Modelling Optional Infinitive Phenomena

S. Croker (sfc@psychology.nottingham.ac.uk)
J. M. Pine (jp@psychology.nottingham.ac.uk)
F. Gobet (frg@psychology.nottingham.ac.uk)
School of Psychology, University of Nottingham,
Nottingham, NG7 2RD, England

Abstract

The Optional Infinitive hypothesis proposed by Wexler (1994) is a theory of children's early grammatical development that can be used to explain a variety of phenomena in children's early multi-word speech. However, Wexler's theory attributes a great deal of abstract knowledge to the child on the basis of rather weak empirical evidence. In this paper we present a computational model of early grammatical development which simulates Optional Infinitive phenomena as a function of the interaction between a performance-limited distributional analyser and the statistical properties of the input. Our findings undermine the claim that Optional Infinitive phenomena require an abstract grammatical analysis.

Introduction

The Optional Infinitive (OI) hypothesis proposed by Wexler (1994) is a theory of early grammatical development which attempts to provide a unified nativist account of children's knowledge of verb movement and verb inflection across languages.

In this paper we present a computational model of early grammatical development which simulates OI phenomena in children's early multi-word speech as a function of the interaction between a performance-limited distributional analyser and the statistical properties of mothers' child-directed speech.

The Optional Infinitive Hypothesis

The Optional Infinitive hypothesis (Wexler, 1994) is an attempt to provide a unified nativist account of young children's knowledge of verb movement and verb inflection across a variety of different languages. 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 OI stage — during which they lack the knowledge that tense is obligatory in finite clauses — knowledge which matures at a later stage of development.

When applied to children learning English, the OI hypothesis can be used to explain a variety of phenomena in early multi-word speech. First, it can be used to explain why children sometimes fail to use appropriately tensed forms in finite clauses (e.g. "that go in there" instead of "that goes in there" or "that going in there" instead of "that's going in

there"). Second, it can be used to explain the pattern of subject-verb agreement in children's speech. Thus, even though English-speaking children do not always use tensed forms in obligatory contexts, the tensed forms that they do use tend to agree in person and number with their subjects (e.g. "I am", "she is", but not "I is" or "she are"). Finally, it can be used to explain the occurrence of pronoun case-marking errors (e.g. "her do it" instead of "she does it" and "him get it" instead of "he gets it") and why such errors tend to occur with non-finite rather than finite verb forms (e.g. "her do it" as opposed to "her does it" and "him get it" as opposed to "him gets it").

The OI hypothesis makes very clear predictions about what children in the OI stage should and should not say. Thus, OI children may produce grammatically incorrect utterances with non-finite verb forms (e.g. "she hide") alongside grammatically correct utterances with finite verb forms (e.g. "she hides"). They may also make certain kinds of case-marking errors (e.g. "him like it" and "her did it"), but should not produce agreement errors (e.g. "she are good"), or case-marking errors with agreeing forms (e.g. "her wants it"). It is worth pointing out, however, that these predictions are qualitative rather than quantitative. That is to say, the hypothesis predicts the presence of certain kinds of errors and the absence of others; it does not predict how often the relevant types of error will occur.

An alternative explanation for the phenomena described by Wexler is that the syntactic patterning of children's speech can be explained with reference to the input that the child receives. In the account we propose here, the child's production of errors with particular lexical forms reflects the distribution of these forms, or of forms that behave in similar ways to these forms, in the language to which the child is exposed. This account capitalises on the fact that most of the patterns of performance predicted by the OI hypothesis have plausible models in mothers' child-directed speech, whereas all of the patterns predicted not to occur do not. For example, sequences such as "she going" (analysed by Wexler as untensed verb forms in contexts where a tensed form is obligatory) occur in maternal questions such as "Where is she going?" and "Is she going to the shops?". Similarly, sequences such as "him go' (analysed by Wexler as case-marking errors with untensed verb forms) occur in maternal utterances such as "Look at him go". On the other hand, sequences such as "she am" and "he are" (analysed by Wexler as agreement errors), and sequences such as "him does" and "her goes" (analysed by Wexler as case-marking errors with agreeing verb forms) do not occur as fragments of correctly formed adult utterances.

