Towards a Unified Model of Language Acquisition

Fernand Gobet (fernand.gobet@brunel.ac.uk)
Centre for the Study of Expertise, Brunel University
Uxbridge, UB8 3PH, UK
Julian M. Pine (julian.pine@liverpool.ac.uk)
Daniel Freudenthal (d.freudenthal@liverpool.ac.uk)
School of Psychology, University of Liverpool
Liverpool, L69 7ZA, UK

Abstract

In this theoretical paper, we first review and rebut standard criticisms against distributional approaches to language acquisition. We then present two closely-related models that use distributional analysis. The first deals with the acquisition of vocabulary, the second with grammatical development. We show how these two models can be combined with a semantic network grown using Hebbian learning, and briefly illustrate the advantages of this combination. An important feature of this hybrid system is that it combines two different types of distributional learning, the first based on order, and the second based on co-occurrences within a context.

Introduction

Distributional approaches to language learning have a long history in psychology and linguistics. Moreover, recent research has demonstrated that an enormous amount of information is present in the statistical distribution of words contained in large text-based and conversation-based corpora (e.g. Brent & Cartwright, 1996; Finch & Chater, 1992; Landauer & Dumais, 1997). However, distributional models of language learning have traditionally faced two specific kinds of criticism.

The first of these is based on a set of logical arguments against the possibility of successful distributional learning derived from learnability theory. While such arguments are useful in illustrating the scale of the problem facing a distributional approach to language learning, they derive much of their power from the way in which they conceptualise language acquisition as a single logical problem rather than as a complex developmental process. Once one accepts the possibility that distributional learning procedures may interact in complex ways with cognitive and developmental constraints, the issue of whether it is possible to build a successful distributional learning model of language acquisition becomes an empirical rather than a logical question.

The second kind of criticism reflects the idea that distributional-learning accounts make unrealistic assumptions about the child's processing abilities. However, this kind of criticism, while valid for the specific models at which it has been aimed (e.g. Maratsos & Chalkley, 1980), does not generalise to all distributional analysers. For example, we have recently built computational models of grammatical development (Croker, Pine & Gobet, 2000; Freudenthal, Pine, & Gobet, 2006) and the acquisition of vocabulary (Jones, Gobet & Pine, 2005) based on the EPAM/CHREST architecture (Feigenbaum & Simon, 1984; Gobet et al., 2001). These models are capable of extracting a great deal of linguistic information from realistic input samples using a relatively simple performance-limited distributional learning mechanism. Moreover, one of the interesting features of these simulations is the extent to which the performance limitations built into the distributional learning mechanism are actually responsible for the similarity between the child's and the model's output (see below).

Distributional approaches have traditionally focused either on syntax (e.g. Finch & Chater, 1992), or on phonology (e.g. Brent & Cartwright, 1996), or on semantics (e.g. Landauer & Dumais, 1997). However, it would obviously be of scientific interest to have a single computational model that covered these three aspects of language, taking as input naturalistic data (i.e. corpora of mothers' child-directed speech). The aim of this theoretical paper is not to present detailed simulations of a particular phenomenon but to show (a) how different aspects of language (phonology and syntax) can be modelled using what is essentially the same system; (b) how a semantic network can be grown incrementally; and (c) how these different sub-models can be brought together into a single unified model. We first briefly present the EPAM/CHREST architecture, and then our models of the acquisition of syntax and vocabulary. We then describe a model that incrementally builds up a semantic network. Next, we show how this semantic network can be connected to our models of syntax and vocabulary acquisition, and provide some examples of the behaviour of this composite system.

The EPAM/CHREST Architecture

The EPAM theory (Feigenbaum & Simon, 1984) is the computational framework behind our models of syntax and vocabulary acquisition. EPAM has a long history of successful simulation of human cognition, including verbal learning behaviour, expert behaviour, and concept formation (see Gobet et al., 2001, for a review).

