

Computational Models of First Language Acquisition Special Issue of Research on Language and Computation

Alexander Clark · William Gregory Sakas

© Springer Science+Business Media B.V. 2011

1 Introduction

How children acquire language has stood, for a long time, as one of the most fundamental, beguiling, and surprisingly open questions of modern science. Recent advances in natural language processing, statistical parsing and machine learning, together with the availability of large corpora of child directed speech and other corpora, make a wide range of computationally-oriented approaches to the study of this problem available. Given a model of some aspect of language acquisition, implementing it as a computational system and evaluating it on naturally occurring corpora has a number of compelling advantages. First of all by implementing the system, we can be sure that the algorithm is fully specified, and the acquisition model does not resort to hand-waving at crucial points. Secondly, by evaluating it on real linguistic data, we can see whether naturally occurring distributions of examples in corpora provide sufficient information to support the studied claims across a divergent range of acquisition theories. Thirdly, study of the system can identify the mechanisms that cause changes in the algorithm's hypotheses during the course of acquisition. Finally, the computational resources required of the model can be concretely assessed and (not so concretely) compared against the resources that might be available to a human language learner.

A. Clark
Department of Computer Science, Royal Holloway, University of London, London, UK
e-mail: alexc@cs.rhul.ac.uk

W. G. Sakas (✉)
Department of Computer Science, Hunter College, PhD Programs in Linguistics and Computer Science,
The Graduate Center, The City University of New York (CUNY), New York, NY, USA
e-mail: sakas@hunter.cuny.edu

It is important not to lose sight of the limitations of these approaches as well. Computational models intrinsically necessitate simplifications of the linguistic and extra-linguistic environments that exist in natural contexts. The child learner has access to a vast range of information—not just the sentences he or she is exposed to, but prosodic contour and situational context, which includes caregiver gaze and numerous other sources of information. Furthermore the child is not passive. The child has the opportunity of interacting with the environment in a number of different ways, both linguistically and through gestures and actions of various types. The extent to which these factors play a role during language acquisition is unsettled and under investigation. As modeling endeavors continue to increase in complexity, putting to use any number of them will require an increasingly important dialogue between computational modelers, psycholinguists and those responsible for gathering and annotating data.

1.1 Proof Versus Experiment

The earliest computationally-oriented studies of language acquisition (Chomsky and Miller 1963; Chomsky 1963) centered on the formalization of the properties of languages and the grammars that generate them. Gold's groundbreaking work (Gold 1967) established the field of *learnability* whose goal is to determine what is learnable in principle. Research in this field primarily makes use of mathematical proof to show which classes of grammars can (or more often, can not) be learned in principle given constraints on what the learner is exposed to. Learnability research typically establishes very broad bounds on what is learnable since the learning paradigms may afford the learner unlimited resources, and consider large classes of languages. While the computational resources attributable to the child learner, and the precise characterisation of a possible human language are yet to be established, learnability research is a reasonable approach and the results yield some insight.

However, just as with computational modeling, there are several shortcomings of establishing learnability through proof. Typically the sample complexity bounds derived from probabilistic models may be very loose, and the probabilistic inequalities hard to deal with (Clark and Thollard 2004). Moreover, some models of acquisition are too complex to readily lend themselves to proof techniques (e.g., connectionist models), and learnability proofs typically don't give insight into the intermediate stages of acquisition. Finally, the risk is very real that the models will be driven by their ease of analysis rather than their relevance or faithfulness to the empirical facts: in order to prove a convergence theorem about an algorithm, one may need to make certain simplifying assumptions that are driven by the requirements of the theorem rather than independently motivated by an analysis of the human learning situation.

Computational modeling is a significant departure from learnability research, and though both have their place, simulation modeling has enjoyed increased attention of language acquisition researchers for the reasons outlined above.

1.2 Syntax (or not)

This special issue was intended to gather together recent approaches to the computational modeling of various aspects of language acquisition. The papers selected

examine this from a variety of different perspectives, and look at various different tasks including word segmentation, verb subcategorisation frame learning, morphology learning and a particular type of learning from acoustic data.

Interestingly none of the papers here deal with computational models of the acquisition of syntax *per se*, a field which has traditionally been considered the locus of the sharpest problems in learnability. This perhaps reflects two factors. The first is the fact that while there are plenty of algorithms that can learn finite state models, the theory of the learning of context free grammars and other non finite state models is comparatively less developed (notwithstanding recent developments in distributional learning [Clark and Eyraud 2007](#)). The second is that there is less of a consensus on the data, and the annotations of that data. There is no argument, at least in English, about the status of inflectional paradigms, or about segmentation apart from perhaps the status of some clitics. The situation in syntax is radically different: there is little consensus about syntactic structure, nor even agreement about whether constituent structure as opposed to dependency structure is the appropriate representation. There are a host of other disagreements both in minor details within particular communities and concerning fundamental representational decisions between different communities. The *de facto* standard for computational linguistics in English of the Penn Treebank annotations is not linguistically defensible.

