Modelling children's negation errors using probabilistic learning in MOSAIC

Steve Croker (s.croker@derby.ac.uk)
Centre for Psychological Research in Health and Cognition
University of Derby, Mickleover, DE3 9GX

Julian M. Pine (jp@psychology.nottingham.ac.uk)
Fernand Gobet (frg@psychology.nottingham.ac.uk)
School of Psychology, University of Nottingham,
University Park, Nottingham, NG7 2RD

Abstract

Cognitive models of language development have often been used to simulate the pattern of errors in children's speech. One relatively infrequent error in English involves placing inflection to the right of a negative, rather than to the left. The pattern of negation errors in English is explained by Harris & Wexler (1996) in terms of very early knowledge of inflection on the part of the child. We present data from three children which demonstrates that although negation errors are rare, error types predicted not to occur by Harris & Wexler do occur, as well as error types that are predicted to occur. Data from MOSAIC, a model of language acquisition, is also presented. MOSAIC is able to simulate the pattern of negation errors in children's speech. The phenomenon is modelled more accurately when a probabilistic learning algorithm is used.

Introduction

Language has proved to be a very rich domain for computational modelling, particularly the modelling of language acquisition. Models of language acquisition have often attempted to explain the occurrence of errors in child speech. In recent years, MOSAIC (Model Of Syntax Acquisition In Children) has been used to explain various error types via a process of extracting distributional information from maternal input, including optional infinitive errors in English (Croker et al., 2000) and in Dutch (Freudenthal et al, 2001, 2002a), case-marking errors (Croker et al, 2001) and subject omission errors (Freudenthal et al, 2002b). In all of these cases MOSAIC provides a good fit to the data. Given MOSAIC's ability to model relatively common errors, an obvious next step is to investigate the extent to which it is able to simulate errors that are rare in child speech. This is a good way of assessing the sensitivity of the model.

Children's Negation Errors

One relatively rare error type in English involves placing an inflected verb form (e.g. 'goes') to the right of the negative particle (i.e. 'not') and producing 'not goes'; the correct form, 'does not go', is inflected to the left of 'not'. Even in the earliest stages of multi-word speech, English-speaking children's use of inflected and uninflected verb forms tends to pattern correctly with respect to placement of the negative particle 'not'. Thus, children regularly produce utterances such as 'he doesn't go' and 'he not go'. The first of these utterance types is grammatically correct. The second is incorrect because of the absence of a tensed form (in this case 'does'). However, in both cases the order of the negative particle and the verb form used by the child conforms to the rules of English grammar. In contrast, children rarely produce utterances such as 'he not goes' or 'he goes not'. Both of these utterance types involve the incorrect use of tensed forms. The first is ungrammatical because it involves the use of a tensed verb form to the right of negation. The second is ungrammatical because it involves the use of a main verb to the left of negation.

The very low frequency of errors involving tensed verb forms in young children's speech has been taken as evidence that, by the time that children begin to produce multi-word utterances, they have already correctly set all the basic clause structure parameters of their language. Thus, Wexler (1994) argues that very young English-speaking children already know that inflected verb forms must be placed to the left of negation in English, and that these forms cannot be main verbs. Harris & Wexler (1996) test this prediction by examining transcripts of 10 children, considering sentences containing negation before a main verb. These utterances were analysed for the presence of inflection – either the present tense ('goes') or the past tense ('went') - or the lack of inflection. The use of inflection in negatives was then compared with the use of inflection in affirmative sentences. Across all 10 children, 43% of affirmative sentences were inflected and 9.6% of negative sentences were inflected. Harris & Wexler use these figures to support the notion that verbs are not tensed after negation, dismissing the inflection rate for negative utterances as a reflection of performance errors in production.