MOSAIC

MOSAIC (Model Of Syntax Acquisition In Children) is a computational model based on the CHREST architecture (Gobet, 1998). CHREST is a member of the EPAM family (Feigenbaum and Simon, 1984). Variants of CHREST have been used to model a number of areas of human cognition including the acquisition of multiple representations in Physics (Lane, Cheng & Gobet, 1999) and the acquisition of vocabulary (Jones, Gobet & Pine, 1999). Knowledge is modelled in CHREST as a discrimination network, which is a hierarchically structured network consisting of nodes and vertical links between nodes. Each node has an 'image', which contains the information available at this node (in MOSAIC, it consists of information regarding the links traversed to arrive at that node).

Croker, S., Pine, J.M. & Gobet, F. (2000). Modelling optional infinitive phenomena. In N. Taatgen & J. Aasman (Eds.) *Proceedings of the 3rd International Conference on Cognitive Modeling,* 78-85. Veenendaal, Holland:Universal Press.

Two processes are used to grow the network. Discrimination adds a new node to the network, and familiarisation adds information to the node image. The discrimination network is grown as input (in our case, maternal speech) 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, the process of discrimination is used to create a new node corresponding to that word. 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 model) are learned and stored. Test links above nodes refer to the 'test' (one word or a sequence of words) that has to be passed to travel down that link to the node, and are represented as the final element in the image of that node. These are created during discrimination, at the same time as a new node. In cases where nodes only consist of one word, the image of the node matches the test link immediately above it, as that is the only link to have been traversed. However, at deeper levels, the image will contain more information relating to the sequence of tests. As noted above, at their first presentation, all words are encoded as primitives at the first layer of the network; a particular word must be 'seen' again in order for it to occupy a second location in the network. Subsequent words in an utterance are represented as nodes below the primitive, as long as they are already encoded as primitives themselves.
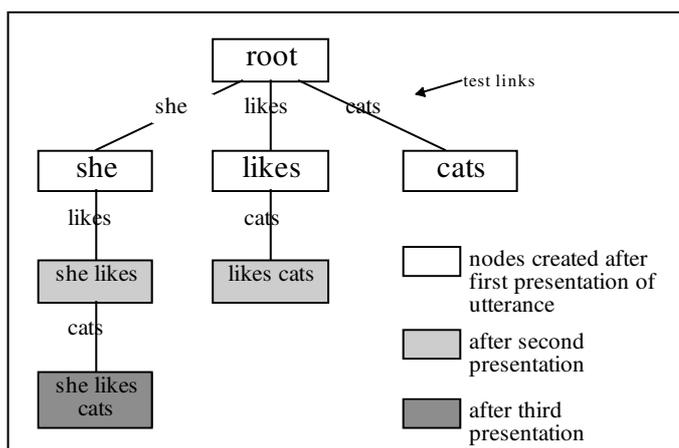
Figure 1: Network formed after the utterance "she likes cats" is presented 3 times.

Figure 1 shows a small network created by presenting the utterance "she likes cats" to the model three times. On the first presentation, the primitives (white nodes) are created. When the model sees the utterance again, the network can be extended as the primitives have already been learned (light grey nodes). The dark grey level 3 "she likes cats" node is created on the third presentation. When an utterance starts with a word already seen by the model, the *image* of the matching node is compared to the utterance. The utterance is then compared at the next level down to see if the second word of the utterance is already in the network below the primitive. The network is followed down as far as possible until one of two possibilities occurs: 1) The entire utterance is already accessible by traversing the

3

Croker, S., Pine, J.M. & Gobet, F. (2000). Modelling optional infinitive phenomena. In N. Taatgen & J. Aasman (Eds.) *Proceedings of the 3rd International Conference on Cognitive Modeling,* 78-85. Veenendaal, Holland:Universal Press.

network; or 2) A point is reached where the utterance can not be traced down the network any further. In this case, discrimination takes place and a new node is created.