EPAM is a self-organising system that models learning as the construction of a discrimination net (see Figure 1a). The nodes in the discrimination net are LTM symbols, having arbitrary subparts and properties, that can be used as processing units, and the links contain tests that must be satisfied in order to reach the next node. The basic mechanisms are as follows. During perception, an object is sorted through a sequence of tests, each relating to some feature of the object. When the description of the object mismatches the internal representation (the *image*) it has been sorted to, a new test-link, which relates to the mismatched feature, is added. When the object is sorted to an internal representation that under-represents it, new features are added to the image by chunking.

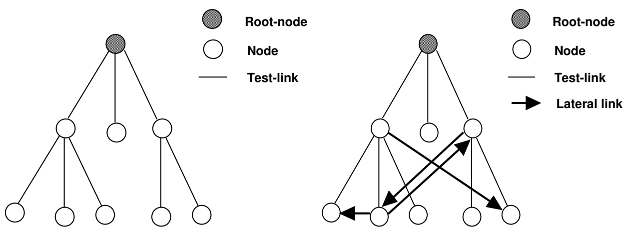

Figure 1: (a) A simple discrimination net, as used in the EPAM models. (b) A discrimination net with lateral links, as used in the CHREST models. Lateral links can be used to connect nodes which share several features, or to create productions, one node serving as the condition, and the other as the action.

Until recently, EPAM has been explored mainly as a theory modelling access to long-term memory (LTM). Simon (1989) has proposed that EPAM nets constitute an index to procedural and declarative memories, but has not given any details about how this should be implemented in a working computational model. CHREST (Gobet & Simon, 2000; Gobet et al., 2001), an extension to EPAM, aims to tackle this question by showing how procedural and declarative knowledge can be created by connecting nodes of the discrimination net by 'lateral' links (see Figure 1b).

MOSAIC

A major aim of our research has been to build a computational model of syntax acquisition in children (MOSAIC), based on the CHREST framework. The basic assumptions are that (a) syntactic categories are actively constructed by the child, using distributional learning abilities; and (b) cognitive constraints in learning rate and memory capacity limit these learning abilities. The major addition to EPAM, as just mentioned, is the presence of lateral links that connect nodes in the discrimination net as a function of similarity in the test links occurring immediately below.

Description

The input given to MOSAIC consists of a set of maternal utterances, taken from the Manchester corpus of the CHILDES database. MOSAIC learns by scanning each utterance in turn, and by adding information to the net using the mechanisms described above. In addition, when a node is accessed by recognition, it is compared with other recently recognised nodes with respect to both preceding and following words. When the overlap is larger than a preset parameter, a lateral link is added that connects the relevant nodes (see Figure 2).

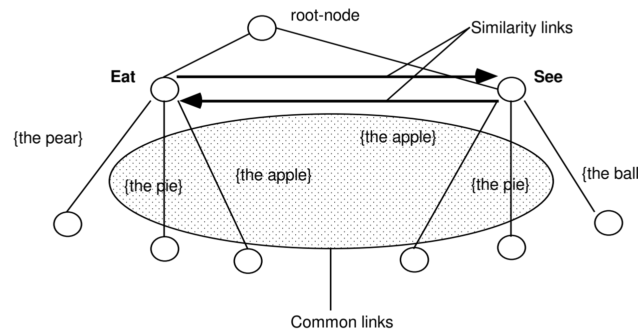

Figure 2: Creation of lateral links in MOSAIC. As the links below 'Eat' and 'See' have sufficient overlap, lateral (or similarity) links are created between them. These links can be used to generate utterances, such as 'Eat the ball'. (Only a subset of the network and only the following context to compute overlap are shown).

Production of utterances can occur in two ways. First, the program can follow a path down using only test links. This will generate utterances that were already in the input. We call this production mechanism *rote output*. Second, in addition to test links, the program can also follow lateral links. This will produce new utterances that were not present in the training input. We call this production mechanism *generation*. Generation can rapidly produce a very large number of new sentences (Jones, Gobet & Pine, 2000).

