Rules and Exemplars in Language Acquisition

Rens Bod (rens@science.uva.nl)
Gideon Borensztajn (gideon@science.uva.nl)
Institute for Logic, Language and Computation
University of Amsterdam, Netherlands

Daniel Freudenthal[a]
(d.freudenthal@liverpool.ac.uk)
Julian Pine[a] (julian.pine@liverpool.ac.uk)
Fernand Gobet[b] (fernand.gobet@brunel.ac.uk)
[a]School of Psychology, University of Liverpool, UK
[b]School of Social Sciences, Brunel University, UK

Carla L. Hudson Kam (clhudson@berkeley.edu)
Department of Psychology, University of California, Berkeley, USA

Alexander Clark (alexc@cs.rhul.ac.uk)
Department of Computer Science, Royal Holloway University of London, UK

Willam G. Sakas (sakas@hunter.cuny.edu)
Department of Computer Science, Hunter College
The Graduate Center, City University of New York

Keywords: language acquisition, computational modeling rules, exemplars

Goals and Scope

There are two main views about the nature of language acquisition. Broadly put, the *nativist* view endorses that human language acquisition is guided by abstract innate rules ("Universal Grammar"), while the *empiricist* view assumes that language acquisition is the product of abstractions from stored exemplars. Despite the apparent opposition between these two views, the essence of the debate seems to lie in the relative contribution of abstract prior knowledge and concrete linguistic experience in learning a language (see Pullum and Scholz 2002; Yang 2004).

One of the major goals for computational models of language acquisition is then to establish the minimal prior knowledge needed for language acquisition to take place. Yet there is surprisingly little agreement on what this minimal knowledge should be: it ranges from a simple chunking mechanism (as in Freudenthal et al. 2007) to rather complex grammar-induction procedures (Clark and Eyraud 2007) up to the assumption of full-fledged constituent structure (Bod 2008).

This symposium aims to discuss differences and convergences across a number of representative models of language acquisition in the context of recent results in human language learning (Hudson Kam et al. 2005). In particular, we want to contribute to a better understanding of the interplay between prior knowledge and linguistic experience in modeling language acquisition. We will compare rule-like and probabilistic behavior in language acquisition models with respect to a variety of phenomena: optional infinitives (Freudenthal et al.), auxiliary fronting (Clark; Bod and Borensztajn) and alternations (Clark). We will discuss how far probabilistic tendencies may lead to categorical behavior (Hudson Kam; Bod and Borensztajn), and to what extent distributional properties of the input language can overcome the need for abstract knowledge in dealing with complex facets such as long-distance and cross-serial dependencies (Freudenthal et al.; Clark; Bod and Borensztajn).

Speakers

Daniel Freudenthal, Julian Pine and Fernand Gobet

Simulating Language Acquisition in MOSAIC

Many computational models of language acquisition focus on showing that it is possible to learn certain aspects of language that have been identified as problematic for general purpose learning mechanisms. That is, many modellers focus on solving particular learnability problems. In this talk we focus on a different approach, implemented in MOSAIC (Model of Syntax Acquisition in Children). Key features of this approach include: learning from realistic, child-directed speech, an emphasis on simulating cross-linguistic data using one, identical model, and a commitment to simulating actual corpora of child utterances. These features make it possible to analyse the output from one model with respect to several, seemingly unrelated phenomena and to investigate how children's early speech is shaped by the interaction between common processing constraints and the distributional properties of the input language. This has allowed us to show that several key phenomena in child speech and their cross-linguistic patterning can be understood in terms of the distributional statistics of the input read through an utterance-final bias in learning without the need to represent abstract rules over grammatical categories.

While MOSAIC's lack of abstract knowledge allows it to identify those areas of the data where abstract knowledge is not required, it does limit its ability to deal with constructions like long-distance dependencies. Possible ways of representing such dependencies are explored through the notions of frames and chunking in the substitution of distributionally similar words.

Carla L. Hudson Kam

Whence rules?

