Cognitive Science 30 (2006) 277–310

Modeling the Development of Children's Use of Optional Infinitives in Dutch and English Using MOSAIC

Daniel Freudenthal[a], Julian M. Pine[a], Fernand Gobet[b]

[a]*School of Psychology, University of Liverpool*
[b]*School of Social Sciences and Law, Brunel University*

Received 7 June 2004; received in revised form 11 July 2005; accepted 11 July 2005

Abstract

In this study we use a computational model of language learning called *model of syntax acquisition in children* (MOSAIC) to investigate the extent to which the optional infinitive (OI) phenomenon in Dutch and English can be explained in terms of a resource-limited distributional analysis of Dutch and English child-directed speech. The results show that the same version of MOSAIC is able to simulate changes in the pattern of finiteness marking in 2 children learning Dutch and 2 children learning English as the average length of their utterances increases. These results suggest that it is possible to explain the key features of the OI phenomenon in both Dutch and English in terms of the interaction between an utterance-final bias in learning and the distributional characteristics of child-directed speech in the 2 languages. They also show how computational modeling techniques can be used to investigate the extent to which cross-linguistic similarities in the developmental data can be explained in terms of common processing constraints as opposed to innate knowledge of universal grammar.

Keywords: Syntax acquisition; Optional infinitives; Computational modeling; MOSAIC

1. Introduction

Children acquiring their native language are faced with a task of considerable complexity that they must solve using only noisy and potentially inconsistent input. Mainstream linguistic theory has addressed this *learnability problem* by proposing the nativist theory that children come into the world with rich domain-specific knowledge of language (e.g., Chomsky, 1981; Pinker, 1984). However, recent research using computational modeling techniques suggests that the amount of information that can be extracted from an analysis of the statistical properties of the input language is considerably greater than has traditionally been assumed by the nativist approach (Cartwright & Brent, 1997; Elman, 1993; Redington, Chater, & Finch, 1998;

Correspondence should be addressed to Daniel Freudenthal, School of Psychology, University of Liverpool, Liverpool, United Kingdom L69 7ZA. E-mail: D.Freudenthal@Liverpool.ac.uk

Solan, Horn, Ruppin, & Edelman, 2003). Moreover, there is now a substantial body of research on infants' distributional learning abilities that suggests that children are sensitive to the statistical properties of the language they are learning from a very early age (Gomez & Gerken, 1999; Saffran, Aslin, & Newport, 1996; Santelmann & Jusczyk, 1998; Shady & Gerken, 1999).

When taken together, the results of these different lines of research raise the possibility that some of the phenomena that have been taken as evidence for innate linguistic knowledge on the part of the child could be explained in terms of the interaction between the child's distributional learning abilities and the statistical properties of the input. One obvious way of investigating this possibility is to use computational modeling techniques to simulate developmental phenomena as a function of a distributional analysis of the language to which children are exposed. However, as Christiansen and Chater (2001) pointed out, a major challenge facing this kind of approach is to develop models that map more directly onto the tasks that human language learners face, use more representative input samples, and are able to make greater contact with the empirical data than has tended to be the case in the past.

In this article, we describe the model of syntax acquisition in children (MOSAIC), a computational model of language learning that attempts to meet this challenge by simulating developmental changes in the characteristics of children's multiword speech as a function of a distributional analysis of mothers' child-directed speech. MOSAIC is a simple resource-limited learning mechanism that takes corpora of orthographically transcribed child-directed speech as input and learns to produce progressively longer utterances that can be directly compared with the speech of children at particular points in development. We show how MOSAIC simulates the developmental patterning of the optional infinitive (OI) phenomenon in two different languages (English and Dutch). The OI phenomenon was chosen as a target for simulation for two reasons. The first is that it occurs in a number of languages that differ considerably in their underlying grammars (Wexler, 1994). It thus constitutes a useful domain in which to explore the extent to which cross-linguistic similarities in the patterns of behavior shown by children learning different languages can be understood in terms of common processing constraints as opposed to innate knowledge of universal grammar. The second is that it shows considerable variation in its developmental characteristics across different languages (e.g., Ingram & Thompson, 1996; Wijnen, Kempen, & Gillis, 2001; Wilson, 2003). Simulating such differences as a function of cross-linguistic differences in the statistical properties of the input is therefore a good way of investigating the relation between the distributional properties of children's early language and the distributional properties of the language to which they are exposed.

1.1. The OI phenomenon

It is a well-established fact that young English-speaking children often fail to provide grammatical morphemes in contexts in which they are obligatory in adult speech (Brown, 1973). For example, between the ages of 2 and 3 years, children learning English produce utterances such as those in 1, 2, 3, and 4:

1. Dolly want a drink
2. Yesterday we go to the park

3. Doggie barking
4. Mummy done it

Utterance 1 is an example of missing third-person singular present tense, Utterance 2 an example of missing past tense, Utterance 3 an example of missing auxiliary "be," and Utterance 4 an example of missing auxiliary "have."

Traditionally, such utterances have been interpreted in terms of the gradual acquisition of the relevant morphemes (L. Bloom, Lifter, & Hafitz, 1980; Brown, 1973), or the dropping of inflections as a result of performance limitations in production (P. Bloom, 1990; Valian, 1991). More recently, however, Wexler (1994, 1998) argued that, rather than reflecting a process of inflection drop, utterances such as those in 1 to 4 reflect children's optional use of nonfinite forms (e.g., "want," "go," "barking," "done") in contexts in which a finite form (e.g., "wants," "went," "is barking," "has done") is obligatory in the adult language.[1] This view has come to be known as the OI hypothesis and has generated a great deal of research over the past 10 years in a variety of different languages.

According to the OI hypothesis, by the time children begin to produce multiword speech, they have already correctly set all the basic inflectional and clause structure parameters of their language. They thus have adultlike knowledge of the word order and inflectional properties of the language they are learning. However, there is a stage of development (the OI stage) during which the abstract features of Tense (TNS) and Agreement (AGR) can be absent from the underlying representation of the sentence. This results in children initially using both finite and nonfinite verb forms in contexts in which a finite form would be obligatory in the adult language and producing errors such as those in 1, 2, 3, and 4. However, because children have adultlike knowledge of the word order and inflectional properties of the language, their use of finite and nonfinite forms tends to pattern correctly with respect to other aspects of the grammar. Thus, although English-speaking children produce OI errors with the negative particle (e.g., "he not go" instead of "he doesn't go"), they rarely produce errors in which a finite or nonfinite verb form is incorrectly placed with respect to the negative particle (e.g., "he goes not" or "he not goes"). Similarly, although English-speaking children produce OI errors with subject pronouns (e.g., "he go" instead of "he goes"), they rarely produce errors in which a finite verb form fails to correctly agree with the sentence subject (e.g., "I goes" instead of "I go") (Harris & Wexler, 1996).

