A THEORY OF LANGUAGE LEARNING

Sadriddinova Mashhura Oktam's girl
Navoi State Pedagogical Institute Foreign language and literature specialist, English language and literature faculty english teacher

Annotation: *A theory of language learning is described, which uses Bayesian induction of feature structures (scripts) and script functions. Each word sense in a language is mentally represented by an m-script, a script function which embodies all the syntax and semantics of the word. M-scripts form a fully-lexicalised unification grammar, which can support adult language. Each word m-script can be learnt robustly from about six learning examples. The theory has been implemented as a computer model, which can bootstrap-learn a language from zero vocabulary.*
Key words: *language, theory, assumption, framework, cognitive linguistic, mechanism*

The theory is in good agreement with many key facts of language acquisition, including facts which are problematic for other theories. It is compared with over 100 key cross-linguistic findings about acquisition of the lexicon, phrase structure, morphology, complementation and control, auxiliaries, verb argument structures, gaps and movement - in nearly all cases giving unforced agreement without extra assumptions. This paper was written in 1996, and has been available in various places on the Internet since then – for instance. I am now posting the paper with no changes except for re-formatting, for the following reason: This paper shows how in any complex domain of cognition (such as language, navigation or 3-D spatial cognition), where knowledge can be represented in feature structures, it is possible to learn those feature structures from a few learning examples each – essentially as fast as is permitted by a Bayesian speed limit on learning. This result is now important because, after more than thirty years of research into neural nets, they are still not capable of fast learning in complex domains. It is time for renewed investigation of computational models which can learn rapidly, as animals and people do.

Some of the notation, notably the use of the term 'Trump links' now seems unfortunate. This term was motivated by the idea of trump cards in Bridge, and has not been changed.

1. Introduction
2. Language Processing
3. Language Learning
4. The Evolution of Mind and Language
5. Comparisons with Observation

1. Introduction

Language acquisition has been high on the agenda of cognitive science for forty years, shaping theoretical linguistics and informing many child language studies. There are many theoretical ideas about parts of the acquisition process. Yet there are very few fullyarticulated, workable models of first language learning — and even fewer have been compared with a wide range of data. This paper describes a new broad-scope theory of language learning, with the following features:

- rapid, robust learning of language from unreliable and noisy data
- most features of adult language acquired by a single learning mechanism
- integrated learning of syntax, semantics and segmentation
- a working computational model, which can bootstrap to learn language from zero vocabulary
- a firm mathematical basis, linking linguistic theory and learning theory
- an evolutionary account of the origins of language and language learning
- good agreement with a lot of data on child language learning

The last point is the most important. I have compared the theory with over 100 key facts about language acquisition, finding good, unforced agreement in the vast majority of cases. In many cases the theory can give a clear, crisp account of facts which are puzzling in most current language learning theories. These comparisons are described in section 5. I believe no other theory of language learning can claim such broad agreement with the facts.

The theory of language learning is part of a larger theory of language evolution, learning and performance. I need to describe other aspects of this theory for three reasons:

1. To establish that what is learnt really is language, and not just a toy subset;
2. To motivate aspects of the learning mechanism.
3. Because understanding and production are the only windows to measure children's language learning

These descriptions of the non-learning aspects are kept as short as possible, to keep the focus on the learning theory (section 3), and the comparisons with data (section 5).

Theories of language learning have been polarised between two camps:

- Chomskian theories, in which the abstract structures of adult language are acquired by innate language-specific mechanisms.
- Broader frameworks such as cognitive linguistics, in which general social/cognitive mechanisms are used to learn language in stages of development.

This theory does not fit neatly in either camp. It does not posit de novo language-specific structures or learning mechanisms in the brain, nor does it rely on broad ill-defined learning mechanisms; it proposes a Bayesian learning mechanism, evolved specifically for primate social intelligence and extended for language, with a precise mathematical structure. This structure underpins the robustness, expressiveness and diversity of languages. The mathematics of the theory are in three linked parts: (1) Script algebra, which is the discrete mathematics of feature structures, and will be familiar to many computational linguists; (2) M-script algebra, which extends this to functions on feature structures (as used, for instance, in categorial grammars) ; and (3) Bayesian learning theory, which is a simple application of probability theory. None of them are complex, or require anything beyond school maths; but the power and self-consistency of the learning theory hinges on them. This mathematical/computational basis is established in sections 2 and 3. While the maths is elementary, it may be unfamiliar and inaccessible to some. Fortunately, many of its important consequences can be understood by a simple analogy to chemistry, which I shall develop alongside the maths, in highlighted paragraphs. It describes the theory of language understanding and generation. It is a unificationbased theory, where sentence meanings are feature structures, built up by successive unifications of meaning elements. Many syntactic constraints are constraints on the unifiability of feature structures. The theory is comparable with other unification-based grammars such as LFG and GPSG, showing that it has similar power - and can handle complex features of many adult languages. Like them, it has a reversible model of language understanding and generation.