When children learn language from non-native speakers, as in cases of the emergence of new contact languages, they impose consistency on input which contains only probabilistic grammatical tendencies. This rapid emergence of consistent grammatical structure languages is often used

as evidence for innate knowledge of language, either rules or structures of particular types, or a more general constraint on language structures as being consistent and rule-governed. However, this is not the only possible explanation for the imposition of consistency or regularization by learners. In several experiments we exposed children and adults to language input containing probabilistic grammatical tendencies characteristic of non-native speakers and assessed learning using production as well as judgment measures. Results from production show that generally, children seem to impose consistent rules on the language. Adults, however, sometimes replicate the probabilistic grammatical tendencies, but when complex enough, adults too will speak the language in a way that is more consistent than their input. However, judgment measures show that both children and adults who regularize are sensitive to the variation present in their input, despite their productions. Thus, the rule-like nature of the productions may stem from aspects of the production process itself, rather than from the imposition of rule-like representations. In this, these cases may provide less evidence for the existence of a formal or structural universal guiding acquisition (resulting in rules) than previously thought.

Alexander Clark

Hierarchical and non-hierarchical models of language

Computational and theoretical analysis of the problem of language acquisition can help cognitive science by providing a clear theoretical understanding of the minimal prior knowledge needed for language acquisition to take place. This minimal prior knowledge is generally assumed to include a bias for hierarchically structured representations. However, the notion of hierarchical structure is never made clear and often confuses several distinct properties: being context free, not being regular, using tree structures or having transformational rules that are sensitive to tree structure. Here we will discuss some of these ideas and present two arguments that call the assumption of hierarchical structure into question.

Prompted by learnability considerations we have developed provably correct algorithms for grammatical inference of context free languages from positive data based on the distributional approach of Zellig Harris, which use associative rather than hierarchical structure. We show that such algorithms can learn from an artificial data set the correct rule for polar interrogatives with embedded relative clauses. We also discuss the extent to which such non-hierarchical models are needed to overcome the limitations of constituent structure as a representational formalism. In particular we will examine phenomena such as free word order, cross-serial dependencies, bracketing paradoxes in morphology, certain forms of ellipsis and local phonological effects such as the a/an alternation in English, and argue that such phenomena are strong evidence for the use of models that don't rely on a strict idea of hierarchical structure.

Rens Bod and Gideon Borensztajn

From Exemplar to Grammar

While rules and exemplars are usually viewed as opposites, the data-oriented parsing (DOP) approach argues that they are end points of the same distribution. By representing both rules and exemplars as (partial) trees, DOP takes into account the fluid middle ground between the two extremes. This insight is the starting point for a new theory of language learning, termed U-DOP, which is based on the following idea: if a language learner does not know which phrase-structure tree should be assigned to a sentence, s/he initially allows implicitly for all possible trees and lets linguistic experience which is the 'best' tree for each sentence. The best tree is obtained by computing the most probable shortest derivation from the frequencies of subtrees in the set of all possible (finite) trees of previous sentences. Thus U-DOP's only prior knowledge is the existence of constituent structure, and lets statistics do the rest.

By having learned the syntactic structures of sentences, we have also learned the grammar implicit in these structures, which can in turn be used to produce new sentences. We show that our model mimicks children's language development from item-based constructions ('holophrases') to abstract constructions ('rules'), and that the model can simulate some of the errors made by children in producing complex questions. It turns out that U-DOP can learn the abstract rules for complex syntactic phenomena, such as agreement and auxiliary fronting, by statistical generalization over Eve's utterances in the Childes database. Finally, we will discuss in how far U-DOP's statistical tendencies result in categorical behavior in the course of language learning.

Commentator: Willam G. Sakas

References

Bod, R. (2008). From exemplar to grammar: Integrating analogy and probability in language learning. Submitted ms (http://staff.science.uva.nl/~rens/analogy.pdf).

Clark, A. & Eyraud, R. (2007). Polynomial Identification in the Limit of Substitutable Context-free Languages, *Journal of Machine Learning Research* , 8, 1725—1745.

Freudenthal, D. Pine, J., Aguado-Orea, J. & Gobet, F. (2007). Modelling the developmental patterning of finiteness marking in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science*, 31, 311-341.

Hudson Kam, C., & Newport, E. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, 1, 151-195.

Pullum, G. and Scholz B. (2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19, 9-50.

Yang, C. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences* 8(10), 451-456.