Harris & Wexler's account is a well-specified, if somewhat complex, nativist account of children's use of negation which assumes that children are born with rich domain-specific knowledge of language. An alternative school of thought is that children learn language by picking up regularities in the speech that they hear; children are essentially distributional analysers. Evidence for this position comes from Tomasello (2000a, 2000b), who argues that children's early language is item-based, organised around particular words and phrases. Children initially learn a restricted morphology for each verb learnt; they are most likely to use the same morphological marker on a novel verb as was presented to them, rather than immediately generalising to other endings, which would suggest that a child's knowledge of language is input-driven. Further evidence for lexical specificity with respect to verb morphology in children's speech is presented by Brown (1973) and Pine, Lieven & Rowland (1998). An argument for the development of language around specific lexical items is provided by Lieven, Pine & Baldwin (1997), who propose that children's novel utterances are generated using positional patterns where a number of variable words or phrases can be combined with positionally constant items. These findings are consistent with recent work in computational modelling (e.g. Cartwright & Brent, 1997; Elman, 1993; Redington, Chater & Finch, 1993, 1998) which has shown that it is possible to derive a significant amount of syntactic information from a distributional analysis of the statistics of the language being learned.

Child Data

We present data obtained from three children (Anne, Aran and Becky), taken from the Manchester corpus (Theakston, Lieven, Pine & Rowland, 2000) of the CHILDES database (MacWhinney & Snow, 1990). This corpus consists of transcripts of tape recordings made twice every three weeks over a period of 12 months, between the ages of approximately 2 and 3 years. Each session consists of two half-hour recordings, one made during free play and the other during structured play. There are two important points that need to be made about the way these data were analysed. First, because we were interested in identifying a corpus of utterances including verbs and comparing this corpus with the output of the model, and because, within the model, there is no way of deciding whether a word is being used as a verb or not, we needed a way of identifying verbs that was independent of the way in which they were used by the child and the model. An analysis was therefore made of the frequencies with which words were used as verbs in the child's speech and, for the purpose of this research, words were classified as verbs if they occurred as verbs in 90% or more of cases in the mother's speech corpus. Second, the data used in analysing both the performance of the children and the performance of the model consisted of types, not tokens. Much of the research in children's speech is based on analyses using tokens as the entire corpus is considered. However, in our analyses it was necessary to use only types as the model does not produce multiple instances of utterances in the same way as the child. All of Anne's, Aran's and Becky's utterances containing both a verb and the word 'not', or one of its contractions ('shouldn't', 'won't' etc.), were analysed for the presence of various patterns. This yielded samples of 478, 360 and 557 utterances for Anne, Aran and Becky, respectively. The patterns used in this analysis are 'correct' (grammatical) utterances such as 'doesn't go' or 'hasn't gone', 'not + untensed verb' (e.g. 'not go'), 'tensed verb + not' (e.g. 'goes not') and 'not + tensed verb' (e.g. 'not goes'). The former two patterns are predicted to occur by Harris & Wexler, whereas the latter two are predicted not to occur. We also found it necessary to include a further category – 'untensed verb + not' ('go not') – as Aran made errors of this type.

Results

The children's use of verbs and negation is shown in Table 1[1]; grammatically correct utterances have been omitted from the table. The figures in brackets refer to the number of errors in the sample, the figures outside brackets represent the error rate expressed as a percentage of the sample. An analysis of these data shows that as well as forming grammatically correct utterances containing the negative particle, all three children use untensed verb forms following negation (e.g. 'I not need this'). In addition, Anne and Becky produced errors in which a tensed verb was used after negation (e.g. 'not fits') and Aran produced the negative particle *preceded* by an uninflected main verb ('go not') – errors which are predicted *not* to occur by Harris & Wexler.

MOSAIC

MOSAIC is a symbolic modelling architecture which consists of a hierarchical discrimination network. The network is grown as input is presented to the model. When an utterance is presented, each word in the utterance is considered in turn, which allows the utterance to be sorted to a given node. If the word currently considered has not previously been seen by the model, a new node corresponding to that word is

Table 1: Children's negation errors.

	Anne	Aran	Becky
not go	(56) 11.7	(59) 16.4	(35) 6.3
goes not	0	0	0
not goes	(2) 0.4	0	(1) 0.2
go not	0	(1) 0.3	0

[1] The utterance types to the left of the table are representative of classes of error rather than errors involving those particular words.

created. The new node is created at the first layer of the network, just below the root node. This first layer may be seen as the layer where the 'primitives' of the network (i.e., the individual words that have been seen by the mo del) are learned and stored.

The model learns the distributional statistics of both words that follow and words that precede a given word - the network contains information about which words have been presented as occurring immediately before a particular item and which words have been presented as occurring immediately afterwards. Figure 1 depicts a fragment of a network created in this manner.

