Incrementality, prediction, and attention in a scaleable network model of linguistic competence and performance.

Marshall Mayberry, Matthew W. Crocker
martym,crocker@coli.uni-sb.de
Saarland University

An ever-increasing body of experimental evidence from psycholinguistic studies indicates that people not only interpret sentences incrementally, dynamically revising their interpretations as they encounter new information, but also that preferred interpretations generate expectations about what is to follow. It has also been established that the human processor has the ability to integrate diverse sources of information, including prosody, syntax and semantics, frequency, discourse, and even from visual scenes containing objects and events.

When taken collectively, these hallmarks of linguistic performance - incremental, dynamical, probabilistic, integrative, and predictive - have led many researchers to explore subsymbolic models of sentence comprehension. Because subsymbolic systems automatically develop distributed representations according to soft constraints, they have been successfully applied to cognitive phenomena for which more data exists than theory. Yet such systems have proven to be very difficult to scale up to realistic levels of linguistic coverage and complexity. Furthermore, the very nature of distributed representations makes it difficult to ascertain whether such models are achieving adequate linguistic competence, let alone the precise form that competence takes.

In this paper we present a network architecture for incremental sentence comprehension that is both more transparent and scaleable, yet also broadly exhibits incremental ambiguity resolution behaviour that is still cognitively plausible. The model is based on a simple recurrent network, but generates explicit semantic representations of input sentences. The network was trained on hand-annotated Minimal Recursion Semantics (MRS) dependency graphs of some 5000 sentences from the recently-released LinGO Redwoods HPSG Treebank (Oepen, Flickinger, Toutanova, & Manning, 2002), and was able to accurately learn to incrementally develop and revise such deep semantic representations. Examination of the network revealed that it also could maintain several alternative interpretations simultaneously, pruning away those which were untenable. When tested on the original VerbMobil sentences from which the Redwoods Treebank corpus was transcribed, the model demonstrated robustness to many of the speech errors, repairs, and dysfluencies of those original sentences. Finally, when trained and tested on a variant of the McClelland and Kawamoto data on the prepositional phrase attachment ambiguity, the model was able to account for the data as well as previous subsymbolic models which had been crafted to model that data only.

We further report recent findings on extending the architecture to modelling language comprehension in context. For example, the revised model permits semantic representations of visual scenes to be input concurrently with the incremental processing of words from a related utterance. Our simulations indicate that the model is able to exploit scene information in a manner suggested by recent visual world studies (Kamide, Scheepers, & Altmann, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2003) to anticipate and resolve an ambiguous sentence-initial noun phrase in favor of the role played by the corresponding character in the scene. A further extension consolidates several experimental results into a single network that directly maps highly-active semantic representations (typically recently encountered or anticipated material) to visually depicted entities and events. Preliminary results suggest this mapping may permit more precise modeling of attention in visual scenes in response to linguistic stimuli, as reflected by visual world experiments. Such a model should help to tease apart the influence of short-term contextual effects of the immediate visual environment from the long-term empirical role of language experience reflected by the models training.

References

Kamide, Y. Scheepers, C., & Altmann, G.T.M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-Linguistic Evidence from German and English. Journal of Psycholinguistic Research. Vol 32(1), pp. 37-55.

Knoeferle, P., Crocker, M.W., Scheepers, C., & Pickering, M.J. (2003). Actions and Roles: using depicted events for disambiguation and reinterpretation in German and English. Proceedings of the 25th Annual Cognitive Science Conference, Boston, Mass.

Oepen, S., Flickinger, D., Toutanova, K., & Manning, C. (2002). LinGO Redwoods: A rich and dynamic treebank for HPSG. In "Beyond PARSEVAL. Workshop of the Third LREC Conference."