Generative Links

As well as learning utterances by rote, MOSAIC is able to generate novel utterances using generative links, an important feature of the model. Generative links are 'horizontal' links between nodes that have contextual similarities. If two words occur frequently in similar contexts (e.g. if they are succeeded by the same items), then a generative or pseudo-category link can be made between these items. These two nodes do not have to be on the same level — a level 2 node can be linked to a level 3 node, for instance. In Figure 2, an example is given in which the model is trained on a data set in which the items "he" and "she" are followed by the verbs "laughs", "likes", "jumps" and "sings". The number of common features needed to create a generative link is a parameter within the model that can be set to any number. In the above example, this parameter has been set to 3; once the model recognises that 2 nodes have at least 3 common test links, a generative link is created. For the purposes of the work outlined in this paper, the parameter was set to 15. It was found that a low parameter setting caused words which had little grammatical similarity to be linked; a higher value resulted in few such links being made (cf. Gobet & Pine, 1997).

Production of Utterances

Once a network has been created, it can be used to produce utterances in two ways: by recognition and by generation. Utterances produced by recognition are essentially rote-
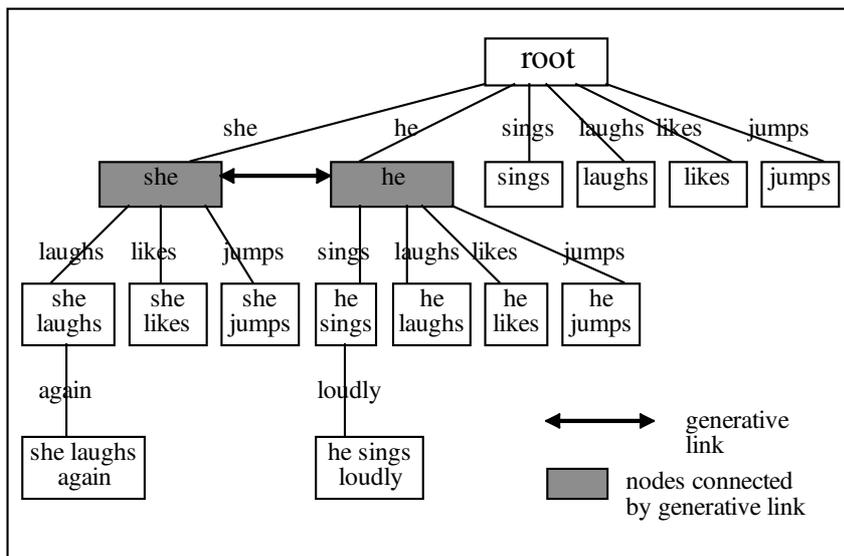
Figure 2: Fragment of a discrimination network showing generative link formation. This network can be used to produce utterances by both generation and recognition

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. For example, from the fragment of a network shown in Figure 2, utterances such as "she laughs again" and "sings loudly" could be produced by recognition.

Croker, S., Pine, J.M. & Gobet, F. (2000). Modelling optional infinitive phenomena. In N. Taatgen & J. Aasman (Eds.) *Proceedings of the 3rd International Conference on Cognitive Modeling,* 78-85. Veenendaal, Holland:Universal Press.

Production by generation utilises the generative links to create utterances not seen in the input. This occurs in a similar way to production by recognition, the difference being that lateral generative links can be traversed as well as vertical test links, although only one generative link can be followed per generated utterance. Thus, from the network in Figure 2, utterances such as "she sings" and "he laughs again" could be produced by generation. It is possible for some utterances to be produced by both recognition and generation. This happens in cases where there are identical test links below both of the nodes connected by a generative link. For example, "he jumps" can be produced by following the "jumps" link down from the "he" node or by following the generative link to "she" and then taking the "jumps" link below "she".

Methods

Child's Data

In this paper we present data obtained from one child, Anne, taken from the Manchester corpus (Theakston, Lieven, Pine & Rowland, in press) of the CHILDES database (MacWhinney & Snow, 1990). This corpus consists of transcripts of audio recordings made twice every three weeks for a period of 12 months, between the ages of 1;10 and 2;9. There are two half-hour recordings for each session, one made during free play and the other made during structured play. The present analysis is limited to utterances including a verb which also began with one of the following set of third person singular subjects: "he", "he's", "him", "his", "she", "she's", "her", "it", "it's", "that", "that's", "Anne", "Anne's", "Mummy" and "Mummy's". Any duplicate utterances were removed to allow direct comparison with the output of the model, which only contains types. Each utterance was coded according to whether certain errors were present: OI errors (e.g. "he have", "it go"), case-marking errors (e.g. "her sit at the table") and other tensing or agreement errors (e.g. "he's can", "she were"). The latter are errors where a tensed form is used but the utterance is not grammatically correct. This last category is a somewhat 'catch-all' category as it is not always clear how such errors should be analysed within Wexler's formalism.

