Syntactic representations as side-effects of a sensorimotor mechanism

Alistair Knott, Dept of Computer Science, University of Otago
alik@cs.otago.ac.nz

This paper describes a computational model of modern natural language syntax in which syntactic structures are defined as descriptions or traces of sensorimotor operations. The syntactic structure of a sentence describing a concrete state or event (e.g. *There is a dog in the garden* or *The man grabbed a cup*) is characterised as a trace of the sensorimotor processes which occur in an agent who directly witnesses it, by observing it or participating in it. The model of syntax is therefore closely grounded in a model of sensorimotor cognition. The model is relevant to work on language evolution, because it provides one way of fleshing out an account of how a language faculty could have evolved as a genetic adaptation of pre-existing sensorimotor capacities.

My system performs a task similar to the L_0 task (Feldman *et al.*, 1996). An agent with perceptual capabilities is given a set of situations to observe, each accompanied by a sentence describing it, and must learn to generate appropriate sentences for similar situations. The agent processes each situation using a sensorimotor mechanism consisting of several different interacting components. Each of these components generates a side-effect of its operation at a linguistic level of representation. The architecture of the perceptual mechanism imposes a (partial) order on these side-effects, which can be construed as encoding certain aspects of the syntactic representation of the sentence to be expressed.

The syntactic framework adopted in the model is a version of Government-Binding (GB) theory (Chomsky, 1981), as extended by Pollock (1989) and Koopman and Sportiche (1991). In this theory, sentences have a deep syntactic structure (DS), from which a surface structure (SS) is derived by movement operations. In my model, the DS of a sentence is directly encoded by the side-effects of sensorimotor operations. Movement operations between DS and SS are (partly) unconstrained; agents have to learn appropriate mappings from exposure to training sentences, using a recurrent neural network architecture similar to that given by Chang (2002).

The sensorimotor model is an integration of several recent biologically-inspired models of object and action recognition (Riesenhuber and Poggio, 1999; Giese, 2000), visual attention (Itti and Koch, 2000) and motor control (Wolpert and Kawato, 1998). It focuses on the idea that the perception of a 'sentence-sized event' involves a sequence of transitions between different attentional states, each of which generates a distinctive side-effect in a medium for assembling linguistic representations. For instance, the perceptual process underlying the sentence *The man grabbed the cup* begins with an action of re-attention to an object already encountered, implemented as the reactivation of an existing object file for the man in question (Kahneman *et al.*, 1992). This then triggers (in parallel) a representation of the man's local environment in a frame of reference centred on the man, and a mechanism for biological motion detection. These two events in turn jointly trigger identification of the action and identification of the target object. The DS structure of this sentence can be reinterpreted quite neatly as an encoding this partially ordered sequence of sensorimotor operations. Each syntactic position in the DS tree denotes one operation. The basic right-branching structure of the tree encodes the sequential dependencies between operations. For instance, subject ([Spec,IP]) position denotes the initial action of attention of the sentence, which precedes the creation of an agent-centred frame of reference, denoted by the structurally lower $\mathrm{Agr_O P}$ position. The possibility of constituent movement arises from partial orderings between operations. For instance, the operation denoted by [Spec,IP] is also denoted by the VP-internal subject position [Spec,VP], because it also has a role in triggering the biological motion recognition system, which is denoted by V. Whether the subject NP appears at [Spec,IP] or [Spec,VP] in SS structure is a learnable parameter.

In summary, the paper presents a novel interpretation of Chomsky's distinction between DS and SS structures, characterising DS as a direct encoding of sensorimotor processing, and the transition from DS to SS as a learnable mapping from this encoding to surface word order.

References

Chang, F. (2002). Symbolically speaking: a connectionist model of sentence production. *Cognitive Science*, **26**, 609–651.

Chomsky, N. (1981). *Lectures on government and binding*. Foris, Dordrecht.

Feldman, J., Lakoff, G., Bailey, D., Narayanan, S., Regier, T., and Stolcke, A. (1996). L_0: The first five years of an automated language acquisition project. *Artificial Intelligence Review*, **10**(1–2), 103–129.

Giese, M. (2000). Neural model for the recognition of biological motion. In G. Baratoff and H. Neumann, editors, *Dynamische Perzeption*, pages 105–110. Infix Verlag, Berlin.

Itti, L. and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, **40**(10–12), 1489–1506.

Kahneman, D., Treisman, A., and Gibbs, B. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, **24**, 175–219.

Koopman, H. and Sportiche, D. (1991). The position of subjects. *Lingua*, **85**, 211–258.

Pollock, J.-Y. (1989). Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry*, **20**(3), 365–424.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, **2**, 1019–1025.

Wolpert, D. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, **11**, 1317–1329.