

Summary of:
References to Named Entities: a Corpus Study

This paper examines how writers refer to named entities within the context of a corpus of news articles. Specifically, they focus on identifying the syntactic properties of first and subsequent mentions of people. The authors wish to further understand these references in order to improve the coherence of automatically generated summaries. They model the references using Markov chains, and then use the results to create rewrite rules for automatically generated summaries. They apply these rules to eleven summaries generated by the Columbia University summarizer, and then ask four human judges whether they prefer the original generation, or the one modified using the rewrite rules. Of the total 44 decisions, 89 percent were decided in favor of the rewritten version.

I was drawn towards this paper because it had a somewhat more abstract evaluation method. I do not have much experience reading NLP papers, but often many CS papers simply look at some sort of stringent numerical evaluation. This works well in some contexts, but falls short in plenty of other ones. It was interesting to see that such simple rewrite rules could make such a difference in which summary someone preferred.