

## **UC Merced**

# **Proceedings of the Annual Meeting of the Cognitive Science Society**

### **Title**

Incorporating Semantic Constraints into Algorithms for Unsupervised Learning of Morphology

### **Permalink**

<https://escholarship.org/uc/item/07v3x3q6>

### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

### **Authors**

Tenenbaum, Abi

Levy, Roger

### **Publication Date**

2019

Peer reviewed

# **Incorporating Semantic Constraints into Algorithms for Unsupervised Learning of Morphology**

**Abi Tenenbaum**

Commonwealth School, Boston, Massachusetts, United States

**Roger Levy**

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

## **Abstract**

A key challenge in language acquisition is learning morphological transforms relating word roots to derived forms. Unsupervised learning algorithms can perform morphological segmentation by finding patterns in word strings (e.g. Goldsmith, 2001), but struggle to distinguish valid segmentations from spurious ones because they look only at sequences of characters (or phonemes) and ignore meaning. For example, a system that correctly discovers  $\text{;add -s}_i$  as a valid suffix from seeing dog, dogs, cat, cats, etc, might incorrectly infer that  $\text{;add -et}_i$  is also a valid suffix from seeing bull, bullet, mall, mallet, etc. We propose that learners could avoid these errors with a simple semantic assumption: morphological transforms should approximately preserve meaning. We extend an algorithm from Chan (2008) by integrating proximity in vector-space word embeddings as a criterion for valid transforms. On the Brown CHILDES corpus, we achieve both higher accuracy and broader coverage than the purely syntactic approach.