The great strength of the OI hypothesis is that it provides a unified account of patterns of finiteness marking in children acquiring a variety of different languages. Thus, children produce OI errors not only in English, where the infinitive is a zero-marked form, but also in a number of other languages, including Dutch, German, French, and Swedish, where the infinitive carries its own infinitival marker. For example, children learning Dutch produce utterances such as "Pappa eten" ("Daddy eat-INF") and "Mamma drinken" ("Mummy drink-INF") where the correct forms would be "Pappa eet" ("Daddy eats") and "Mamma drinkt" ("Mummy drinks"). Moreover, children's use of finite and nonfinite verb forms in these languages tends, as in English, to respect the word-order patterns of the language being learned. Thus, although children learning Dutch use both finite and nonfinite verb forms in finite contexts, they nevertheless tend to correctly place finite verb forms before their complements and nonfinite verb forms after their complements and, hence, produce finite utterances such as: "Pappa eet appels" ("Daddy eats apples") and nonfinite utterances such as "Pappa appels eten" ("Daddy apples eat-INF").

However, although the OI hypothesis is able to explain a number of regularities in the patterning of children's early multiword speech across different languages, it also has two important weaknesses. The first of these is that it attributes a great deal of highly structured linguistic knowledge to the child on the basis of rather weak empirical evidence. Thus, although children's tendency to position finite and nonfinite verb forms correctly in their utterances is consistent with the claim that young children already know the word order and inflectional properties of their language, it is not at all clear that this is the only possible explanation of the data. Indeed, as Joseph and Pine (2002) point out, any learning mechanism sensitive to the relation between finiteness marking and utterance position in the input is likely to capture this effect.

The second weakness is that the theory makes very limited quantitative predictions about the developmental patterning of children's use of finite and nonfinite verb forms in the different languages to which it has been applied. Thus, the theory predicts that there will be a stage in which children will produce both finite and nonfinite verb forms in finite contexts and that children's use of nonfinite forms in finite contexts will decrease over time, until such errors eventually disappear from their speech. However, it has little to say about the rate at which children will make OI errors at particular points in development, or about differences in the rate at which children will make OI errors with different kinds of verbs. For example, it provides no explanation of why there is a stage in early child Dutch when finite verb forms are virtually absent from the child's speech (Wijnen & Verrips, 1998), or why, when Dutch children do begin to use finite verb forms more frequently, the verbs that they use in finite and nonfinite form tend to come from different populations with different distributional characteristics in the input (Jordens, 1990; Wijnen, 1998).

In view of these weaknesses, this study represents an attempt to investigate the extent to which it is possible to explain the OI phenomenon in terms of a simple resource-limited distributional analysis of the language to which children are exposed. This is done by using MOSAIC to simulate developmental changes in children's use of finite and nonfinite verb forms in two languages (English and Dutch) with considerable differences in their underlying grammars. MOSAIC is a simple distributional learning mechanism with no built-in domain-specific knowledge that takes as input corpora of orthographically coded child-directed speech and learns to produce as output utterances that become progressively longer as learning proceeds. As a result of these characteristics, MOSAIC can be used to model the behavior of children learning different languages across a range of different mean length of utterance (MLU) values. The aim of this study is therefore to use MOSAIC to investigate whether it is possible to simulate the developmental patterning of finiteness marking in English and Dutch as a function of the interaction between a single learning mechanism and the distributional characteristics of English and Dutch child-directed speech.

1.2. MOSAIC

MOSAIC is an instance of the CHREST architecture, which in turn is a member of the EPAM (elementary perceiver and memorizer) family (Feigenbaum & Simon, 1984). CHREST models have been used successfully to simulate phenomena such as novice–expert differences in chess (Gobet & Simon, 2000) and memory for computer programs (Gobet & Oliver, 2002), as well as phenomena in diagrammatic reasoning (Lane, Cheng, & Gobet, 1999) and language acquisition (Croker, Pine, & Gobet, 2000, 2001; Freudenthal, Pine, & Gobet, 2001, 2002a, 2002b; Gobet et al., 2001; Gobet & Pine, 1997; Jones, Gobet, & Pine, 2000a, 2000b).

The basis of the model is an n-ary discrimination network, consisting of nodes connected by test links. The network is headed by a root node that has no contents. The other nodes in the network encode words or phrases. Test links encode the difference between the contents of two nodes. MOSAIC employs two learning mechanisms. The first mechanism, based on discrimination, adds new nodes and test links to the network in a probabilistic fashion. The second mechanism, based on similarity, adds "generative" links between nodes encoding phrases encountered in similar contexts. In its present form, the model learns from text-based input. That is, the model assumes that the phonological stream has been segmented into words.

1.2.1. Adding new nodes to the network

The model encodes utterances by parsing them in a left-to-right fashion. When the network is given input, it creates new nodes under the root node. We call the nodes just under the root *primitive nodes.* When additional input is received, new nodes at increasingly deeper levels are created. The model therefore encodes phrases of increasing length. This process is now illustrated by an example. To keep this example simple, we assume that a node is created with a probability of 1 (and hence that nodes for sentence-initial words and phrases are as likely to be created as words for sentence-final words and phrases). In fact, the probability of creating nodes, which depends on a number of factors, is quite low (see Section 1.2.3.), and there is a bias in the model to create nodes for sentence-final words and phrases. This means that, in practice, nodes for sentence-final words and phrases are often created before nodes for sentence-initial words and phrases.

1.2.2. An example

Assume that an empty network receives the utterance "did he go." The model first encounters the word *did.* Because the network is empty, there are no test links. The model therefore creates a test link and a node under the root node. Essentially, the model is *learning* the word. The new node has *did* as its contents, as does the test link. The model then proceeds to the next part of the input (i.e., *he*). When the model now examines the test links from the root node, it has only one link that encodes the word *did.* Because this is different from the word it is considering, the model now creates a second test link and node under the root node, which encode the word *he.* In the same way, it subsequently creates a node and link that encode the word *go.* Fig. 1 shows the network after one presentation of this input.

When the model sees the same sentence again, it first encounters the word *did.* When considering the test links from the root node, it finds a link encoding the word *did;* the model *recognizes* this word. The model now follows the test link and moves to the next part of the input (*he*). When encountering *he,* the model considers the test links emanating from the "did" node. As there are no such test links, and as *he* has already been learned as a primitive, the model creates a new test link and node under the "did" node. The test link encodes the word *he,* and the node encodes the phrase *did he.* When the model has created a new node, it moves to the next part of the input, and starts from the root node again. It now recognizes the word *go,* but does not learn, because there is no input remaining.

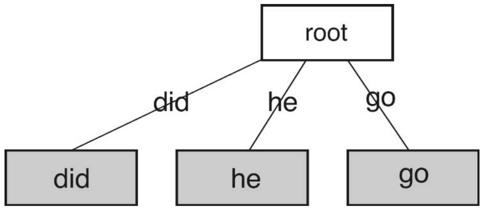

Fig. 1. MOSAIC after it has seen the utterance "did he go" once.

When the model sees the same sentence a third time, it parses the utterance until reaching the "did he" node and finds that there is no "go" link. It now creates a "go" link under the "did he" node, thus encoding the fact that it has seen the utterance "did he go." At this point it also copies the information that "he" has been followed by "go" into the primitive node "he." On this pass, the model has thus encoded the fact that the phrase *did he* has been followed by the word *go,* as well as the fact that the word *he* has been followed by the word *go.* Fig. 2 shows what the network looks like at this point. The model has now seen the utterance three times. On the first pass, it has created the three primitive nodes. On the second pass it has created "he" under "did". On the third pass it has created "go" under "did he," as well as "go" under "he."

