A Computational Approach to Metalanguage and the Use-Mention Distinction

In linguistic communication it is sometimes necessary to refer to features of language, such as orthography, vocabulary, structure, pragmatics, or meaning. Metalanguage enables a speaker to select a linguistically-relevant referent over (or in addition to) other typical referents. It is both pervasive and, paradoxically, the subject of limited attention in research on language technologies. Metalanguage encodes unusually direct and salient information about language, but simple examples thwart parsers and other common language analysis tools.

In this talk, I will first present on a framework for identifying and analyzing instances of metalanguage, in an effort to reconcile the many theoretical treatments of the phenomenon for empirical use. This will include a definition of mentioned language, a common form of metalanguage with many practical roles in communication. I will then describe the creation of the first tagged and delineated corpus of English metalanguage, built by applying a combination of stylistic and lexical heuristics to Wikipedia article text. Finally, I will present preliminary results from using NLP methods to automatically identify mentioned language in text. These contributions validate the feasibility of building language technologies that can exploit the salient information about language that metalanguage encodes.

Shomir Wilson is a Postdoctoral Associate with the Mobile Commerce Lab at Carnegie Mellon University’s Institute for Software Research. He is also an NSF International Research Fellow, and will spend a year at the University of Edinburgh starting this July. He received his PhD in Computer Science from the University of Maryland in 2011, and during graduate school he twice received grants from the NSF's East Asia and Pacific Summer Institutes. His most recent research is in usable privacy and security; he is also interested in how people speak about language, and what computers can learn from metalanguage and metadialogue.

http://www.cs.cmu.edu/~nlp-lunch/