

## **UC Merced**

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

**Title**

Data-Oriented Modeling of Infant Vocal Babbling and the Emergence of Syntax

**Permalink**

<https://escholarship.org/uc/item/55k503hp>

**Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

**ISSN**

1069-7977

**Author**

Cochran, Dave

**Publication Date**

2007

Peer reviewed

# Data-Oriented Modelling of Infant Vocal Babbling and the Emergence of Syntax

**Dave Cochran (davec@cs.st-andrews.ac.uk)**

School of Computer Science, University of St. Andrews  
North Haugh, St. Andrews, Fife KY16 9SX, Scotland

**Keywords:** First Language Acquisition; Data-Oriented Parsing; Babbling; Early Syntax.

## Introduction

This poster outlines and motivates a novel theory of syntactic bootstrapping, Exemplar-Based Phonological Bootstrapping, which may be understood as an interpretation from within the Data-Oriented Parsing paradigm (an exemplar-based computational formalism used mostly in linguistic parsing tasks, wherein analyses of new inputs are constructed by extracting and recombining fragments from stored representations of previously analysed inputs: Scha 1990, Bod 1998) of Morgan & Demuth's (1996a) "phonological bootstrapping" hypothesis, that infants acquire structured representations of ambient language phonological structure as a first step towards lexicon and syntax; that the ontogenetic origin of syntactic structure lies in phonological structure. The theory is outlined, its motivation from the empirical research literature on the schedule of early phonological and syntactic development is given and some of the challenges ahead for the computational simulation and empirical evaluation of this model are outlined.

## Vocal Babbling

The model comes in two principal parts. The first attempts to explain the origin of the phonological structures by accounting for the development of vocal babbling. This is modeled as the emergence of crossmodal connections (associative connections between particular nodes in pairs of trees in different modalities of cognition) between sets of acoustic and motory exemplars. Both the initial single-modality trees and the later crossmodals emerge as the result of early randomly-generated exemplars (random motor gestures, underpinned by tree-structured representations, random phonological parses of adult and self-generated speech, random node-to-node connections between co-occurrent phonological and motory trees), combined with the principle that, if novel outputs and novel analyses of inputs are always added back into the system's store of exemplars, and always contain new instances of the fragments (subtrees) that were used to construct them, subtrees are replicators; and if more generalizable subtrees are more likely to be re-used, subtrees are Darwinian replicators, under a selection pressure for generalizability. This should allow for a model of babbling to be built displaying the same developmental schedule as infant babbling; starting with an early stage of random productions modulated by the structure of the vocal tract, moving from

there to babbling patterns showing statistical regularities in common with the adult ambient language.

## Early Syntax

The second part models the stages of the emergence of syntax. It is motivated by Jusczyk, Houston and Newsome's (1999) finding that early word-learning is modulated by phonological patterning; the hypothesis is that meanings are first attached to phonological constituents, then as associations between meanings and phonological constituents start to be recognized, areas of parallel tree-structure are constructed to account for them. These areas of parallel structure would initially be isolated, but as the child acquires more experience of the language, areas of parallel structure would come to be "coupled" by being found in collocation with each other in utterances; as these pairings increase in frequency, a phase-transition is predicted, from most clusters being pairs to the rapid emergence of larger clusters, as the system of syntactic representations peels away almost entirely from the phonological representations. I predict that these stages - parallel structures absent or isolated; parallel structures pairing off, pre-phase-transition; and larger clusters, post-phase-transition - correlate to the documented stages of syntactic development - one-word, two-word and multi-word

## Acknowledgments

With thanks to Rens Bod, Tecumseh Fitch, Simon Kirby, Jim Hurford, Barbora Skarabela and Monica Tamariz for helpful comments and discussions. This work is supported by the Principal's Personal Fund of Professor Rens Bod, University of St. Andrews.

## References

- Bod, R. (1998), *Beyond Grammar; An Experience-Based Theory of Language*, Stanford, California: Centre for the Study of Language and Information.
- Jusczyk, P., Houston, D., and Newsome, M., (1999), "The Beginnings of Word-Segmentation in English-Learning Infants". *Cognitive Psychology* 39:159-207.
- Morgan, J. and Demuth, K. (1996) *Signal to syntax: bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Lawrence Erlbaum.
- Scha, R. (1990). "Taaltheorie en Taaltechnologie: Competence en Performance", in Q. de Kort and G. Leerdam (eds.), *Computertoepassingen in de Neerlandistiek*, Almere: Landelijke Vereniging van Neerlandici (LVVN-jaarboek).