The performance of MOSAIC has been tested in detail on different sets of phenomena in early syntactic development, of which we describe two here (a) 'Verb-Island' phenomena (i.e. the verb-specific nature of children's early use of word order patterns; Tomasello, 1992) and (b) Optional-Infinitive phenomena (i.e. children's tendency to use finite and non-finite verb forms interchangeably in contexts where a finite verb form is obligatory) (Wexler, 1994).

Verb-Island Phenomena

One of the most important recent constructivist models of early grammatical development is Tomasello's (1992) Verb-Island hypothesis. According to this view, children's early grammars consist of inventories of lexically-specific predicate structures (or 'Verb Islands'). The Verb-Island hypothesis can account for a number of phenomena in children's early multi-word speech. For example, it can explain the lexically-specific patterning of children's early verb use, i.e. the fact that in the early stages of grammatical development, children's ability to generate longer sentences builds up piecemeal around particular verbs, and fails to generalise to new verbs. It can also explain differences in the flexibility with which children use nouns and verbs in their early multi-word speech. For example, young children will readily slot novel nouns into familiar verb structures but tend to restrict their use of novel verbs to the structures in which they have heard those same verbs modelled in the input (e.g. Akhtar & Tomasello, 1997). However, one problem for a strict version of the Verb-Island hypothesis is the fact that, in addition to verb- or predicate-islands, young children also appear to be acquiring structures based around high-frequency items that would not normally be considered predicates such as proper nouns and case-marked pronouns (Pine, Lieven & Rowland, 1998)

MOSAIC is able to simulate the basic Verb-Island phenomenon as the product of a performance-limited distributional analysis of real child-directed speech (Jones et al., 2000). That is to say, it acquires generative structures based around particular lexical items by linking together high frequency words that behave in similar ways in the input. Since such words are more likely to be verbs than nouns, verbs tend to function as structuring items in the model's output whereas nouns tend to function as slot-fillers. Interestingly, however, MOSAIC also acquires structures based around high-frequency words other than nouns (e.g. proper nouns such as 'Mummy' and the child's own name and case-marked pronouns such as 'I' and 'You'). It is therefore also able to simulate the kind of 'other-island' effects reported by Pine, Lieven & Rowland (1998). Moreover, the performance limitations built into MOSAIC's distributional learning mechanism result in MOSAIC generating output which is more similar to the speech of the target child than it is to the speech of the mother on which it has been trained.

Optional-Infinitive Phenomena

One of the most influential recent nativist models of early grammatical development is Wexler's (1994) Optional Infinitive hypothesis. According to this view, by the time that children begin to produce multi-word utterances they have already correctly set all the basic inflectional/clause structure parameters of their language. However, there is an initial stage — the Optional-Infinitive stage — during which they lack the knowledge that tense and agreement are obligatory in finite clauses.

The Optional-Infinitive hypothesis makes very clear predictions about what children in the OI stage will and will not say. Thus, it predicts, first, that children will use tensed and untensed forms interchangeably in contexts where tensed forms are obligatory (e.g. producing 'she going' and 'she go' as well as 'she's going' and 'she goes'); second, that children will make various kinds of case-marking errors (e.g. producing 'her go' and 'her did' instead of 'she goes' and 'she did'); and third, that children will not make case-marking or agreement errors with agreeing forms (e.g. 'him goes' instead of 'he goes' or 'he are' instead of 'he is').

MOSAIC is able to simulate the basic Optional-Infinitive phenomenon by learning sequences such as 'he going' and 'he go' from questions such as 'Is he going?' and 'Does he go?' and then acquiring generative patterns such as 'he + untensed verb' by forming lateral links between pronouns (Croker et al., 2000; Freudenthal et al., 2006). It is also able to reproduce the basic pattern of errors seen in young children. This includes the occurrence of case-marking errors such as 'her go' and 'her did', but also the occurrence of other low frequency errors (e.g. 'him goes' and 'he are') that are problematic for a strict version of the Optional-Infinitive hypothesis. Although some of these errors are produced by rote learning (e.g. by learning sequences such as 'her go' from 'let her go'), most of them are produced by generating across lateral links. This includes errors for which there is no direct model in the input (e.g. 'her goes') and errors for which there is such a model (e.g. 'her go'). Note that some of the simulations on the Optional-Infinitive phenomenon have been done in four languages (English, Dutch, German, and Spanish).

