Contribution ID: **85**   Type: **Talk**

Vague and vacuous morphological decomposition

Saturday, June 24, 2017 9:30 AM (20 minutes)

Empirical tests of productivity and decomposition use etymological, semantic, and distributional criteria to classify items into categories, such as transparently affixed, opaquely affixed, pseudoaffixed, or unaffixed. Such classifications require analysts, as well as language learners, to know semantic and syntactic properties of forms, identify potential base forms, and determine which (base, derived) pairs share the same relationship. Language learners often have sparse or incomplete information about semantic and syntactic properties of words and related bases; indeed, positing potential morphological decompositions is a useful first step in decoding the category and meaning of words. Numerous models of morphological segmentation have been developed using distributional techniques such as MDL or Bayesian inference; however, these models focus almost exclusively on segmenting words into morphemes, and very little on establishing productivity or selectional restrictions. Furthermore, segmentation in these models is all or nothing.

In this talk, I describe the application of a supervised morphological learning model (the Minimal Generalization Learner; Albright and Hayes 2003) to data with incomplete or missing information about pairs, with the goal of discovering the pairs and their decomposition simultaneously. Given no information about syntactic or semantic properties, the model resembles other distributional learners, discovering recurring pieces. However, the model can also identify potential morphological relations with varying degrees of semantic specificity and predictive power. Thus, the grammar may simultaneously contain semantically vague or vacuous rules encoding frequently recurring formal relations, alongside rules for narrow prescribed morphosyntactic functions. The model makes several interesting predictions. The first concerns gradient decomposition: pseudoaffixed forms may be decomposed with varying probabilities. The model also favors "vague" decomposition, in which the base and derived form share some morphosyntactic properties with transparently related pairs, over "vacuous" decomposition, in which the relation is purely formal. Finally, the model favors decomposition for strings that represent multiple homophonous affixes.

Primary author: ALBRIGHT, Adam (Massachusetts Institute of Technology)

Presenter: ALBRIGHT, Adam (Massachusetts Institute of Technology)

Session Classification: Symposium 3 – Theoretical linguistics

Track Classification: Symposium 3 –Theoretical linguistics