Suppose the model now sees the phrase *he walks.* It first recognizes the word *he.* When it comes to the word *walks,* it tries to create a new test link under *he.* However, there is no primitive "walks" node and therefore the model creates one. When encountering the phrase *he walks* again, it creates the test link "walks" (and node "he walks") underneath the "he" node. At this point, the "he" node has two test links, encoding the fact that "he" has been followed by "go" and "walks." Fig. 3 shows the network at this point.

Although, in the example so far, consecutive nodes differ by only one word, the model can also treat larger phrases as one unit. If the model in Fig. 3 were to learn the word *does* and then sees *does he go,* it can create a "does he go" node underneath the "does" node. Because the

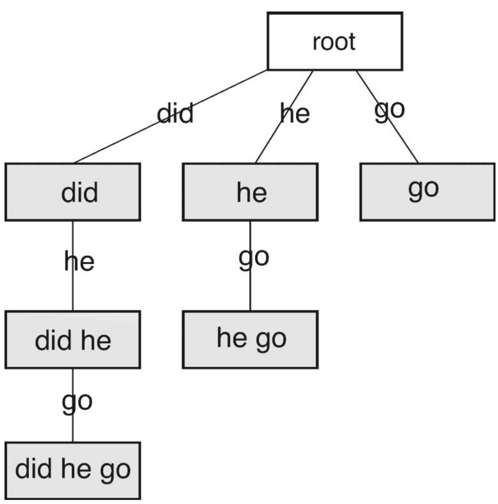

Fig. 2. MOSAIC after it has seen the utterance "did he go" three times.

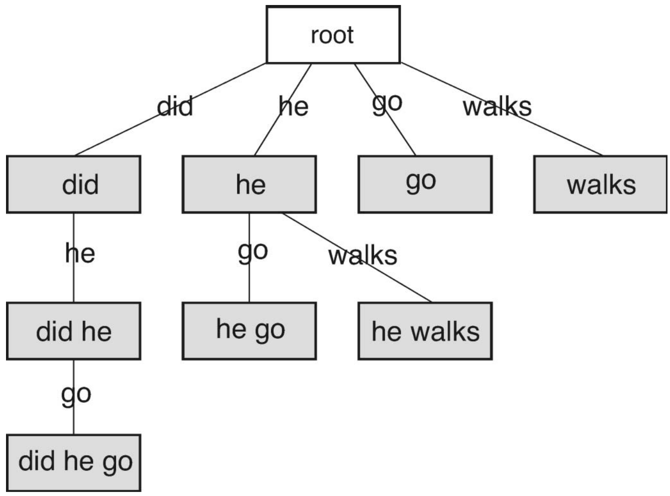

Fig. 3. MOSAIC after it has seen the utterance "did he go" three times and the utterance "he walks" twice.

model already has a node encoding the phrase *he go,* it can recognize this phrase as one unit. This mechanism (known as chunking) enables the model to learn frequent phrases quickly. Because the chunking mechanism does not appear to play an important role in the simulations reported here, it is not discussed any further in this article.

1.2.3. Probability of creating a node

So far, we have assumed that nodes are created whenever there is an opportunity to do so. In fact, however, the creation of nodes is directed by a *node-creation probability* (NCP) that can vary between 0 and 1. When this probability is equal to 1, a node is always created when the opportunity arises, as in our example. However, when the NCP is less than 1, a node may or may not be created; the outcome is decided by the model generating a random number between 0 and 1 and determining whether or not this number is less than or equal to the NCP. Thus, the lower the NCP, the less likely it is that a node will be created.

Making node creation probabilistic in this way has two important consequences. First, using lower NCP values reduces the rate at which the model learns from its input and hence prevents it from learning to produce long utterances too quickly. Second, using lower NCP values makes the model more frequency sensitive. Thus, because nodes are no longer created whenever the opportunity arises, nodes for high-frequency items are more likely to be created than nodes for low-frequency items because opportunities to create nodes for high-frequency items arise more frequently (and the probability of generating a random number that is less than or equal to the required NCP in several attempts is obviously higher than the probability of generating such a number in one attempt).

To simulate the range of MLUs displayed by young children, and generate enough output to perform meaningful analyses, a decision was made to set the NCP to gradually increasing values as a function of the size of the network (i.e., to make it easier to create nodes as the size of the network increases). This decision is consistent with empirical data showing that children learn new words more readily as their vocabulary size increases (Bates & Carnavale, 1993). A further constraint on learning is the relative ease with which phrases of various lengths can be learned. In these simulations, nodes that encode long phrases have a lower likelihood of being created.

The specific formula for calculating the NCP is the following:

$$\mathrm{NCP} = (N/M)^L \tag{1}$$

M is a parameter arbitrarily set to 50,000 for Dutch and to 70,000 for English,[2] N = number of nodes in the network ($N \leq M$), and L = length of the phrase being encoded (in words).

According to the previously mentioned formula, NCP depends on the number of nodes in the network. If this number is small, the first term (the ratio N/M) is small and the resulting NCP is small. Thus, in a small network, learning is slow. When the number of nodes increases, NCP increases as well, and the learning rate goes up. The first term of the equation is raised to the power of L (the length of the phrase being encoded for this node).[3] This exponent makes it more difficult to create nodes that encode longer utterances. However, as the number of nodes in the network increases, the relative difficulty of encoding longer utterances decreases. When the first term approaches 1, the weight of the exponent diminishes, because the ratio of .9 raised to the power 2 or 3 is smaller than the ratio of .1 raised to the power 2 or 3.

1.2.4. Creation of generative links

The second type of learning used by MOSAIC is the creation and removal of *generative* links. Generative links are created between phrases[4] that share a certain percentage overlap between both the preceding and following context. Because new nodes are constantly created in the network, the percentage overlap between two phrases is likely to vary over time. As a result, the percentage overlap between two nodes may drop below the threshold and the generative link be removed. Thus, unlike nodes, generative links can be unlearned.

The rationale behind generative links is the following. When two words belong to the same *word class,* they are likely to take the same position in the sentence and, hence, to be preceded and followed by similar kinds of words. For example, in English, nouns are likely to be preceded by articles and adjectives and followed by verbs. MOSAIC will pick up on this similarity by linking words that are preceded and followed by overlapping sets of words. Note that MOSAIC does not know anything about the class of nouns as a linguistic construct; it only knows that there are words that tend to take the same position relative to other words in the sentence. The development of a class of nouns is thus an emergent property of MOSAIC's distributional analysis of the input.

The percentage overlap between nodes necessary to create a generative link is an important parameter in MOSAIC. A typical value for this parameter is 10%. Setting this parameter to a lower value results in more generative links being created. Setting it to a higher value results in fewer generative links being created.

1.2.5. Producing utterances

MOSAIC produces output by traversing the network from the root node and outputting the contents of the test links. If the model only traverses test links, the utterances it produces must have been present in the input either as entire utterances or as utterance fragments (and can be seen as *rote-learned* utterances). However, MOSAIC is also able to traverse generative links

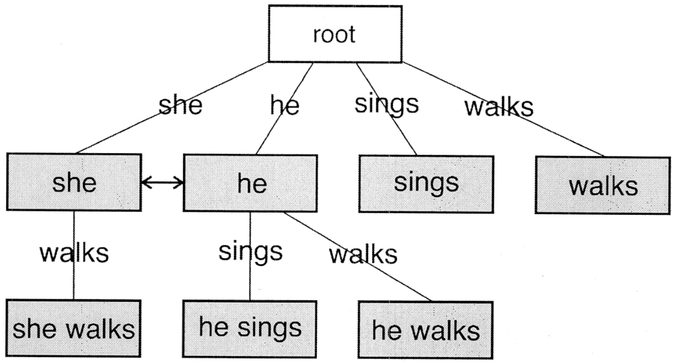

Fig. 4. Generating an utterance. Because *she* and *he* have a generative link, the model can output the novel utterance "she sings." (For simplicity, preceding context is ignored in this figure.)

during production. When the model traverses a generative link, it is able to supplement the utterance produced up to that point with the contents of the test links following the generative link. It is thus able to produce novel or *generated* utterances. Fig. 4 illustrates the mechanism by which generated utterances are produced.

