A learnable version of set theoretic merge

Alexander Clark, Royal Holloway, University of London

Learnability does not occupy the central role in biolonguistics that it does in the principles and parameters program, but the problem is arguably more acute. Positing a rich language faculty is evolutionarily implausible, and thus the burden of learnability must fall more sharply on domain-general learning mechanisms or on theird factor principles. In this context, in order to understand what merge is we need to know how it is acquired.

As Chomsky defines it, MERGE is the name given to the primitive computational procedure that combines two syntactic elements together to make a larger unit. If this is unconstrained, then it is vacuous. Standardly it is assumed that there is a finite set of features that limit this mechanism. The exact nature of these features is not clear – are they part of UG or acquired? Lexicalising a grammatical formalism does not solve the acquisition problem: the acquisition of the lexicon – the mapping of words to features – then becomes the key problem.

In the modern theory of distributional learning [3, 1] we now have a plausible idea of how MERGE could be acquired. The syntactic objects, X, correspond merely to sets of substrings $S(X)$ – the set of all strings that are generated by that synatctic object. The abstract, and unlearnable, syntactic features are replaced by the distributional lattice structure of the sets of strings that is observable and this learnable. Recently [2] we have a constraint on which merges are trivial and which are not based on whether the resulting object is "greater" than the sum of the parts. This gives rise to a natural notion of hierarchical syntactic structure by limiting merge to those non-vacuous combinations. Given two syntactic objects X and Y and a merge operation $\circ$ we say that the merge is non vacous if $S(X \circ Y) \supsetneq S(X)S(Y)$: i.e. if the inclusion is strict. This "minimal" assumption serves to define a notion of syntactic structure that is objective in the sense that it is based on the structure of the language, and is thus learnable using modern techniques.

This schema is applicable both to simple merge, where $S(X)$ and $S(Y)$ are sets of strings, and to cases where they are sets of tuples of strings, which gives rise to a notion of internal merge similar to that in minimalist grammars [5] and multiple context free grammars [4]. This allows a unified treatment of displaced constituents and cross-serial dependencies, while limiting the size and complexity of UG to a minimum.

References

[1] Alexander Clark. Towards general algorithms for grammatical inference. In *Proceedings of ALT*, pages 11–30. Springer, Canberra, Australia, October 2010. Invited Paper.

[2] Alexander Clark. A language theoretic approach to syntactic structure. In *Proceedings of the 12th Meeting on the Mathematics of Language (MOL)*, Nara, Japan, 2011. to appear.

[3] Alexander Clark and Rémi Eyraud. Polynomial identification in the limit of substitutable context-free languages. *Journal of Machine Learning Research*, 8:1725–1745, August 2007.

[4] H. Seki, T. Matsumura, M. Fujii, and T. Kasami. On multiple context-free grammars. *Theoretical Computer Science*, 88(2):229, 1991.

[5] E. Stabler. Derivational minimalism. *Logical aspects of computational linguistics*, pages 68–95, 1997.