Generative Links

Once the network has been trained, it can be used to produce utterances in two ways: by recognition and by generation. Utterances produced by recognition are essentially rote-learned (i.e. they are utterances or portions of utterances presented to the model in the input corpus). These are produced by starting at each node in turn, and following the test links down the network. Novel utterances can also be produced by a process of generation – the ability to follow 'horizontal' links in the network as well as 'vertical' ones. These horizontal, generative links are created as follows: If two words in the network occur frequently in similar contexts (e.g. they are preceded and succeeded by the same items), then a generative link can be made between these items. The number of common features needed to create a generative link (the similarity measure) is the degree of overlap between items that precede and succeed any two nodes. This is calculated by taking all the children of any two nodes and assessing whether the proportion of children shared by both nodes exceeds a certain threshold with respect to the total number of child nodes. A number of variations with different values have been assessed. In this paper, values of 4%, 8% and 10% have been used. The value is the same in both directions (e.g. 8% is the critical value for shared nodes both above and below the nodes under consideration). Figure 2 contains an example of generative link creation. In this figure, 'does', can' and 'will' all precede 'he' and 'she'. 'Goes', 'likes' and 'jumps' all follow 'he' and 'she'. A generative link is formed between 'he' and 'she' as a result of this contextual similarity.

Probabilistic Learning

In the version of MOSAIC used to model optional infinitive errors and case-marking errors (Croker et al., 2000, 2001), any item presented to the model is learnt at once. The probability of learning any given word is 1 at all times. In addition to data from simulations with p=1, we present data from simulations in which a parameter is set which determines how likely it is that any given word in the input is learnt. This parameter was set to 0.1. This means that, on average, a word must be seen 10 times before it is added to the network. This form of learning gives a positive bias to words and phrases that occur many times in the input corpus. The output is therefore a reflection of the frequency with which words occur, rather than a reflection of which items are present in the input. To offset the fact that little is learned from one presentation of the input corpus, the input is presented several times before an output is produced. For the p=0.1 condition, analyses of the output produced by MOSAIC after the input set had been presented both 5 and 10 times are included here.

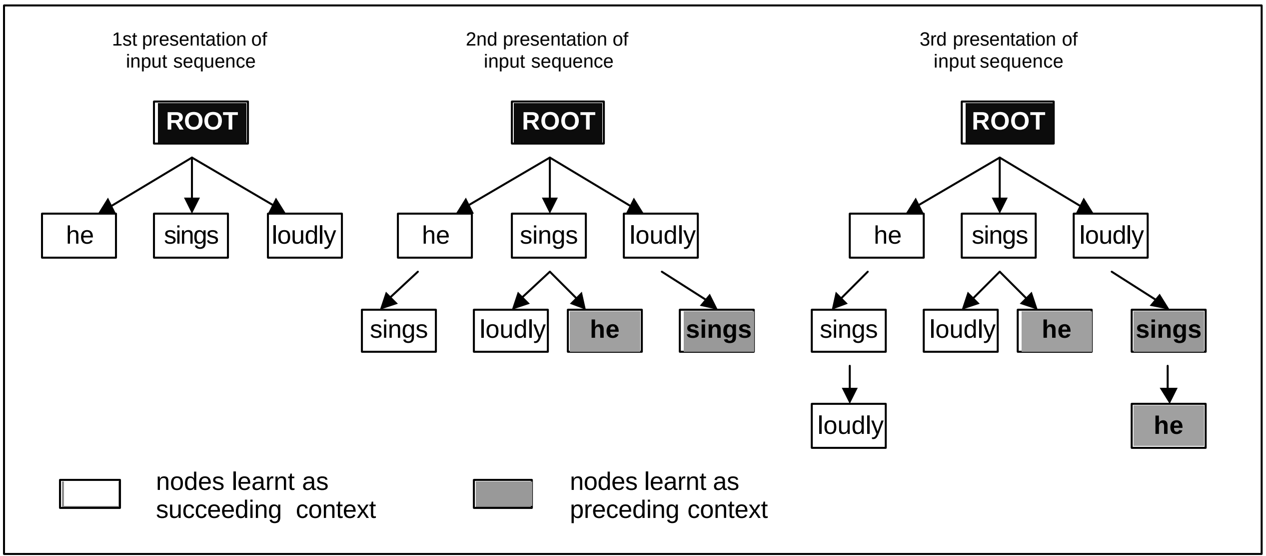

Figure 1: Network formed after the phrase 'he sings loudly' has been presented to MOSAIC 3 times.

