

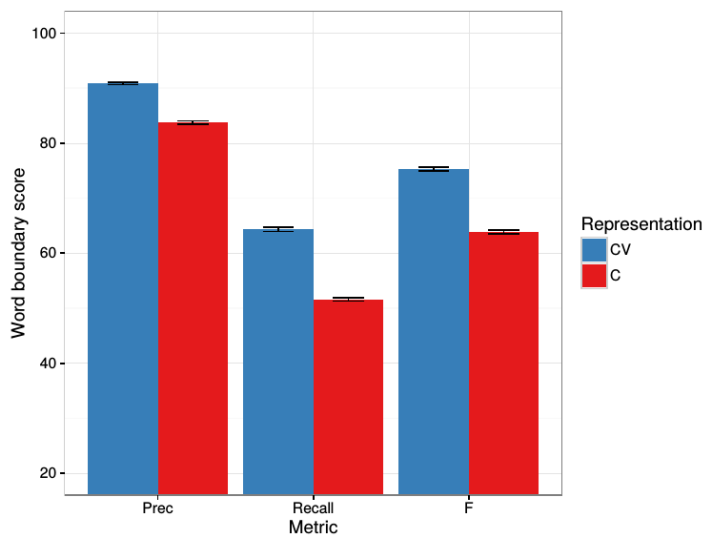
## Learning and modeling Semitic roots

For an infant acquiring language, a fundamental task is to identify individual words in the continuous speech/sign stream. This segmentation problem is an issue for concatenative languages like English and perhaps even more so for nonconcatenative, Semitic languages like Arabic. Recent work has suggested that if the learner “guesses” at possible words in the input and keeps track of them, she will induce word boundaries and an initial lexicon (Goldwater et al 2009). Independently, it has also been observed that infants are able to distinguish vowels from consonants from very early on (e.g. Werker and Tees 1984, Hochmann et al 2011). However, it is unknown whether the same learning mechanism is used for languages with different morphological systems.

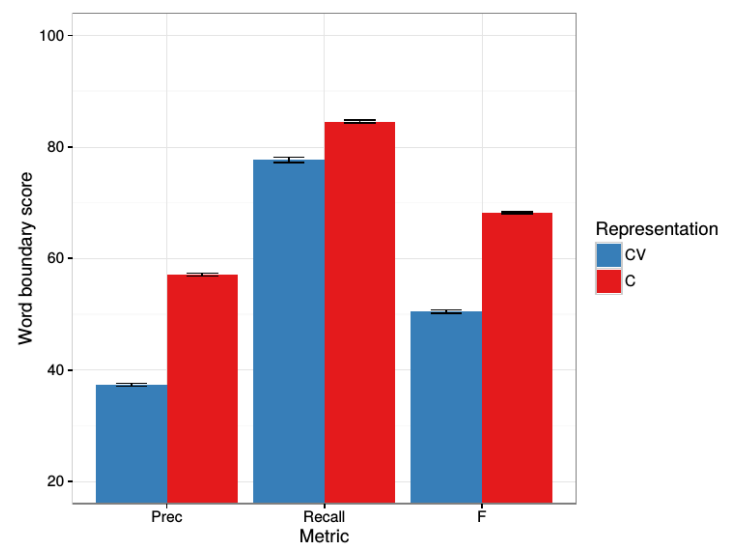
We hypothesize that for Arabic, the learner first keeps track of consonant cooccurrence patterns rather than whole words. These patterns would then be combined with vowels and map onto “roots” and “templates” (McCarthy 1979). The child would learn two things at once: how to segment the input stream, as well as morphological patterns which feed additional generalizations.

I survey the results of computer simulations which show that this difference in representation – consonant-only rather than consonant-and-vowel – hinders the learner of English, (1a), but is indeed beneficial for a learner of Arabic, (1b). Figure (1) plots the performance of a Bayesian segmentation model (Goldwater et al 2009) evaluated by Precision, Recall and F-measure.

(1) a. English CHILDES



b. Arabic Gigaword



Similar results obtain when learning lists of roots (the lexicon) and basic phonotactics (OCP-Place). With these results in mind, I summarize recent developmental findings in another Semitic language, Modern Hebrew. I will discuss how the Semitic system can be learned by the infant, and how the learning mechanism differs between the acquisition of a Semitic language and a non-Semitic language. An explicit process is thus proposed for how meaning follows form.

### Selected References

- Goldwater, Sharon, Thomas L Griffiths, and Mark Johnson (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition* 112: 21–54.
- Hochmann, Jean-Remy, Silvia Benavides-Varela, Marina Nespou, and Jacques Mehler (2011). Consonants and vowels: different roles in early language acquisition. *Developmental Science* 14 (6): 1445–1458.
- Werker, Janet F., and Richard C. Tees (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7:49–63.