

## **What counts as useful, relevant data? An I-Language perspective**

Carson Schütze

*Linguistics Department, University of California, Los Angeles*

In this talk I strive to address the title question of the workshop from the perspective of generative grammar. Since our object of study obviously affects our use of data, I begin by explicating “I-language” (in contrast to “E-language”), which is what we are trying to investigate in “Chomskyan” linguistics. It is vital to distinguish this notion—a domain of tacit knowledge (a.k.a. competence)—from the ways in which we may try to study it: A so-called grammaticality judgement, for example, does not inherently bear on I-language any more than does, say, a sentence extracted from a corpus of recorded texts. Both are simply small pieces of human behaviour that involve language, and hence could in principle be relevant to most any question one might want to pose about language. As such, they are (*ceteris paribus*) equally worthy of documentation. Whether either one is in fact relevant depends on what the specific question is. And whether either one is actually useful in answering the question is not a property of the datum itself but rather of its relationship to other relevant available data. In the second half of the talk I exemplify these points by examining some particular kinds of data that are used in generative grammar, and I consider how to determine when they are relevant and useful. Among these will be the growing trend of using Web searches in theoretical argumentation. I show that while some such uses are reasonable, even insightful, too many employ data whose usefulness and relevance are highly suspect. Furthermore, claims that this and other “cutting edge” techniques have rendered classic data sources such as grammaticality judgements no longer useful are thoroughly misguided.