In spite of these problems, research in the acquisition of syntax continues. There is very active research in the computational linguistics community on unsupervised learning of dependency structures ([Klein and Manning 2004](#); [Gillenwater et al. 2011](#)). Though these models are motivated primarily by engineering concerns rather than the modeling of language acquisition, as evidenced by the use, typically, of newswire corpora rather than more realistic corpora of child directed speech. More linguistically oriented research continues in a number of different paradigms, including the principles and parameters program ([Yang 2002](#); [Fodor and Sakas 2004](#); [Sakas and Fodor, to appear](#)), as well as work in more empiricist frameworks, for example [Alishahi and Stevenson \(2008\)](#). Still, given the paradigmatic obstacles noted earlier, much of the work focuses on specific subproblems of the acquisition of syntax; we note with some regret that this issue does not include an article in this area.

2 Contents

The special issue consists of four papers.

The first paper is “Online Learning Mechanisms for Bayesian Models of Word Segmentation” by Lisa Pearl, Sharon Goldwater and Mark Steyver (PGS). This is an examination, within a Bayesian paradigm, of word segmentation: in many models this is one of the first tasks that a child learner must carry out. Word segmentation is by now a well studied task—the novelty of this paper is that it makes more cognitively realistic assumptions about the computational limitations on the learner. Standard methods in the field tend to work in a batch mode: the learner stores the whole data set in memory and operates globally on this data set, typically trying to optimise some quantity defined over the whole set. This, from a modeling point of view is convenient but it is highly unrealistic from a cognitive viewpoint. PGS compare this “ideal” learner, with

learners that are constrained to operate only on a fraction of the data at any one time. They contrast an ideal learner, that uses Gibbs sampling over the entire corpus with various online learners that only process one data point at a time, or have only a limited recall for recent utterances. One of the conclusions from this work is that this processing limitation does not significantly affect the overall performance of the learner. These limited learners still outperform the naive transitional probability learners that are considered to be cognitively more plausible.

This paper addresses an interesting methodological point. The problem of how children go about segmenting a stream of sounds into words is certainly not yet solved but a range of different computational models exist each of which achieve high accuracy. Once a threshold of accuracy is passed, emphasis switches to matching the developmental trajectory of acquisition, and cognitive limitations; a point at which the use of additional data sources becomes more important. No longer can a pure modeling perspective be taken—evaluating modeling performance becomes much more tied to research in psycho- and neurolinguistics. We return to this point briefly in the discussion section below.

The second paper is “A computational model of unsupervised speech segmentation for correspondence learning” by Hinrich Schütze, Daniel Duran, Michael Walsh and Bernd Möbius. This paper addresses a novel problem that the authors call the *correspondence problem*. In an early stage of language acquisition the child must learn both that different acoustic sequences must correspond to the same phonological object, and how to relate the sequence of articulations that the child makes to the sequence of sounds that it hears: in short, to learn the correspondence between speech production and speech perception. Studying this problem requires working on raw acoustic data—a particularly difficult challenge. The authors argue for the importance of this task and propose a hypothesis about how this could be performed that they call the Correspondence-by-Segmentation Hypothesis (CSH). This hypothesis is, roughly speaking, that the speech is segmented first and then aligned. The paper presents some computational experiments that support the CSH.

“Learning the raising control distinction” by William Garrett Mitchener and Misha Becker (M&B) looks at the acquisition of raising and control verbs in English. The acquisition problem here is caused by the ambiguity both of sentential frames which may fail to distinguish between raising and control verbs, and ambiguity of the verbs themselves which may, such as in the case of “begin”, be used either in control or raising constructions. M&B conduct a careful study looking at two particular semantic cues, animacy of the subject and eventivity of the predicate, which have been shown to be helpful for learners in direct psycholinguistic experiments. M&B experiment with various learners on a corpus of data annotated with semantic cues. Rather than a subset strategy, they conclude that various learners including a Bayesian learner were capable of acquiring these distinctions, and offer arguments that the learners are biologically plausible.

Finally, “Investigating the Relationship Between Linguistic Representation and Computation Through an Unsupervised Model of Human Morphology Learning” by Erwin Chan and Constantine Lignos (C&L) is a study of the acquisition of morphology. This has been a particularly active area of research since Goldsmith’s influential

LINGUISTICA model (Goldsmith 2001). The model focuses on trying to match the course of development which seems to be cross linguistically stable.

This work is characterised by an attention to subtle properties of the data—the learner exploits a Zipfian distribution in the data, but without explicitly modeling it. The learner uses a greedy iterative algorithm: again rather than trying to find a global optimum which can be computationally expensive, the learner makes a series of decisions based purely on local information. This has two benefits—it reduces the computational cost and makes the storage more cognitively plausible. The paper also addresses one of the most subtle and difficult issues in language acquisition—the relationship between learnability and linguistic representation. C&L argue that a type of morphophonological rewrite rule is important: in other words that models must be sufficiently expressive to account for acquisition.