Note that because there is a distinction in the model between test links and generative links, it is possible to separate the model's output into utterances that were produced by traversing only test links and utterances that were produced by traversing test and generative links. This makes it possible to distinguish between rote-learned and generated utterances in the model's output.

1.2.6. MOSAIC's utterance-final constraint

Although MOSAIC could, in principle, output everything that has been encoded during learning, there is an important restriction on what the model does actually output. Thus, an utterance is produced only if the final word in the utterance was the final word in an utterance in the input when it was encoded. This information is encoded in the model by adding an *end marker* to utterance-final phrases as they are learned. Using the end marker to constrain the model's production prevents MOSAIC from outputting utterances that end "halfway through the sentence" and thus ensures that MOSAIC's output consists entirely of utterance-final phrases and utterances produced by substituting a word into an utterance-final phrase across a generative link. Although technically part of the production mechanism, MOSAIC's utterance-final constraint is intended as a way of implementing a conceptual model of language learning that involves gradually building up representations of sentence structure from the right edge of the sentence (i.e., learning progressively longer utterance-final sequences and frames). Several authors have argued, on the basis of empirical data, that children are particularly sensitive to material in utterance-final position. That is, children comprehend and learn words and phrases that occur in utterance-final position more readily than words and phrases that occur in other positions (Naigles & Hoff-Ginsberg, 1998; Shady & Gerken, 1999). It has also been argued that this kind of utterance-final bias in learning may provide a plausible explanation of the OI phenomenon itself (at least in languages such as Dutch and German). For example, Wijnen et al. (2001) argued that the high proportion of OI errors in early child Dutch can be at least partly explained by the fact that in Dutch nonfinite verb forms are restricted to sentence-final position and are hence easier for children to learn than finite verb forms, which tend to occur earlier in the sentence.

1.3. The OI phenomenon in Dutch and English and a preview of the simulations

The central aim of this study is to investigate the extent to which it is possible to explain the OI phenomenon in Dutch and English as a function of a resource-limited distributional analysis of the language to which children are exposed. That is to say, to what extent is it possible to explain the OI phenomenon by assuming that children are learning progressively longer sequences and lexically specific frames from the input in a way that is limited both by a restriction on the amount of information that can be learned from one exposure of any particular utterance and by a constraint on the order in which information from different utterance positions is learned. In the following sections we describe the central characteristics of the OI phenomenon in Dutch and English together with those features of the model that play an important role in allowing MOSAIC to simulate these characteristics. Only later do we present detailed simulations of the phenomenon.

1.3.1. OIs in Dutch

Dutch is a subject–object–verb/verb second language, which means that, in main clauses, finite verb forms take second position and precede their complements, whereas nonfinite verb forms take sentence-final position and are preceded by their complements.

5. *Ik wil een koekje* (I want a biscuit)
6. *Ik wil de bal trappen* (I want the ball kick-INF)

For example, in Sentence 5, the verb *wil* is finite and precedes its complement *een koekje,* and in Sentence 6, the verb *trappen* is nonfinite, and is preceded by its complement *de bal.*

Dutch is a particularly good example of an OI language because, in contrast to English, where the infinitive is a zero-marked form, Dutch has an infinitival morpheme that makes the Dutch infinitive readily distinguishable from the first, second, and third-person singular present tense verb forms (although not from the first, second, and third-person plural present tense verb forms) (see Table 1 for examples of the conjugation of a regular Dutch and a regular English verb).

During the OI stage, Dutch children produce utterances such as those in 7 and 8. Utterance 7 is an example of a correct finite verb form in a main clause. Utterance 8 is an example of an incorrect nonfinite verb form in a main clause (i.e., an OI error).

7. *Hij werkt* (He works)
8. *Hij werken* (He work-INF)

Although Dutch-speaking children produce both finite and nonfinite verb forms in main clauses, they tend to do so in a way that respects the word-order pattern of their language. Thus,

Table 1
Examples of the conjugation of a regular English and a regular Dutch verb

	Dutch		English	
	Present Tense	Past Tense	Present Tense	Past Tense
1st singular	Ik werk	Ik werkte	I work	I worked
2nd singular	Jij werkt	Jij werkte	You work	You worked
3rd singular	Hij werkt	Hij werkte	He works	He worked
1st plural	Wij werken	Wij werkten	We work	We worked
2nd plural	Jullie werken	Jullie werkten	You work	You worked
3rd plural	Zij werken	Zij werkten	They work	They worked
Infinitive	Werken		Work	
Past participle	Gewerkt		Worked	

when they attempt to express relations between verbs and their complements, they tend to produce utterances such as those in 9 and 10.

9. *Wil een koekje* (Want a biscuit)
10. *De bal trappen* (The ball kick-INF)

Utterance 9 consists of a finite verb form *wil* followed by its complement *een koekje,* whereas utterance 10 consists of a nonfinite verb form *trappen* preceded by its complement *de bal.*

Detailed developmental data on the prevalence of OI errors in Dutch are provided by Wijnen et al. (2001). They analyzed the speech of 2 Dutch children (Matthijs and Peter) with respect to their use of finite and nonfinite verb forms in main clauses over a period of roughly 18 months. During this time, the children moved from the *one-word stage,* through stages dubbed the *early two-word stage* and *OI stage* to the *end stage.* Wijnen et al.'s analysis shows that the children displayed a developmental pattern in which the proportion of OIs decreased from roughly 90% during the early two-word stage to less than 10% during the end stage. It also shows that the children's speech during the early stages was very different from that of their mothers. Thus, root infinitives (which are permissible in Dutch, but only in certain contexts such as elliptical answers to questions) made up less than 10% of the mothers' utterances with verbs, and *compound finites* (which include a finite verb form in second position and an infinitive in sentence-final position and are therefore plausible models for OI errors) made up only 30% of the mothers' utterances with verbs. Wijnen et al. suggested that these differences could be explained, at least in part, in terms of an utterance-final bias in learning.