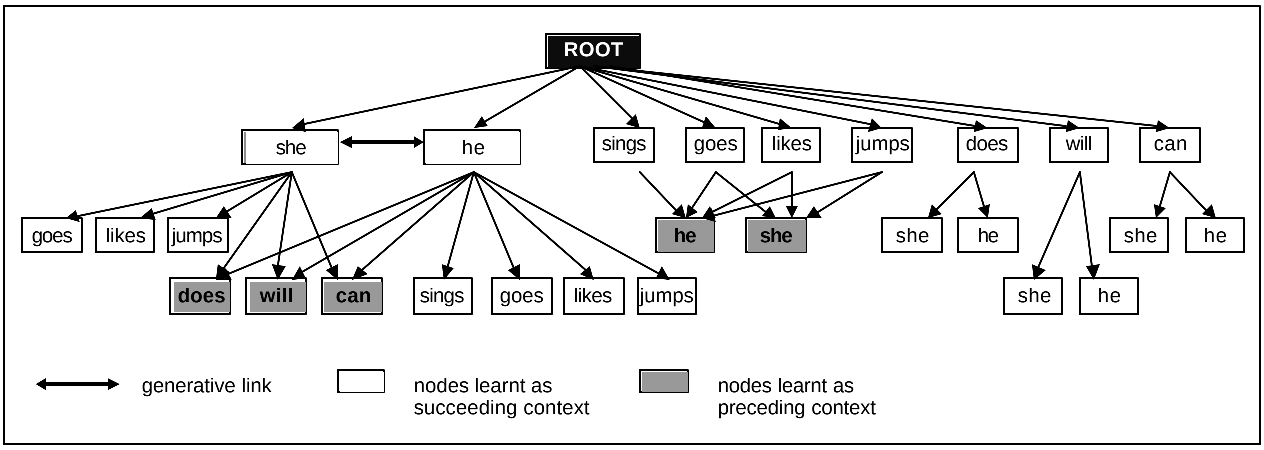

Figure 2: Generative link creation.

Model Data: p=1

Method

The model was run on input data from the mother of Anne, one of the children in the CHILDES database, collected in the same sessions as the child's speech. This is a very large corpus of naturalistic input, consisting of 33,390 utterances. Three different versions of the model were trained, with varied input sets and overlap parameters – two versions with the full input corpus and one version with only the first half of the corpus. The overlap parameter was set to 10% for the first of these and 4% for the others. A 4% overlap allows more generative links to be made as only 4% of the child nodes of any two items need to be shared to create a link; this means that MOSAIC becomes much more generative than with a 10% overlap. With the full input presented and an overlap value of 10%, MOSAIC produced 35,758 utterance types, of which 18,852 were produced by recognition and 16,906 by generation. With the full input presented and an overlap value of 4%, MOSAIC produced 586,015 utterance types, of which 21,999 were produced by recognition and 564,016 by generation. With just half the input and an overlap value of 4%, MOSAIC produced 194,303 utterance types, of which 10,451 were produced by recognition and 183,852 by generation.

Results

Results for MOSAIC-Anne, with a learning probability of 1 are presented in Table 2. With an overlap value of 10%, some 'not + untensed verb' errors are produced, although no 'negation + tensed verb' errors are made. When the overlap percentage was reduced to 4%, allowing many more links to be made, errors of this type were produced, albeit rarely. The percentage of 'not + untensed verb' errors is lower, just 2%, than the 11.7% produced by Anne or the 5.2% produced by MOSAIC-Anne with a 10% overlap. Also, 3 'untensed-verb + negation errors' are made; Anne does not make any errors of this type. When a smaller input corpus was presented to MOSAIC with the overlap percentage still set to 4% the number of 'not + tensed verb' errors was increased to 3.8% which, whilst higher, is still nowhere near the quantity produced by Anne. In addition, the number of 'untensed verb + negation' errors also increased to 3.6%. This latter type of error has been observed in child speech (Aran makes a few of them), but Anne does not produce such errors.