The theory is fully lexicalised; every word (or word sense) is represented in the brain by a structure called an m-script, which embodies all the syntax and semantics of the word. There are no separate phrase structure rules, transformations or parameters. Therefore if we can learn the m-scripts for words, we can learn a language. Since our knowledge of child language acquisition comes only from studies of language production and comprehension, we need a theory of production and comprehension in order to compare the learning theory with data. Effects which have been attributed to learning limitations can often be understood as production effects - arising from children's strategies for speaking with limited vocabularies. The model of language production is a key part of the theory.

Section 3 describes the process for learning the m-script for each word, and how this leads to `bootstrap' learning of a language. Many of the background assumptions are as in Pinker's theory - for instance, that the child learns by hearing sentences in contexts where he can infer their meaning non-linguistically. But the statistical and mathematical basis of learning is different. It is a Bayesian learning theory, which can be shown to give optimum learning performance. The learning procedure projects out common structure from examples (rejecting random extra noise), and has a Bayesian criterion of sufficient evidence. This means it can learn the m-script for any word from a few noisy examples. It can gather implicit negative evidence and learn from it.

Section 4 discusses the evolutionary origins of language, and the processes of historic language change; there are parallels between the two. I propose that the capacity to use scripts (which underlie language meanings) evolved to support primate social intelligence; so they have a 20 million year evolutionary history and require a fast, robust learning mechanism. M-scripts arose more recently, in part to support a primate theory of mind. Language learning allows word m-scripts to reproduce and propagate through a population of speakers, and so to evolve (as a form of Dawkins' `memes'). They evolve to maximise the speed and efficiency of communication, and evolve much faster than the brains which use them. This accounts for many prominent features of language (such as approximate regularity, grammatical subjects, and the Greenberg-Hawkins universals) as the results of language change (m-script evolution) rather than innate features of the human brain.

Section 5 is the longest section in the paper, and compares the predictions of the learning theory with observations. I first discuss some general properties of language acquisition (such as its speed, robustness, and approximate order of acquisition). I then discuss particular observations, in the order: acquisition of the lexicon, phrase structure, morphology, complement-taking verbs, auxiliaries, alternating verb arguments, pronouns and movement, and finally bilingual language acquisition.

For the majority of these 101 comparisons, the m-script theory is in good unforced agreement with the data, not requiring extra assumptions. Where extra assumptions are required, they do not strain credibility. I have found no major conflicts between the theory and the data. However, I have not been able, in the time and space, to examine the data as thoroughly as, for instance, Pinker (1984) does in his comparisons; much work remains to be done for a full evaluation of the theory. Nevertheless, the initial indications from these comparisons are positive.

Section 6 compares this theory with other theories of language learning, I discuss Pinker's theories (which have much in common with the m-script theory, and from which I have borrowed the treatments of some phenomena), then discuss Principles & Parameters theories, Connectionist theories, Slobin's Operating Principles, and Siskind's computational model of lexical acquisition.

Section 7 concludes, summarising the main results of this work.

There are four appendices: (A) describing algorithms for the m-script operations which underpin the theory; (B) showing that Bayesian learning gives optimal performance; (C) deriving a fundamental theorem of language learning; and (D) describing the computer program which implements the theory.

REFERENCES:

1. Aksu-Koc, A. A. and Slobin, D. I. (1985) The acquisition of Turkish, in The cross-linguistic study of language acquisition, Volume 1: the data, D. Slobin (ed) Lawrence Erlbaum, Hillsdale.
2. Austin, P. and Dalrymple, M. (1995) A bibliography of Lexical Functional Grammar. Electronic manuscript available from dalrymple@parc.xerox.com
3. Baker, C. L. (1979) Syntactic theory and the projection problem. Linguistic Enquiry 10: 533581
4. Barrett, M. (1995) Early lexical development, in P. Fletcher and B. MacWhinney (Eds) The handbook of child language, Blackwell, Oxford
5. Barwise, J. and J. Perry (1983) Situations and attitudes, MIT press, Cambridge, MA.
6. Bates, E. Dale, P. S. and Thal, D. (1995) Individual differences and their implications for theories of language development, in P. Fletcher and B. MacWhinney (Eds) The handbook of child language, Blackwell, Oxford
7. Bates, E. and MacWhinney, M. (1989) Functionalism and the competitive model, in MacWhinney, B. and Bates, E. (Eds) The crosslinguistic study of sentence processing, Cambridge University Press, Cambridge.
8. Beckwith, R. Tinker, E., and Bloom, L. (1989) The acquisition of non-basic sentences, paper presented at the Boston child language forum, Boston, USA.