1.3.2. OIs in MOSAIC trained on Dutch

MOSAIC simulates developmental changes in children's early multiword speech as the result of a resource-limited distributional analysis of child-directed speech according to which the model learns to reproduce and generate progressively longer strings that have occurred in utterance-final position in the input. In Dutch, this results in OI errors being learned from utterances including compound finites (compound verbs that include a finite and nonfinite verb form). For example, an utterance such as "ijs eten" ("ice cream eat-INF") might be produced

Table 2
Examples of simple-finite and compound-finite verb forms in declaratives and questions in Dutch and English

Dutch	English
Yes–No Questions	Yes–No Questions
Declaratives	Declaratives
Simple Finites	Simple Finites
Hij springt	He jumps
Hij gaat naar het park	He goes to the park
Hij eet ijs	He eats ice cream
Hij legt het boek op de tafel	He puts the book on the table
Compound Finites	Compound Finites
Hij kan springen	He can jump
(He can jump-INF)	
Hij kan naar het park gaan	He can go to the park
(He can to the park go-INF)	
Hij kan ijs eten	He can eat ice cream
(He can ice cream eat-INF)	
Hij kan het boek op de tafel leggen	He can put the book on the table
(He can the book on the table put-INF)	
Yes–No Questions	Yes–No Questions
Simple Finites	Compound Finites
Springt hij?	Does he jump?
(Jumps he?)	
Gaat hij naar het park?	Does he go to the park?
(Goes he to the park?)	
Eet hij ijs?	Does he eat ice cream?
(Eats he ice cream?)	
Legt hij het boek op de tafel?	Does he put the book on the table?
(Puts he the book on the table?)	

when the model outputs the final part of a declarative including a compound finite such as "Ik wil ijs eten" ("I want ice cream eat-INF"). Similarly, a phrase such as "hij eten" ("he eat-INF") might be learned from the final part of a question including a compound finite such as "Wil hij eten?" ("Wants he eat-INF?").

Examples of the way in which simple-finite and compound-finite verb forms pattern in Dutch in different types of declaratives and questions are presented in Table 2 (together with English glosses). It can be seen from these examples that compound finites in Dutch always consist of a nonfinite verb form preceded by a finite main verb or auxiliary. This feature of Dutch means that, when trained on Dutch, MOSAIC will not only produce OI errors, but it will also likely produce proportionally fewer of these errors as the model learns to produce longer utterances (and hence more compound finites).

It can also be seen from the examples in Table 2 that, whereas nonfinite verb forms are restricted to utterance-final position in Dutch, finite verb forms can occur in a range of different utterance positions, in both simple and compound-finite utterances. Thus, finite verb forms can be the last word in the utterance in short intransitive utterances (e.g., "Hij springt"), the penulti-

Table 2
(Continued)

Dutch	English
Compound Finites	
Kan hij springen? (Can he jump-INF?)	Can he jump?
Kan hij naar het park gaan? (Can he to the park go-INF?)	Can he go to the park?
Kan hij ijs eten? (Can he ice cream eat-INF?)	Can he eat ice cream?
Kan hij het boek op de tafel leggen? (Can he the book on the table put-INF?)	Can he put the book on the table?
Wh- Questions	Wh- Questions
Simple Finites	Compound Finites
Waar springt hij? (Where jumps he?)	Where does he jump?
Waar gaat hij naar toe? (Where goes he to?)	Where does he go?
Wat eet hij? (What eats he?)	What does he eat?
Wat legt hij op de tafel? (What puts he on the table?)	What does he put on the table?
Compound Finites	
Waar kan hij springen? (Where can he jump-INF?)	Where can he jump?
Waar kan hij naar toe gaan? (Where can he to go-INF?)	Where can he go?
Wat kan hij eten? (What can he eat-INF?)	What can he eat?
Wat kan hij op de tafel leggen? (What can he on the table put-INF?)	What can he put on the table?

mate word in the utterance in short transitive utterances (e.g., “Hij eet ijs”), or a word that occurs several words to the left of the last word in the utterance in longer transitive and intransitive utterances (e.g., “Hij legt het boek op de tafel” or “Hij gaat naar het park”). This feature of Dutch is likely to interact with the model’s utterance-final bias to increase the size of any developmental effect on the proportion of OI errors in the model’s output. This is because finite verb forms that occur several words to the left of utterance-final position in the input are only likely to be produced by the model at relatively high MLUs.

It is clear from the previously mentioned description that, when trained on Dutch, MOSAIC is likely to produce a relatively high proportion of OI errors early in development that will decrease as the model learns to produce longer utterances (and hence more simple and compound finites). What is less clear, however, is whether MOSAIC will be able to simulate the very pronounced decrease in the proportion of OI errors reported by Wijnen et al. (2001). The answer to this question will depend on a number of factors. However, given the way in which the length of MOSAIC's output increases over time, the most important of these is likely to be the relative frequency with which finite and nonfinite verb forms occur in utterance-final and near-utterance-final positions in the input. The central question addressed in the Dutch simulations is therefore how closely MOSAIC is able to simulate developmental changes in children's provision of OI errors as a function of the interaction between the distributional properties of finite and nonfinite verb forms in Dutch child-directed speech, an utterance-final bias in learning, and the ability to produce longer utterances as development proceeds.

1.3.3. OIs in English

English is a subject–verb–object language in which the position of verbs with respect to their complements is not dependent on the finiteness of the verb.

11. I want a biscuit
12. I want to kick the ball

For example, in Sentence 11, the verb *want* is finite and precedes its complement *a biscuit,* and in Sentence 12, the verb *kick* is nonfinite and precedes its complement *the ball.* Like Dutch-speaking children, English-speaking children go through an OI stage in which they produce both finite and nonfinite verb forms in contexts in which a finite verb form is obligatory. For example, English-speaking children produce utterances such as those in 13 and 14.

13. That goes there
14. That go there

Utterance 13 is an example of a correct finite verb form in a main clause. Utterance 14 is an example of an incorrect nonfinite verb form in a main clause (i.e., an OI error). However, as can be seen from Table 1, the number of unambiguous finite forms in English is relatively low. There is therefore an inherent difficulty in classifying English utterances as finite or nonfinite. For example, if a child produces the utterance "I walk," the verb form *walk* could be either a correct first-person singular finite form or an incorrect infinitive. Similarly if an English-speaking child produces the utterance "He played," the verb form *played* could be either a correct finite past tense form, or an incorrect (nonfinite) past participle (with *has* omitted). As a result of this ambiguity, there is no published quantitative description of the developmental patterning of the OI phenomenon in English that is detailed enough to serve as a benchmark against which to test the performance of MOSAIC. However, it is clear that, like Dutch-speaking children, English-speaking children's ability to provide correct finite verb forms increases with development. One would therefore expect English-speaking children's speech to include an increasing proportion of unambiguously finite forms as development proceeds.

1.3.4. OIs in MOSAIC trained on English

As in Dutch, MOSAIC learns OI errors in English from compound finites and questions including compound finites. For example, an utterance such as "eat ice cream" might be produced when the model outputs the final part of a declarative including a compound finite such as "I want to eat ice cream." Similarly an utterance such as "He sing" might be learned from the final part of a question including a compound finite such as "Can he sing?"

Examples of the way in which simple-finite and compound-finite verb forms pattern in English in different types of declaratives and questions are presented in Table 2. Because compound finites in English consist of a nonfinite verb form preceded by a finite main verb or a finite auxiliary verb, MOSAIC is likely to produce proportionally fewer OI errors as the model learns to produce longer utterances. However, it is also likely that nonfinites will be replaced by compound finites more quickly in English than in Dutch because in English the finite and nonfinite verb forms in compound finites tend to occur closer together in the sentence. The central question addressed in the English simulations is therefore how closely it is possible to simulate the developmental patterning of OI errors in early child English with the same version of the model used to simulate the data on early child Dutch.

