

AI and Complex Systems for Human Languages

William A. Kretzschmar, Jr.
University of Georgia and Uppsala University

Natural Language Processing (NLP), the field concerned with interactions between computers and human language, especially the processing of large corpora of authentic language, has proceeded through three main stages. The first, symbolic NLP, was essentially rule based, following generative principles. The second, beginning in the 1990s and still in use today, applies advanced statistics to corpora of language for machine learning. The third stage, neural NLP, applies neural network models in order to improve machine learning. Completion of language tasks has become more and more successful over time, and yet we are still some distance from the goal of fluent processing of human language by computers. In this paper I will argue that part of the problem is our understanding of the underlying status of human language, which is not essentially rule-based but instead probabilistic based on hyperbolic asymptotic frequency distributions of linguistic variants (see Kretzschmar 2009, 2015). Neural networks are a better match than statistical approaches for this distributional pattern, the output of the complex system of human interaction in language, so long as we understand that meaning is inherently probabilistic in line with the frequency profiles of linguistic variation. Human speakers perceive a best-fit meaning from their experience with the frequency profiles of language production. In order for an AI system to be most like human speakers, the AI must also be able work with asymptotic hyperbolic frequency profiles and construct possible meanings from them. Because meaning in human language is not binary, either right or wrong, but instead multidimensional, the best AI systems should be able to evaluate constructed meanings not just as right or wrong, but as more or less probable, subject to reevaluation as a stream of language continues. Neural network outputs enable just such a process, as shown in Kretzschmar 2015: Ch. 6. The most natural AI system for processing human language will thus be interactive, in such a way that the AI can make new evaluations as new data is received and at each point recognize the meaning(s) at the top of asymptotic hyperbolic distributional pattern.

References

- Kretzschmar, William A., Jr. 2009 *The Linguistics of Speech*. Cambridge: Cambridge University Press.
- Kretzschmar, William A., Jr. 2015. *Language and Complex Systems*. Cambridge: Cambridge University Press.