Discussion

Although not a close match for Anne, MOSAIC is still producing the sorts of errors children (as a group) make and, importantly, MOSAIC does not produce the one error type ('goes not') that is not made by any of the children. Whilst this is a positive result, the model ought to provide a closer fit to Anne than to Aran or Becky as it was Anne's mother's data that MOSAIC was trained on. It is possible that when the probability of learning a word is 1, MOSAIC is sensitive enough to the distributional properties of the input to provide a

Table 2: Negation errors (MOSAIC-Anne and Anne, p=1).

	MOSAIC-Anne			Anne
	10% Overlap	4% Overlap	Half-input, 4% Overlap	
not go	(26) 5.2	(10) 2.0	(19) 3.8	(56) 11.7
goes not	0	0	0	0
not goes	0	(2) 0.4	(2) 0.4	(2) 0.4
go not	0	(3) 0.6	(18) 3.6	0

model of general child data, but not sensitive enough to pick up on the characteristics of the input which distinguish Anne from the other children. This motivated the move to a lower learning probability. If every word or phrase only has a relatively low probability of being learned on each occasion that it is encountered in the input, then words are more likely to be learned in contexts in which they occur frequently, making the model more sensitive to the statistical properties of the input language.

Model Data: p=0.1

Method

As before, the model was trained on input data from Anne's mother. In this simulation, the overlap parameter was set to 8%. The reason for this was that when output data from this version of MOSAIC was analysed for case-marking errors (Croker, 2002), 8% proved to be the optimum parameter value. Any lower and too many errors were made, any higher and too few were made. Although the results presented in this paper are not concerned with case-marking errors, this value was retained as we wanted to explain the occurrence of multiple phenomena without parameter fitting. Similarly, the number of times the input corpus was presented before generating an output set was a number arrived at in the analysis of case-marking errors. After 5 presentations of the input corpus, MOSAIC produced 98,533 utterance types, of which 8,974 were produced by recognition and 89,559 by generation. After 10 presentations of the input corpus, MOSAIC produced 187,574 utterance types, of which 15,772 were produced by recognition and 171,802 by generation.

Results

Results for MOSAIC-Anne, with a learning probability of 0.1 are presented in Table 3. After both 5 and 10 presentations of the input corpus, MOSAIC produces 'not + untensed verb' errors at a rate of 10%, consistent with Anne's rate of 11.7% (see Table 1). Similarly, MOSAIC produces 2 'not goes' type errors: 'not goes an egg' and 'not he goes'. Both of these errors are possible as a result of the creation of a generative link between 'not' and 'there' (which explains how a pronoun can follow negation in the second of these errors). After 10 input presentation cycles, this link ceases to exist and so no errors of this type are produced. In both cases, the number of 'not + untensed verb' errors is explained in terms of generative links between first- or second-person pronouns and third-person pronouns. For example, both 'I' and 'you' are linked to 'he' and 'it', enabling utterances such as 'it don't like it' and 'he don't want' to be produced.

Table 3: Negation errors (MOSAIC-Anne, p=0.1).

	MOSAIC-Anne 5 x input	MOSAIC-Anne 10 x input
not go	(50) 10.0	(50) 10.0
goes not	0	0
not goes	(2) 0.4	0
go not	0	0

Discussion

With probabilistic learning, MOSAIC is able to provide an almost exact fit to Anne's data. As theorised, this learning algorithm enables MOSAIC to encode more frequent word combinations and to fail to learn infrequent combinations. As a result, 'go not' type errors are not made – Anne does not produce errors of this type, but MOSAIC did produce such errors when the learning probability was set to 1. This ability to discriminate between items in the input set means that MOSAIC is much more sensitive to the distributional characteristics of the data. With a lower learning probability, the data from MOSAIC ceases to look like an amalgamation of the data from all three children, producing a different pattern of errors from that shown by Aran and Becky.

Conclusion

Harris & Wexler (1996) argue that children produce errors involving the use of untensed verbs after negation, but do not produce errors involving the use of tensed verbs after negation. Analysis of our data shows that as well as forming grammatically correct utterances containing the negative particle, all three children used untensed verbs following negation. In addition, Anne and Becky produced errors in which an inflected verb was used after negation and Aran produced the negative particle *preceded* by an inflected verb – errors that are predicted *not* to occur by Harris & Wexler. All of the MOSAIC simulations produce errors with untensed verbs following negation, and several of them produce errors with tensed verbs after negation. When the probability of learning any new word is set to 1, MOSAIC consistently under-produces 'not go' type errors, although with an overlap of 4%, 'not goes' errors are produced at the same rate as they are produced by Anne. When the probability is reduced to 0.1, MOSAIC produces a lot more 'not go' type errors, and hence provides a good fit to Anne's data. This modification to the model therefore appears to increase its sensitivity to patterns in the input.