2. Simulations: Method

Simulations are run in MOSAIC by training the model on corpora of orthographically transcribed child-directed speech. Given that the majority of publicly available child-directed speech corpora are orthographically rather than phonetically transcribed, MOSAIC's ability to accept such corpora as input obviously has certain advantages. However, it should be noted that the use of orthographically transcribed data also means that MOSAIC is insensitive to information that is not included in this format, such as information about intonation and relative stress. As a result, MOSAIC is unable to simulate aspects of the data that depend on such factors. For example, MOSAIC is insensitive to the difference between stressed and unstressed morphemes and will learn sequences including unstressed function words (e.g., "kick the ball") as readily as sequences of stressed content words (e.g., "Anne likes chocolate").

When training MOSAIC, input corpora are fed through the model several times, and an output file is created after each cycle through the input corpus. Because the network grows as a function of the amount of input to which it has been exposed, the model's ability to produce utterances increases with every cycle through the input corpus, both in terms of the number and length of the utterances that it can produce. The model's output can therefore be analyzed at various stages of development and compared to the speech of children at similar stages of development (i.e., with similar MLUs). Indeed, it is possible to perform identical (automated) analyses on the model's output and the speech of the child on whose input the model has been trained and hence to evaluate directly the degree of correspondence between the model's output and the child's speech.

2.1. Data sets used in these simulations

In this article we present the results of simulations of 2 Dutch- and 2 English-speaking children. The data sets used in the Dutch simulations were those of Matthijs and Peter from the Groningen corpus (Bol, 1996). The data sets used in the English simulations were those of Anne and Becky from the Manchester corpus (Theakston, Lieven, Pine & Rowland, 2001). Both the Groningen and the Manchester corpora are publicly available data sets contained in the CHILDES database (MacWhinney, 2000). The Groningen corpus consists of a series of 1-hr recordings of parent–child interaction made at regular fortnightly intervals over a period of approximately 2 years. Recording started for Matthijs at the age of 1;10 and for Peter at the age of 1;5. The Manchester corpus consists of a series of 1-hr recordings of parent–child interaction made approximately twice every 3 weeks over a period of approximately 1 year. Recording started for Anne at the age of 1;10 and for Becky at the age of 2;0.

2.2. *Running the simulations*

The simulations described in this article were run by repeatedly feeding the relevant child's input corpus through the model and creating an output file after each cycle through the input.

2.3. *Input corpora*

Input corpora were created by extracting all the maternal utterances from a given child's transcripts, concatenating these utterances into a single input file, and removing incomplete utterances (i.e., false starts and interrupted utterances) and unintelligible or partially intelligible utterances (i.e., cases where the transcriber had been unable to identify one or more of the words in the utterance). The input corpora were also (automatically) filtered to remove the following kinds of material: filler words such as *oh* and *um;* the repeated and corrected material in retraced utterances such as "That's a … that's a dog" and "I want … I need a coffee"; and words and sequences tagged onto the end of utterances, such as the vocative in "What would you like, Anne?" and the tag in "You like chocolate, don't you?" (although it should be noted that vocatives and tags were not removed if they occurred as isolated utterances). The input corpora for Matthijs and Peter consisted of approximately 14,000 and 13,000 utterances, respectively. The input corpora for Anne and Becky consisted of approximately 33,000 and 27,000 utterances, respectively. All four corpora contained a wide range of different utterance types, including fully formed sentences, such as those presented in Table 2, single-word utterances, and sentence fragments (provided that they appeared in the original transcripts as complete utterances).

2.4. *Output files*

Output files were created by outputting all of the utterances that the model was able to produce at each point in its development. This included all the rote-learned utterances that the model was able to produce (i.e., all the utterance-final phrases encoded in the model at each point in development), together with all the generated utterances that the model was able to produce (i.e., all the utterances that could be produced by substituting a word into an utterance-final phrase across a generative link).

2.5. *Coding and data analysis*

To assess the overall quality of MOSAIC's output, a randomly selected 500-utterance sample was drawn from the output files of each child's model when the MLU was approximately 3.0. These 500-utterance samples were then coded for errors of commission (i.e., errors that appeared to reflect the use of a word in an inappropriate grammatical context as opposed to errors that appeared to reflect the omission of a word or phrase from an utterance that would have otherwise been considered grammatically correct). Coding of the Dutch samples was performed by the first author (who is a native speaker of Dutch). Coding of the English samples was performed independently by the first and the second author (who is a native speaker of English). The overall level of agreement for the English samples was 95% ($\kappa = .80$).

The results of the error analysis are presented in Table 3 and show that the rate of errors of commission in MOSAIC's output is similar in English and Dutch and relatively low in both cases (less than 20% in all four samples).

To assess the fit between the output of the model and the speech of the child on whose input the model had been trained, four output files were selected from the output files of each child's model on the basis that they had MLUs as close as possible to 1.5, 2.5, 3.0, and 4.0. These output files were then compared with data sets for the relevant child matched as closely as possible for MLU to the corresponding output file. All of the data sets (i.e., the four data sets for each of the 2 Dutch- and 2 English-speaking children and the four data sets for each of their respective models) were analyzed in exactly the same way by applying the following automated procedure. First, each data set was searched for utterances that contained at least one verb form other than the copula. Then, each of the utterances identified in this way was classified as a simple-finite, a compound-finite, or a nonfinite utterance. Note that the use of an automated coding procedure meant that there was no need to exclude any utterances on the grounds that they could not be coded. All utterances that contained at least one relevant verb form were included in the analysis.

Simple-finite utterances were defined as utterances that only included unambiguously finite verb forms (e.g., utterances containing first-person singular, second-person singular, or third-person singular verb forms in Dutch and utterances containing third-person singular verb forms and irregular past tense verb forms in English[5]).

Compound-finite utterances were defined as utterances containing both an unambiguously finite verb form and a verb form that was not unambiguously finite (e.g., utterances containing a singular present tense verb form and an infinitive in Dutch, and utterances containing a modal and an infinitive or an auxiliary and a perfect or progressive participle in English).

Nonfinite utterances were defined as utterances that did not include an unambiguously finite verb form (e.g., utterances containing infinitive or plural present tense verb forms in Dutch and utterances containing zero-marked verb forms in English).

The previously mentioned classification was then used to determine the proportion of finite, compound-finite, and nonfinite utterances for each child and its respective model at each stage of development.

An important feature of this coding scheme is that it treats all ambiguous verb forms as if they were nonfinite verb forms. This feature of the coding scheme is necessary because there

Table 3
Rates of errors of commission in the model output of Anne, Becky, Matthijs, and Peter

	English		Dutch	
	Anne	Becky	Matthijs	Peter
Error Rate	.16	.14	.14	.19

are some finite verb forms in both Dutch and English that are indistinguishable from the infinitive verb forms. Thus, although there is strong evidence that Dutch and English children do produce infinitive verb forms in contexts in which a finite verb form is required, it is actually impossible to be sure whether the verb form included in any particular utterance is an infinitive verb form as opposed to a finite plural present tense verb form (in Dutch) or a zero-marked finite present tense verb form (in English).

