

Contextual Bootstrapping for Grammar Learning

Eva H. Mok (emok@icsi.berkeley.edu)

International Computer Science Institute and University of California, Berkeley
1947 Center Street Suite 600, Berkeley, CA 94704 USA

Nancy Chang (nchang@icsi.berkeley.edu)

International Computer Science Institute and University of California, Berkeley
1947 Center Street Suite 600, Berkeley, CA 94704 USA

We present a computational model of grammar learning that combines domain-general learning mechanisms with rich representations of linguistic knowledge, world knowledge and situational and discourse context. These representations support processes of language understanding and inference (including both constructional analysis and reference resolution) that help the learner make sense of utterances in context. The learner then draws on generalization and statistical induction techniques to form new constructions that better capture correlations between linguistically identified and contextually inferred information.

Our work is part of the larger Neural Theory of Language (NTL) project, whose goal is to build models of cognition and language that satisfy convergent constraints from biology, psychology, linguistics and computation (Chang, Feldman, & Narayanan, 2004). We assume that the basic unit of linguistic analysis is a *construction* (form-meaning pair), and further that representations for both linguistic and non-linguistic knowledge are *embodied* (i.e., motivated by sensorimotor experience). These assumptions are captured by the Embodied Construction Grammar formalism (Bergen & Chang, 2005), which serves as the basis for a simulation-based model of language understanding. Our model includes several interleaved processes. *Constructional analysis* takes a sentence and its surrounding context and produces an embodied meaning representation, called a *semantic specification* (Bryant, 2004). *Reference resolution* takes a semantic specification and infers how its various components map to aspects of the current context.

Language learning is driven by the interaction of these analysis processes with more general mechanisms of statistical and rule-based learning (Figure 1). Partial analyses produced by incomplete grammars motivate the hypothesis of new form-meaning mappings to bridge the gap between linguistically derived and contextually available information. Constructions are also reinforced through usage and reorganized based on shared structure to form larger and more general constructions (Chang & Gurevich, 2004).

Recent work has focused on a richer and more tightly integrated model of situational and discourse context that facilitates the learning of context-dependent constructions, as necessary for pro-drop languages such as Mandarin Chinese (Mok & Bryant, 2005). The resulting model is designed to acquire a wide range of grammatical constructions, demonstrating in essence how grammar learning can be bootstrapped by context.

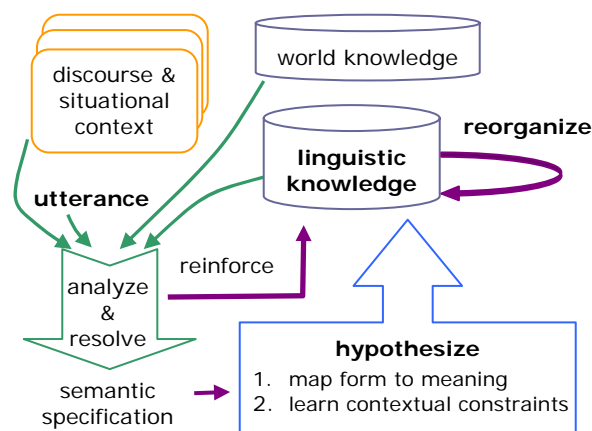


Figure 1. Construction learning model: Each utterance is analyzed using the current grammar and resolved against the current context to produce a semantic specification. New constructions are hypothesized to explain information inferred through resolution, and existing constructions are reorganized via generalization over constructions based on shared constituents or constraints.

References

- Bergen, B. K., & Chang, N. (2005). Embodied Construction Grammar in simulation-based language understanding. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive Groundings and Theoretical Extensions*. Philadelphia, PA: John Benjamins.
- Bryant, J. (2004). *Recovering coherent interpretations using semantic integration of partial parses*. Paper presented at the 3rd workshop on Robust Methods in Analysis of Natural Language Data, Geneva, Switzerland.
- Chang, N., Feldman, J., & Narayanan, S. (2004). *Structured Connectionist Models of Language, Cognition and Action*. Paper presented at the Ninth Neural Computation and Psychology Workshop, Plymouth, UK.
- Chang, N., & Gurevich, O. (2004). *Context-driven construction learning*. Paper presented at the 26th Annual Meeting of the Cognitive Science Society, Chicago, IL.
- Mok, E., & Bryant, J. (2005). *A best-fit approach to productive elision of arguments*. Paper presented at the 32nd Annual Meeting of the Berkeley Linguistics Society, Berkeley, CA.