There are two important points to be made about the data from the transcripts. First, words used as verbs may often also be used as instances of other syntactic categories. For the purpose of this research, words were classified as verbs if they appeared as verbs on 90% or more of occasions in the mother's speech corpus. Second, the data used in analysing both human performance and the performance of the model consist of types, not tokens. Much of the research on children's speech is based on analyses using tokens. However, we have found it necessary to use types as the model does not produce multiple instances of utterances in the same way as the child.

Model's Data

The model was presented with input data from Anne's mother taken from the same sessions as Anne's speech. This was a very large corpus of naturalistic input, consisting of 33,390 utterances. After the model had been trained, a list of all the utterances that the model was capable of producing, both by recognition and generation, was obtained. As with the child data, the model's output was reduced to those utterances beginning with a third person

Croker, S., Pine, J.M. & Gobet, F. (2000). Modelling optional infinitive phenomena. In N. Taatgen & J. Aasman (Eds.) *Proceedings of the 3rd International Conference on Cognitive Modeling,* 78-85. Veenendaal, Holland:Universal Press.

singular subject + verb pair. The model produced more of these types of utterances than the child, so random samples matching the number of utterances in Anne's data were taken for analysis. Two samples of 555 utterances were taken and coded for errors in the same way as the child's utterances.

Results

Certain error types are predicted to occur in the OI hypothesis, including non-finite verb form (or OI) errors, and case-marking errors with untensed verb forms or with tensed verb forms that do not carry agreement. Other error types, including agreement errors and case-marking errors with agreeing verb forms are predicted not to occur. There are also certain kinds of errors about which no clear predictions are made, in particular overtensing errors (e.g. "he didn't went"). The literature in the field (see Harris & Wexler, 1996; Schutze & Wexler, 1996; Rispoli, 1998) reports instances of all of these error types in children's speech. MOSAIC also produces instances of all of these error types (see Table 1). Thus, although the model embodies a very simple learning mechanism, it captures aspects of the data which, at best, the OI hypothesis cannot explain and which, at worst, count directly against it.

Figure 3 shows that the errors made by the model after presentation of the full input corpus occur with a similar frequency to those made by the child during the latter half of the period during which taping took place.

Table 1: Types of error predicted in the OI hypothesis and their occurrence in children's speech and MOSAIC.

	examples	predicted by Wexler	occurs in child speech	occurs in the model
OI errors	"that go there"	√	√	√
agreement errors	"he are big"		√	√
overtensing errors	"that didn't went down"		√	√
case marking errors with untensed verb	"her sit by herself"	√	√	√
case marking errors with tensed verb	"him did it"	√	√	√
case marking errors with agreement	"her does it"		√	√

Croker, S., Pine, J.M. & Gobet, F. (2000). Modelling optional infinitive phenomena. In N. Taatgen & J. Aasman (Eds.) *Proceedings of the 3rd International Conference on Cognitive Modeling,* 78-85. Veenendaal, Holland:Universal Press.
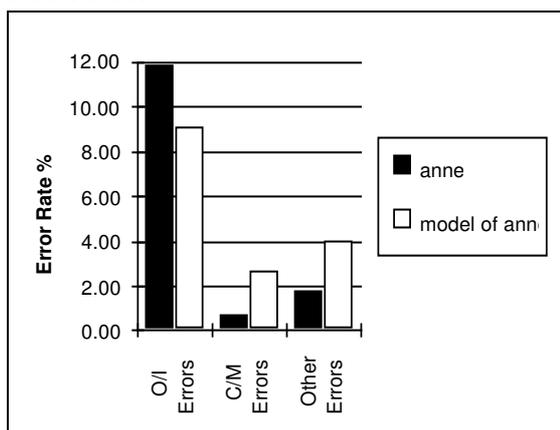
Figure 3: Comparison of the error rates shown by the model and the child.