An obvious disadvantage of coding the data in this way is that the measures are always likely to underestimate to some degree the child's and the model's ability to produce correct finite forms. This will be particularly true at later stages of development, when ambiguous forms are much more likely to be finite than nonfinite because OI errors are, by then, virtually absent from the children's speech. It will also be particularly true in English where ambiguous verb forms are much more likely to be finite than they are in Dutch because the relevant finite verb forms are high-frequency singular verb forms in English and low-frequency plural verb forms in Dutch. These facts should obviously be borne in mind when interpreting the absolute values of the measures reported in the simulations. However, it is important to realize that they do not affect the validity of any analysis of the closeness of the fit between the data of the children and their respective models because these data were analyzed in exactly the same way. Indeed, we would argue that they illustrate one of the strengths of our modeling approach, which is that it allows us to measure the closeness of fit between the child's and the model's output in a way that is independent of any assumptions that we might have about the knowledge underlying the child's use of particular utterances at particular points in development.

Nevertheless, it could be argued that the level of ambiguity in English is so great that the decision to deal with this ambiguity by treating all zero-marked forms in the same way makes it much easier to achieve a close fit to data on early child English than it does to data on early child Dutch and, hence, makes the model's ability to simulate English data much more difficult to evaluate than its ability to simulate Dutch data. To deal with this problem, a set of subsidiary analyses were performed on the English data that focused exclusively on contexts with third-person singular subjects (i.e., contexts such as "he like(s) sweets" in which the use of a zero-marked form is clearly incorrect). These analyses were conducted by searching in both the children's and the models' data for utterances in which the relevant verb form was preceded by one of the third-person singular subjects: *he, she, it, this* (*one*), or *that* (*one*). Only pronominal subjects were used in this analysis for the simple reason that it is much easier to identify all instances of utterances with pronominal subjects by lexical search than it is to identify all instances of utterances with lexical subjects. Each of the utterances identified in this way was classified as a simple-finite, a compound-finite, a nonfinite, or an ambiguous utterance. The categories of simple- and compound-finite utterance were defined in exactly the same way as in the coding scheme presented previously. However, restricting the analysis to utterances with third-person singular subjects meant that it was possible to define the category of nonfinite utterance more conservatively than in the earlier coding scheme, by distinguishing between utterances in which the potentially nonfinite verb form was clearly incorrect (e.g., "he like sweets") and utterances in which the potentially nonfinite verb form was not clearly incorrect (e.g., "he dropped the sweets").[6] Thus, for the purposes of this analysis a more fine-grained distinction was made between nonfinite utterances and ambiguous utterances.

Nonfinite utterances were defined as utterances that did not contain an unambiguously finite verb form and in which the verb form used was clearly incorrect (e.g., utterances including a zero-marked verb form such as *go,* or a bare present or past participle such as *going* or *gone*).

Ambiguous utterances were defined as utterances that did not contain an unambiguously finite verb form and in which the verb form used was not clearly incorrect (e.g., utterances including a verb form such as *dropped* that could be either a correct past tense verb form or an incorrect past participle).

This classification was then used to determine the proportion of simple-finite, compound-finite, nonfinite, and ambiguous utterances for each child and its respective model at each stage of development. It should be noted that this particular analysis was only conducted at the three later developmental points because it requires utterances that contain at least two words (i.e., a subject and a verb form), so it can only be performed on data in which the MLU is reasonably high.

3. Simulations: Results

3.1. Results for the Dutch simulations

Fig. 5 shows the data for the 2 Dutch-speaking children: Matthijs and Peter and their respective models. It is clear from Fig. 5 that although MOSAIC does not capture all of the fine detail of the children's data, it does simulate the central feature of the OI stage (i.e., the substantial drop in the proportion of nonfinite verb forms in finite clauses across the relevant MLU range). Thus, Matthijs and Peter's models show a decrease in the proportion of nonfinite utterances from an average of .73 at the earliest stage to an average of .30 at the final stage, compared with a decrease in the children's speech from an average of .84 at the earliest stage to an average of .19 at the final stage.

A more detailed look at the data in Fig. 5 reveals that, for both children, the model underestimates the proportion of nonfinite utterances at the earliest stage and overestimates the proportion of nonfinite utterances at the final stage. The model also overestimates the ratio of compound-finite to simple-finite utterances across the MLU range. Nevertheless, the model seems to provide a relatively good fit to the developmental data. Thus, the correlations between the children's and the models' data are .92 for Matthijs and .89 for Peter, and the root mean square errors for the four stages depicted in Fig. 5 are .11, .11, .16, and .10 for Matthijs's model, and .06, .09, .21, and .12 for Peter's model.

Descriptive statistics for the model's output are presented in Table 4. It can be seen from Table 4 that MLU increases with additional runs of the model, as does the number of utterances containing a verb. However, although both models show a sharp increase in the proportion of generated as opposed to rote-learned utterances across the first two data points, the proportion of generated utterances levels off or decreases in the later stages. The initial rise is the result of an increase in the size of the net. As relatively high-frequency items are added to the net, the number of generative links increases and, as a result, so does the proportion of generated utterances. The leveling off and decrease in the later stages results from relatively low-frequency items being added to the net. As a result of their low frequency, these items tend to share little

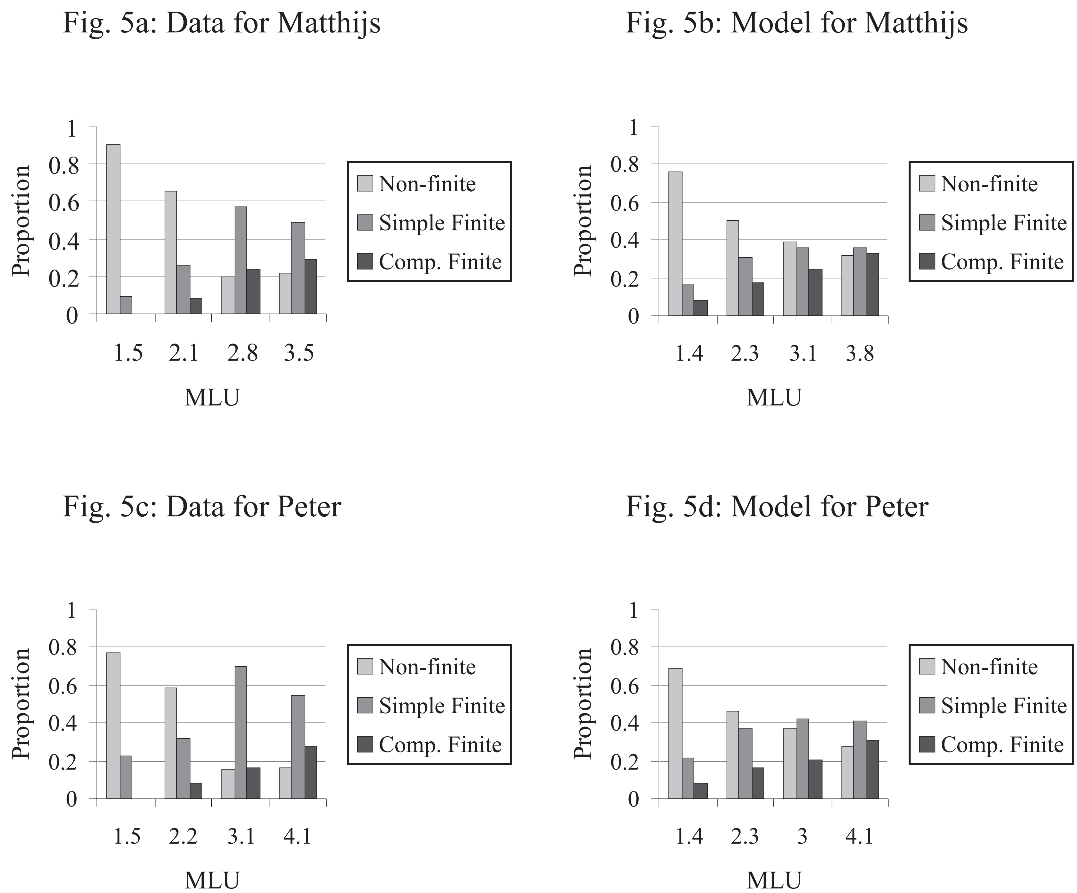

Fig. 5. Proportion of nonfinite, simple-finite, and compound-finite utterances for Peter, Matthijs, and their respective models, as a function of MLU.

overlap with other nodes. The number of generative links therefore ceases to rise so rapidly, whereas the number of nodes in the net continues to increase. This results in a leveling off or decrease in the proportion of generated utterances.