It is apparent from the results of this study that Harris & Wexler's account of negation errors is incomplete, which raises doubts about the claim that children have already correctly set all the clause structure parameters of their language. In contrast, MOSAIC is able to predict children's negation errors quite accurately. With a learning probability of 1, MOSAIC predicts the *types* of errors made by all three children, but does not provide an exact quantitative fit. With a learning probability of 0.1, MOSAIC predicts Anne's errors almost exactly, both in terms of error type and error rate. This version of MOSAIC also performs better with respect to case-marking errors and optional infinitive errors than earlier incarnations (Croker, 2002). The data from this simulation also shows that errors can be 'unlearned' over time – 'not goes' type errors are no longer produced after the input has been presented 10 times. Children, of course, cease to produce these errors as they get older. Although we do not attempt to assess the developmental aspects of MOSAIC in any detail in this study, the fact that MOSAIC is able to unlearn this error can be seen as a positive feature of the model.

In conclusion, the simulations reported in this study show that it is possible to model the pattern of negation errors in children's speech without assuming any domain-specific knowledge of linguistic structure. This suggests that the claim that this pattern can be taken as evidence for innate grammatical knowledge on the part of the child is too strong. It also suggests that at least some of the patterning of children's early language can be explained in terms of a distributional analysis of the statistics of the language being learned.

References

Brown, R. (1973). *A first language* . Cambridge, MA: Harvard University Press.

Cartwright, T. A. & Brent, M. R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. Cognition, 62, 121-170

Croker, S., Pine, J. M. & Gobet, F. (2000) Modelling optional infinitive phenomena. In *Proceedings of the Third International Conference on Cognitive Modeling*. Veenendaal: Universal Press.

Croker, S., Pine, J. M. & Gobet, F. (2001) Modelling children's case-marking errors with MOSAIC. In *Proceedings of the 2001 Fourth International Conference on Cognitive Modeling.* Mahwah, NJ: Erlbaum.

Croker, S. (2002). *Modelling optional infinitive phenomena in children's speech.* PhD Thesis, University of Nottingham.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. Cognition, 48, 71-99

Freudenthal, D, Pine, J. M. & Gobet, F. (2001). Modelling the optional infinitive stage in MOSAIC: A generalisation to Dutch. In *Proceedings of the 2001 Fourth International Conference on Cognitive Modeling*. Mahwah, NJ: Erlbaum.

Freudenthal, D., Pine, J. M., & Gobet, F. (2002a). Subject omission in children's language: The case for performance limitations in learning. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Freudenthal, D., Pine, J. M., & Gobet, F. (2002b). Modelling the development of Dutch Optional Infinitives in MOSAIC. In W.D. Gray & C.D. Schunn (Eds.), *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Harris, T. & Wexler, K. (1996). The optional-infinitive stage in child English. In H. Clahsen (Ed.) *Generative perspectives in language acquisition*. Philadelphia: John Benjamins.

Lieven, E.V.M., Pine, J.M. & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187-219.

MacWhinney, B., & Snow, C. (1990). The Child Language Data Exchange System: An update. *Journal of Child Language*, 17, 457-472.

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

Redington, M., Chater, N. & Finch, S. (1993). Distributional information and the acquisition of linguistic categories: a statistical approach. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Redington, M., Chater, N. & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science 22*, 425-469.

Theakston, A. L., Lieven, E. V .M., Pine, J .M., & Rowland, C. F. (2000). The role of performance limitations in the acquisition of 'mixed' verb-argument structure at stage 1. In M. Perkins & S. Howard (Eds.) *New directions in language development and disorders.* New York: Plenum.

Tomasello, M. (2000a). Do young children have adult syntactic competence? *Cognition,* 74 (3), 209-253.

Tomasello, M. (2000b). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences* Vol. 4, No. 4, 156-163.

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: Cambridge University Press.