Learning in Minimalism-based Language Modeling

Deryle Lonsdale (lonz@byu.edu)

Department of Linguistics and English Language, Brigham Young University

4039 JFSB, Provo, UT 84602 USA

Keywords: cognitive modeling; syntax; Minimalist Program; learning.

Background

The natural language version of the Soar cognitive modeling system (Newell, 1990) has enabled a number of language modeling applications from on-line parsing behavior (Lewis, 1993) to simultaneous interpretation (Lonsdale, 1997, 1998) to robotic control (Benjamin, Lonsdale, & Lyons, 2004). The system supports an integrated approach to incremental comprehension and generation. Learning mechanisms account for processes in language performance from deliberate, explicit reasoning to automatic, recognitional expertise.

Syntactic processing in prior versions of the system followed the Principles and Parameters approach to syntax.

Minimalist incremental parsing

This paper introduces new results: how the syntactic model of an incoming utterance is constructed incrementally according to principles of the newer Minimalist Program (Chomsky, 1995). Ambiguities can be resolved subject to certain constraints on backtracking and exhaustivity. Lexical information is based on WordNet-specified properties (Lonsdale & Rytting, 2001). Interesting issues concerning memory usage and processing time arise within this new paradigm versus within the earlier syntactic model.

Semantic interpretation

We also explain how the system is capable of performing incremental semantic model construction of the incoming utterance. The basic representation, a lexical-conceptual semantics (LCS) construction, is mapped step-by-step opportunistically from the syntactic component (Rytting & Lonsdale, 2001). Leveraging more semantic information (e.g. thematic grids from an external knowledge source) has helped constrain and specify licit structures. Interestingly, the newer syntactic model enables better treatment of such semantic phenomena as scope and quantification.

Generation, learning, and bootstrapping

This functionality for syntactic and semantic processing in comprehension also supports incremental natural language generation (Lonsdale, 2001). We map an LCS to a minimalist parse tree, linearize the leaf node lexical content, and buffer the sentence for strategically-timed output.

The system's operator-based architecture and subgoaling assure that syntactic and semantic processing can be learned during performance of the task. Operators that chunk up prior experience in each realm discussed above (syntax and semantics), as well as others (lexical access and discourse comprehension) can be interleaved with each other and indeed with operators associated with other tasks and even other processing modalities (e.g. Lonsdale & Manookin, 2004) to produce interesting accounts of interactive processing. Previously processed utterances produce lexical access, semantic, and syntactic construction operators that can bootstrap related aspects of language generation.

This paper, then, explores the various processing and learning modalities of the new linguistic components, compares the adequacy of the new framework to the pre-minimalist implementation, and sketches future exploration of the new theoretical assumptions.

References

Benjamin, P. & Lonsdale, D. & Lyons, D. (2004). Designing a Robot Cognitive Architecture with Concurrency and Active Perception; *AAAI Fall Symposium: The Intersection of Cognitive Science and Robotics*; American Association for Artificial Intelligence.

Chomsky, N. (1995). The Minimalist Program. MIT Press.

Lewis, R. L. (1993). An Architecturally-based theory of Human Sentence Comprehension. Doctoral dissertation, Carnegie Mellon University.

Lonsdale, D. (1997). Modeling Cognition in SI: Methodological Issues; *International journal of research and practice in interpreting*, 2(1/2):91-117; John Benjamins Publishing Company, Amsterdam.

Lonsdale, D. (1998) An architectural account of simultaneous interpretation. PhD thesis, Carnegie Mellon.

Lonsdale, D. (2001). An Operator-based Integration of Comprehension and Production; *LACUS Forum,* XXV; Linguistic Association of Canada & the United States.

Lonsdale, D. & and Manookin, M. (2004). Combining learning approaches for incremental on-line parsing. *Proceedings of the 6th International Conference on Cognitive Modeling* (pp. 160-165). Lawrence Erlbaum.

Lonsdale, D. & Rytting, C.A. (2001). Integrating WordNet with NL-Soar; *WordNet and other lexical resources: Applications, extensions, and customizations*; Proceedings of NAACL-2001; Association for Computational Linguistics; pp. 162-164.

Newell, A. (1990). Unified Theories of Cognition. Harvard University Press.

Rytting. C.A & Lonsdale, D. (2001). An Operator-based Account of Semantic Processing; *The Acquisition and Representation of Word Meaning*; European Summer School for Logic, Language, and Information; pp. 84-92.