

Title: Investigating The Properties Of Speech Using Linguistic Corpora

Presented by: Ariel Goldberg

Authors: Ariel Goldberg, Department of Psychology, School of Arts and Sciences, Tufts University

Abstract:

Psycholinguistics is the sub-field of psychology that studies the cognitive processes involved in producing and comprehending language. Psycholinguistics has long used big datasets to assist in the design of experiments. For example, by analyzing large written texts, researchers have been able to estimate the frequency with which words are used in a language and experiments have shown that the way the mind processes words is strongly influenced by this frequency. These and related findings have strongly influenced cognitive theories of language processing. In recent years, the creation of corpora of spontaneous speech has allowed researchers to investigate in exquisite detail the cognitive mechanisms involved in producing speech. Rather than simply serving as a tool in experimental design, contemporary corpora have become the data themselves.

In this talk I will describe the central role that big data played early in the history of psycholinguistics, the challenges faced by researchers when such data are not available (such as is the case with American Sign Language (ASL)), and the efforts by my lab to collect such information for ASL by alternative means. I will also present two studies conducted in my lab that used corpora of spontaneous speech to investigate how the mind stores words with multiple parts (e.g., running; run+ing) and how this internal structure subtly changes the way these words are pronounced.