EPAM-VOC

There has recently been a great deal of interest in vocabulary acquisition, with Baddeley and Hitch's working-memory model being adapted to account for vocabulary learning (e.g. Gathercole & Baddeley, 1989). Gathercole and Baddeley claim that the phonological loop part of the model is a critical mechanism for learning new words. The phonological loop has two linked components: the phonological short-term store, and the sub-vocal rehearsal mechanism.

A key experimental task for investigating Gathercole & Baddeley's model is the nonword repetition test. This test involves two sets of nonwords, one with single consonants (e.g. 'rubid') and one with clustered consonants (e.g. 'glistow'). Several studies using these types of nonwords have found that repetition accuracy decreases as the number of syllables in the nonword increases, excepting one-syllable nonwords (e.g. Gathercole & Adams, 1993), and that accuracy decreases for clustered consonant nonwords. Performance on this test correlates strongly with vocabulary knowledge.

Vocabulary acquisition is another domain to which we have applied EPAM/CHREST (Jones et al., 2005). As with syntax acquisition, learning is seen as the development of a discrimination network. The model (EPAM-VOC) also makes assumptions about verbal working memory.

Description

We give as input to the model utterances from mothers' child-directed speech so that it can learn phonemes and combinations of phonemes. The mothers' utterances are converted into a sequence of phonemes before being used as input. This is done using the CMU Lexicon database which cross-references words with their phonemic representations. The use of phonemic input assumes that some form of phonemic feature primitives already exist to distinguish one phoneme from another.

EPAM-VOC begins with an empty root node. When it sees an input (a sequence of phonemes), new nodes and links are created. At first, most of the new nodes and links are just for single phonemes. As learning progresses, the information at nodes will become sequences of phonemes and therefore segments of speech (e.g. specific words) rather than just individual phonemes.

The model offers a specification of the phonological loop and a method by which the loop interacts with long-term memory. The storage part of the phonological loop is a decay-based store which allows items to remain in the store for about 2000 ms (i.e. consistent with the phonological loop estimates). The input is cut-off as soon as the time limit is reached, because there is no rehearsal to refresh the input representations.

The cumulative time required to encode the input provides a theory of how the amount of information in the phonological store is mediated by long-term memory. When an input is heard, long-term memory (the EPAM-VOC network) is accessed and the input is represented using the minimum number of nodes possible. Rather than the actual input being placed in the phonological store, the nodes which capture the input are used. The length of time taken to represent the input is therefore calculated on the number of nodes that are required to represent the input. The time allocations are based on estimates from Zhang and Simon (1985), who estimate about 400 ms to match each node, and about 84 ms to match each syllable in a node except the first (which takes 0 ms).

Simulations

When trained on speech addressed to 2-3 year old children (4,000 maternal utterances), EPAM-VOC provides a good approximation to the performance of 2-3 year olds on the nonword repetition test (Jones et al., 2005). However, attempts to simulate the performance of 4-5 year olds simply by increasing the size of the input sample (to 25,000 maternal utterances) result in a much poorer fit to the data, with the model seriously underestimating children's level of performance on 3- and 4-syllable non-words. Interestingly, however, when the model is trained on a smaller but more varied input sample (consisting of 5,000 words selected at random from the CMU lexicon), its performance improves beyond that of 4-5 year old children so that it now performs at ceiling on 3-syllable non-words. These results illustrate the critical role that input characteristics play in determining the model's level of performance, and suggest that non-word repetition performance may be highly sensitive to the lexical diversity of the input which different children receive.