As discussed above, MOSAIC can produce utterances either by recognition or generation. Utterances produced by recognition are essentially rote-learned. However, the model can also produce novel utterances by traversing generative links. Some of the errors produced by the model have been rote-learned directly from the input (e.g. "he look away" can be learned from "Did he look away?" and "him do it" can be learned from "Let him do it"). However, the model also produces instances of these kinds of error by generation. Other errors such as "her does" and "he am" can only be produced by generation. "Her does" type errors are especially interesting in this respect because they are a particular problem for Wexler's theory. Such errors are produced because the model has formed a generative link between "her" and "that" on the basis that these lexical items occur before overlapping groups of nouns and non-finite verbs in the input.

Discussion

This study utilises a computational model to implement a low-level theory of OI phenomena in English. The output of the model trained on a set of maternal utterances was compared to the output of the child. The results show that after being trained on the input corpus once, the model was able to produce OI errors , other tense/agreement errors and case-marking errors in similar proportions to the child. Some of these errors were produced by *recognition* (i.e. by virtue of being present in the input corpus). For example, OI errors can be picked up directly from questions such as: "Did he go?" and case-marking errors can be picked up directly from double verb constructions such as "I saw her go". However, not all errors have direct models in the input. These kinds of errors were produced by *generation*, an important feature of the model which provides us with an explanation of utterances which cannot be produced by rote-learning. These findings suggest that Wexler's account of OI phenomena in which the child is credited with underlying knowledge of 'tense' and 'agreement' may be an overinterpretation of the data.

MOSAIC produces many of the kinds of errors made by young children, some of which the OI hypothesis fails to correctly predict. Moreover, while the model makes too many overtensing, agreement and case-marking errors, these errors still occur relatively infrequently. In cases where the OI hypothesis does correctly predict the presence of errors it does not predict how often the relevant error types will occur.

Croker, S., Pine, J.M. & Gobet, F. (2000). Modelling optional infinitive phenomena. In N. Taatgen & J. Aasman (Eds.) *Proceedings of the 3rd International Conference on Cognitive Modeling,* 78-85. Veenendaal, Holland:Universal Press.

Conclusion

Overall, the output of the model approximates the data from the child's speech reasonably well. Although it does not provide a precise *quantitative* fit, the model is able to produce all the phenomena predicted by the OI hypothesis, together with some phenomena that the Optional Infinitive hypothesis fails to predict, and therefore cannot explain. MOSAIC thus provides data consistent with a low-level account of OI phenomena and illustrates how these phenomena could arise as a function of the interaction between a performance limited distributional analyser and the statistical properties of the input received by the child.

References

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

Gobet, F. (1998) Memory for the meaningless: How chunks help. *Proceedings of the 20th Meeting of the Cognitive Science Society* pp.398-403. Mahwah, NJ: Erlbaum.

Gobet, F. & Pine, J.M. (1997) Modelling the acquisition of syntactic categories. *Proceedings of the 19th Annual Meeting of the Cognitive Science Society*. pp.265-270 Hillsdale, NJ: Erlbaum.

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

Jones, G., Gobet, F., & Pine, J.M. (1999) Modelling vocabulary acquisition: An explanation of the link between the phonological loop and long-term memory. (CREDIT Technical Report No. 61, University of Nottingham).

Lane, P.C.R., Cheng, P. C-H., & Gobet, F. (1999) Learning perceptual schemas to avoid the utility problem. In *Proceedings of the 19th SGES International Conference on Knowledge Based Systems & Applied Artificial Intelligence.* Cambridge, UK: Springer-Verlag.

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

Rispoli, M. (1998) Patterns of pronoun case error. *Journal of Child Language*, 25, 533-544.

Schutze, C. & Wexler, K. (1996) Subject case licensing and English root infinitives. In A. Stringfellow, D. Cahma-Amitay, E. Hughes & A. Zukowski (Eds.) *Proceedings of the 20th Annual Boston University Conference on Language Development.* pp.670-681. Somerville, MA: Cascadilla Press.

Theakston, A.L., Lieven, E.V.M., Pine, J.M., & Rowland, C.F. (in press) 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.* Plenum

Wexler, K. (1994) Optional infinitives, head movement and the economy of derivations in child grammar. In D. Lightfoot & N. Hornstein (Eds.) *Verb movement* . pp.305-350. Cambridge, MA: Cambridge University Press.