3 Discussion

Some common themes emerge from these papers. First, there seems to be a gap opening between two classes of sub-problems in the computational study of language acquisition. The first class consists of those problems where there is a range of adequate learning algorithms, and attention is now shifting to selecting which algorithms will more closely mimic either the developmental trajectory of the infant learner, or human computational limitations, either in time or space. This line of research is inherently interdisciplinary and as noted above will require an increasing collaboration between computational and cognitive science researchers. With the exception of the article by Schütze et al., the papers in this issue fall in this category.

The second class of problems, exemplified by syntax, is where satisfactory algorithms have yet to be found. In this case, research focuses on finding suitable algorithms regardless of their cognitive implausibility, and with a blind eye towards psycholinguistic or developmental data. This seems a reasonable strategy—data can be used to choose *between* theories, but where there are no theories that meet the most fundamental boundary conditions, data will not help. The principles and parameters framework, for example, is best viewed as a meta-theory rather than a fully specified theory of grammar.

In its stead computational study of so-called “ideal learners” (Chater and Vitányi 2007) or other comparatively unconstrained algorithms, can help shed light on what is learnable in principle admitting a wide range of model parameters, within which it might be hoped that cognitive resources, and the domain of human language lie.

Another is a point almost too obvious to be worth noting: the central use of probabilistic methods in all of the papers in this special issue. From the standpoint of computational linguistics and natural language processing, this has been completely standard for over 20 years. The increased use of Bayesian methods is perhaps noteworthy, as these techniques, well established in cognitive science, are more frequently being put to task on linguistic problems. But from a linguistics point of view, the use of the probabilistic paradigm is still controversial. Though Yang (Yang 2002) has argued for the importance of information theoretic methods in limited circumstances, and increasing numbers of linguists are using such methods at least in part (Bresnan

2007), it is noticeable that for example, a recent handbook surveying part of the field (Boeckx 2011) does not even contain an index entry for “probability”, and indeed the only discussion in the volume of probabilistic methods is by Yang. Information theoretic methods and mainstream generative linguistics still seem as far apart as ever (Pereira 2000).

Acknowledgments We would like to thank all of the authors who submitted to the special issue, the reviewers and Shuly Wintner for suggesting the topic and dealing with various issues.

References

- Alishahi, A., & Stevenson, S. (2008). A computational model of early argument structure acquisition. *Cognitive Science*, 32(5), 789–834.
- Boeckx, C. (Ed.). (2011). *The Oxford handbook of linguistic minimalism*. Oxford (New York): Oxford University Press.
- Bresnan, J. (2007). Is syntactic knowledge probabilistic? experiments with the English dative alternation. In S. Featherston & W. Sternefeld (Eds.), *Roots: Linguistics in search of its evidential base* (pp. 75–96). Berlin: Mouton de Gruyter.
- Chater, N., & Vitányi, P. (2007). ‘Ideal learning’ of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology*, 51(3), 135–163.
- Chomsky, N. (1963). Formal properties of grammars. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 2 (pp. 323–418). New York: Wiley.
- Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 2 (pp. 269–321). New York: Wiley.
- Clark, A., & Eyraud, R. (2007). Polynomial identification in the limit of substitutable context-free languages. *Journal of Machine Learning Research*, 8, 1725–1745.
- Clark, A., & Thollard, F. (2004). PAC -learnability of probabilistic deterministic finite state automata. *Journal of Machine Learning Research*, 5, 473–497.
- Fodor, J. D., & Sakas, W. G. (2004). Evaluating models of parameter setting. In *Boston University Conference on Language Development* (pp. 1–27). Cascadilla Press.
- Gillenwater, J., Ganchev, K., Graça, J., Pereira, F., & Taskar, B. (2011). Posterior sparsity in unsupervised dependency parsing. *The Journal of Machine Learning Research*, 12, 455–490.
- Gold, E. M. (1967). Language identification in the limit. *Information and Control*, 10, 447–474.
- Goldsmith, J. A. (2001). Unsupervised learning of the morphology of a natural language. *Computational Linguistics*, 27(2), 153–198.
- Klein, D., & Manning, C. (2004). Corpus-based induction of syntactic structure: Models of dependency and constituency. In *Proceedings of the 42nd Annual Meeting of the ACL*.
- Pereira, F. (2000). Formal grammar and information theory: Together again? *Philosophical Transactions of the Royal Society Series A*, 358, 1239–1253.
- Sakas, W. G., & Fodor, J. D. (to appear). *Language Acquisition*.
- Yang, C. (2002). *Knowledge and learning in natural language*. Oxford (New York): Oxford University Press.