Combining EPAM-VOC with MOSAIC and Adding a Semantic Network

In order to keep the analysis of our simulations relatively simple, we have so far treated MOSAIC and EPAM-VOC as separate models. However, it is worth noting that the two models are conceptually very similar. Moreover since EPAM-VOC can take phonemically coded utterances and use them to learn both lexical items and strings of lexical items (which is basically what MOSAIC is doing), it is possible to collapse EPAM-VOC and MOSAIC into a single model, MOSAIC-VOC, that is capable of learning both vocabulary and syntactic structure in the same way.

Even so, when compared with children, MOSAIC-VOC lacks access to several important kinds of linguistic information (in particular, semantic, pragmatic and communicative information). We now consider how semantic knowledge (or more accurately, an approximation to semantic knowledge) can be added to MOSAIC-VOC.

Approximating Semantics

A fully-fledged model of semantics would require a theory of how semantic information is linked to perceptual, motor and proprioceptive information. Although research with autonomous robots is making impressive progress (e.g. Roy & Pentland, 2002), we are far from even approximating how semantics is derived from these types of information. We will therefore have to settle for a lesser goal and use an approximation to semantics.

As noted above, recent work has shown that large text-based corpora contain a vast amount of syntactic and semantic information that can be extracted with surprisingly simple techniques. In their ground-breaking article, Landauer and Dumais (1997) propose a method for extracting semantic information from such texts. Latent Semantic Analysis (LSA) is a mathematical method for extracting the similarity of meanings of words and passages from the analysis of large text-based corpora. Using a general form of factor analysis known as singular-value decomposition, LSA reduces large matrices of word-by-context data into 100-500 dimensions.

The central rationale is that the contexts in which a given word does and does not appear powerfully constrain and determine the similarity of word meanings and sets of word meanings to each other. As indicated by the term "Latent Semantic Analysis," the similarity values estimated by LSA are not simply based upon co-occurrence frequencies, but depend on a deeper statistical analysis.

What LSA is doing is computing correlations between words within different contexts. The method represents an efficient solution, but there are other ways of achieving this goal as well. The method we use here is to dynamically create a semantic network capturing correlations between words belonging to the same context (utterance) using Hebbian learning. We first describe how the semantic network is created, and then how its creation is combined with MOSAIC-VOC.

Creating a Hebbian Semantic Network

The semantic network is made up of a set of units with sparse connections. As in the previous simulations, utterances from a mother speaking to her child are used as input. At the beginning of learning, the semantic network is empty. When words unknown to the network are presented, units are created for each of these words. Connections are also created between units denoting words belonging to the same utterance (context). When words appear again in the same context, the connections between them are updated using a simple Hebbian learning rule. Thus, this method creates connections only for words that have co-occurred in the same context.

Units have a default activation of 0. Spreading activation within the semantic network occurs as follows. The words in the input set have an activation of 1. Activation then spreads from one unit to another by multiplying the activation of the unit by the weight of the connection. For each unit, the difference in activation is the sum of all weight/activation products. The final activation of the unit is 'squashed' using a sigmoid function.

Linking the Syntactic and Semantic Networks

The learning of the MOSAIC network[1] and of the semantic network occur in parallel, with, in addition, the creation of links joining the nodes of the former with the units of the latter (see Figure 3). When an utterance is presented to MOSAIC, the nodes traversed during sorting are activated. At the same time, units in the semantic network that refer to words mentioned in the utterance are activated. Interlinks are created between activated nodes and units, fully connecting them, and later on, updated using Hebbian learning. When activation is spread from the semantic network, the activation of a node in MOSAIC is computed as follows. First, for each node, the products of the connected units by their interlink weights are summed. Second, this sum is squashed using a sigmoid function. Third, this squashed sum is multiplied with the (squashed) size of the image associated with that node. During the production of utterances (either by rote or by generation), activation is used to bias the choice of words or sequences of words.

