Seminck
Université Paris Diderot
 olga.seminck@cri-paris.org

Pascal Amsili
Université Paris Diderot
 pascal.amsili@linguist-univ-paris-diderot.fr

Donnerstag
 09.03.2017
 10:00 – 10:30
 B3 2, 0.03

AG1

Anaphora resolution, for human speakers, can be more or less costly depending on various factors like ambiguity, syntactic complexity and semantic plausibility. The variation of cost has been measured by many studies in psycholinguistics, through experimental paradigms like self-paced reading, or eye-tracking. Our project aims at devising a system, inspired by current NLP coreference resolution systems, that can predict a processing cost for anaphora resolution, which can be evaluated by running our system on human data coming from psycholinguistic experiments, or eye-tracking corpora e.g. the Dundee Corpus (Kennedy 2003). Inspired by surprisal theory (Hale 2001) and the entropy reduction hypothesis (Hale 2006), we propose a continuous, incremental measure that assigns processing cost to anaphora. Our measure reflects how certain a probabilistic anaphora resolution system is about its decisions. To do so, with a simple anaphora resolution tool, we compute a probability distribution over all antecedent candidates of an anaphor and calculate entropy over it. We hypothesize that the entropy over this distribution can be seen as the processing cost of the resolution of the anaphor. So the smaller the entropy, the less processing cost that is predicted. A first study we conducted on two biases that were discovered by psycholinguists (Subject Assignment Strategy and Parallel Function Hypothesis (e.g. Crawley 1990)) showed that our model was able to simulate human performance in these matters: it assigned the pronouns in a way comparable to human participants and the cost it predicted corresponded to reading times recorded in self-paced reading experiments.

References: • Rosalind A Crawley et al. (1990). The use of heuristic strategies in the interpretation of pronouns. In: *Journal of Psycholinguistic Research* 19 4, 245–264. • John Hale (2001). A probabilistic Earley parser as a psycholinguistic model. In: *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*. Association for Computational Linguistics, 1–8. • John Hale (2006). Uncertainty about the rest of the sentence. In: *Cognitive Science* 30.4, 643–672. • Alan Kennedy et al. (2003). The dundee corpus. In: *Proceedings of the 12th European conference on eye movement*.