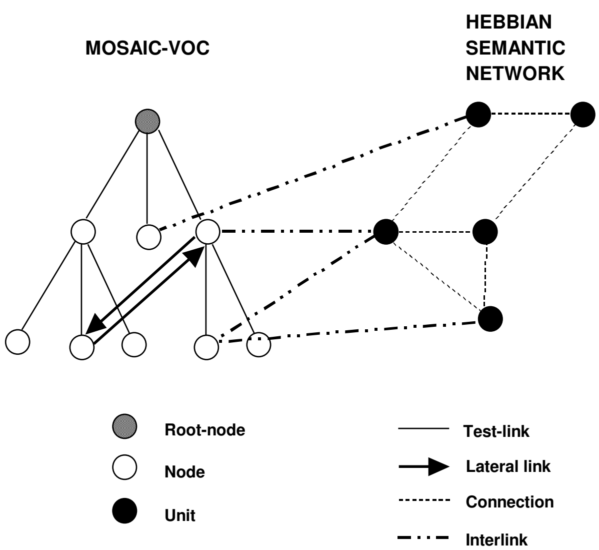

Figure 3. How MOSAIC-VOC is connected to a sparse Hebbian semantic network in the system described in the paper. Hebbian learning operates on connections and interlinks. (Only a subset of interlinks are shown.)

Generating Output

At the end of the learning stage (actually, at any time during learning), we have an EPAM-like discrimination network linked to a semantic neural network. There are several ways this hybrid system can be used: (a) the semantic network can be used to spread activation to MOSAIC and generate sentences; (b) MOSAIC can be used to parse an utterance and to propagate the activation to the semantic network, thus approximating the "understanding" of a verbal utterance; (c) finally, two hybrid networks can interact together, combining the operations outlined above. An utterance produced by the first network is parsed by the second network, and its semantic memory is activated accordingly. This activation leads to the generation of a new utterance through the MOSAIC network of the second network.

An Example: Using the Semantic Network to Generate Sentences

A corpus of 21,329 utterances was used as input, taken from a mother interacting with her child. Each utterance in the corpus was learned by the method described above.

To generate an utterance, we activated one or several node(s) in the semantic network, spread activation through the semantic network, spread activation to MOSAIC, and used MOSAIC to produce a sentence. When outputting an utterance, MOSAIC was biased in favour of nodes having a high activation, and nodes with activation below a threshold cannot be used.

At the end of one pass through the corpus, MOSAIC contained 49,640 nodes and 367 lateral links. 2,797 units were created in the semantic network, with an average of 22 connections within the semantic network (minimum = 0; maximum=1,182). 261,100 interlinks were created (average 93 per unit; minimum=0; maximum=7,213).

In general, the semantic network proves itself useful in filtering out the utterances generated by MOSAIC. For example, MOSAIC generates about 20 times less utterances when the semantic network is used.

The semantic network allows some semantic generalisation. For example, when activating 'DADDY', the model may produce utterances containing 'MUMMY' or 'BABY'. Or, activating 'CAT' may yield utterances with 'DOG', 'COW', or even 'ZOO'. The semantic network also allows some 'weak-contextualized' generalisation. For example, activating the set {ME EAT DRINK} produces utterances like 'ME PLEASE' or 'ME NOW'. However, not all utterances can be categorised as semantic generalisation or as weak-contextualized' generalisation. Finally, some fairly sophisticated utterances can be generated by the model, such as 'I NEARLY TORE IT' or 'I DRYED IT FOR ME'. In general, the same results apply when MOSAIC produces rote outputs.

In spite of these positive features, the Verb-Island and Optional-Infinitive phenomena we have described above are still present in the model's output.

Conclusion

The most original feature of our approach to the study of vocabulary and syntax acquisition is our attempt to use computational modelling based on unsupervised learning with naturalistic input data and to carry out detailed comparison of the model's output with children's data. Until

[1] To keep the presentation simple, we focus here on the connections between the semantic network and MOSAIC.

now, we have randomly selected utterances produced by MOSAIC, with the difficulty that some of them may be semantically anomalous. We are confident that the addition of the semantic network will alleviate this problem and that the hybrid system may turn out to be useful for selecting utterances during simulations. Two characteristics of our approach—use of naturalistic data as input and detailed comparison with children's data—single it out from other attempts to develop computational models addressing both syntax and semantics, such as the neural net used by Hadley and Hayward (1995). Such models are typically limited to small artificial grammars.

While systematic evaluation is obviously needed, we speculate that this hybrid system, which scales up well in preliminary simulations, may obtain interesting results because it combines two different types of distributional learning, the first based on order, and the second based on co-occurrences within a context. We plan to test the plausibility of this hypothesis by looking at the extent to which patterns of semantic activation can be used to constrain MOSAIC's performance and thereby reduce the frequency of certain kinds of errors which, although present in children's speech, occur less often than they do in the model's output.

While our main interest is in simulating in detail language development in children, it is also worth pointing out that this hybrid system may have wider relevance, to the study of language in general. For example, a similar approach could be used in the field of human-computer interaction. Alternatively, it could be used in the study of text understanding. The way the semantic network is created is consistent with the type of networks used by one of the prominent theories in this field, the construction-integration model (Kintsch, 1998), and seems psychologically more plausible than LSA. In addition, the MOSAIC module could be extended to act as an adult syntactic parser — something that currently requires hand coding in the construction-integration model.

References

Akhtar, N. & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology, 33*, 952-965.

Brent, M.R. & Cartwright, T.A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*, 93-125.

Croker, S., Pine, J. & Gobet, F. (2000). Modelling optional infinitive phenomena: A computational account of tense optionality in children's speech. *Proc. of the 3rd International Conference on Cognitive Modelling* (pp. 78-85). Veenendaal (NL): Universal Press.

Feigenbaum, E.A. & Simon, H.A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8,* 305-336.

Finch, S. & Chater, N. (1992). Bootstrapping syntactic categories. *Proc. of the 14th Annual Conference of the Cognitive Science Society* (820–825). Hillsdale, NJ: Erlbaum.

Freudenthal, D., Pine, J., & Gobet , F. (2006). Modelling the development of children's use of optional infinitives in English and Dutch using MOSAIC. *Cognitive Science, 30*, 277-310.

Gathercole, S.E. & Adams, A.M. (1993). Phonological working memory in very young children. *Developmental Psychology, 29*, 770-778.

Gathercole, S.E. & Baddeley, A.D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. Journal of *Memory and Language, 28*, 200-213.

Gobet, F., Lane, P.C.R., Croker, S., Cheng, P.C.H., Jones, G., Oliver, I., & Pine, J.M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*, 236-243.

Gobet, F. & Simon, H.A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24*, 651-682.

Hadley, R.F. & Hayward, M.B. (1995). Strong semantic systematicity from unsupervised connectionist learning. *Proc. of the 17h Annual Meeting of the Cognitive Science Society*, pp. 358-363. Hillsdale, NJ: Erlbaum.

Jones, G., Gobet, F. & Pine, J.M. (2000). A process model of children's early verb use. *Proc. of the 22nd Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum

Jones, G., Gobet, F., & Pine, J. M. (2005). Modelling vocabulary acquisition: An explanation of the link between the phonological loop and long-term memory. *Journal of Artificial Intelligence and Simulation of Behaviour, 1,* 509-522.

Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: CUP.

Landauer, T.K. & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Maratsos, M.P. & Chalkley, M.A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K.E. Nelson (Ed.), *Children's Language*. (Vol. II). NY: Gardner.

Pine, J.M., Lieven, E.V.M. & Rowland, C.F. (1998). Comparing different models of the development of the English verb category. *Linguistics, 36*, 807-830.

Roy, D., & Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science, 26*, 113-146.

Simon, H.A. (1989). *Models of thought*. (Vol. II). New Haven: Yale University Press.

Tomasello, M. (1992). *First verbs*. Cambridge: CUP.

Wexler, K. (1994). Optional infinitives, head movement and the economy of derivations in child grammar. In D. Lightfoot & N. Hornstein (Eds.) *Verb movement.* Cambridge, MA: CUP.

Zhang, G. & Simon, H.A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory and Cognition, 13*, 193-201.