3.1.1. Percentage overlap

The data reported in Fig. 5 were obtained with the percentage overlap required for the creation of a generative link set to 10%. The effect of manipulating this variable was also assessed. In general, the result of these manipulations was that increasing the required percentage overlap resulted in a lower proportion of finite utterances and a lower MLU. Decreasing the percentage resulted in a higher proportion of finite utterances and a larger MLU. Setting the percentage overlap to a different value had a greater effect for the later developmental stages. This is not surprising, as the proportion of generated utterances increases with developmental phase. For Matthijs's model, reducing the percentage overlap from 13% to 7% resulted in a decrease in the proportion of nonfinite utterances of roughly .20 for the final stage. For Peter's

Table 4
Descriptive statistics for Matthijs's and Peter's models

Run	MLU	Number of Utterances Containing a Verb	Proportion of Generated Utterances	Root Mean Square Error
Matthijs's model				
6	1.44	251	.15	.11
11	2.32	5,492	.38	.11
13	3.06	11,353	.41	.16
14	3.80	20,467	.40	.10
Peter's model				
8	1.39	158	.10	.06
14	2.33	5,459	.46	.09
18	3.01	9,162	.40	.21
20	4.11	19,690	.35	.12

Note. MLU = mean length of utterance.

model, the decrease amounted to .08. Clearly, the percentage overlap parameter (and hence generativity) influences the production of finite utterances. However, there is also some variability in the behavior of Matthijs's and Peter's models. Causes for this variability may include the size of the input set and the relative distribution of low- and high-frequency items. For the purposes of this article, the value of the overlap parameter was set to 10%. All data reported hereafter (including the English data), were obtained using this value.

3.1.2. Word order

Of course, the fact that children use nonfinite verb forms in finite contexts is not the only feature of children's early multiword speech that the OI hypothesis is designed to explain. Thus, as Wexler (1994) pointed out, one of the most striking features of children's behavior during the OI stage is that, although children may produce a large number of OI errors, their use of finite and nonfinite verb forms tends to respect the word-order patterns of the language being learned. For example, although children learning Dutch use both finite and nonfinite verb forms in finite contexts, they nevertheless tend to correctly place finite verb forms before their complements and nonfinite verb forms after their complements.

To assess whether the model is able to simulate this aspect of the data, samples of utterances of the various types (i.e., nonfinite, simple-finite, and compound-finite utterances) were coded with respect to object–verb order. Utterances were selected on the basis that they included a constituent that could be considered a direct, indirect, or prepositional object. For compound finites, which consist of a finite auxiliary or modal and a nonfinite lexical verb, object position was coded relative to the nonfinite lexical verb. Samples were coded at each developmental stage. However, because the percentage of object–verb or verb–object orders did not differ drastically as a function of developmental stage; the proportion of word-order errors was calculated by collapsing across all developmental phases. The results of this analysis are pre-

Table 5
Proportion of correct object–verb orderings for the model as a function of finiteness (averaged over developmental phase)

	Finites	Nonfinites
Matthijs	.94	.91
Peter	.96	.93

sented in Table 5 from which it can be seen that, for the models of both children, the proportion of word-order errors was very low indeed.

These results show that the model is not only able to simulate developmental changes in children's use of finite and nonfinite verb forms, but is also able to simulate the fact that children's use of finite and nonfinite verb forms tends to respect the word-order patterns of the language being learned. Of course, in one sense, this finding is not particularly surprising, given the model's sensitivity to the distributional characteristics of the language to which it is exposed. However, the fact that such a simple learning mechanism can simulate this effect does illustrate the dangers of assuming that low rates of grammatical error in children's speech imply adultlike grammatical knowledge on the part of the child. More specifically, it undermines the claim that the low frequency of word-order errors in children's use of finite and nonfinite verb forms constitutes evidence that young children have already correctly set the word-order parameters of their language and suggests that any learning mechanism sensitive to the relation between finiteness marking and utterance position in the input would be likely to simulate this aspect of the data.

3.1.3. The eventive–stative asymmetry

An additional feature of the OI stage in Dutch is that there is a well-documented asymmetry in the kind of lexical verbs that occur as correct finite verb forms and incorrect OIs in children's speech. Thus, whereas the finite lexical verb forms that Dutch children produce during the OI stage tend to be a mixture of stative verbs such as *wil* (wants) and eventive verbs such as *eet* (eats), the nonfinite verbs that children produce as OI errors are almost exclusively eventive verbs (Jordens, 1990; Wijnen, 1998).

To assess whether MOSAIC is able to simulate this aspect of the data, samples of 250 nonfinite and 250 simple-finite utterances were drawn from the models' output files at the point at which the proportion of OI errors in the models' output was closest to .50 (MLU = 2.3 for Matthijs's model and 2.3 for Peter's model). Those utterances that included a lexical verb were then coded for whether the verb was eventive or stative. The results of this analysis are presented in Table 6, from which it can be seen that for both models the proportion of nonfinites classified as statives is very low and much lower than the proportion of simple finites classified as statives (.06 vs. .29, $\chi^2 = 38.8$, $p < .01$ for Matthijs's model and .03 vs. .25, $\chi^2 = 43.2$, $p < .01$ for Peter's model). MOSAIC is thus able to simulate the eventive–stative asymmetry.

In fact, given the way in which MOSAIC produces OI errors, this pattern of results is not particularly surprising, because, as Wijnen at al. (2001) pointed out, stative verbs are much less likely than eventive verbs to occur as the nonfinite verb form in Dutch compound finites. How-

Table 6
Proportion of simple-finite and nonfinite utterances with lexical verbs in the Dutch models' output that were classified as stative (as opposed to eventive)

	Finites	Nonfinites
Matthijs	.29	.06
Peter	.25	.03

ever, the fact that the model is able to simulate the eventive–stative asymmetry so readily does provide further support for the view that OI errors in Dutch are learned from compound finites in the input and, hence, for the idea that the pattern of finiteness marking in Dutch children's speech can be understood in terms of the interaction between an utterance-final bias in learning and the distributional characteristics of Dutch child-directed speech.

3.1.4. Mechanisms of developmental change

The results reported so far suggest that it is possible to simulate the central characteristics of the OI stage in Dutch in terms of a resource-limited distributional analysis of Dutch child-directed speech. However, they do not tell us which features of MOSAIC are most important for simulating these characteristics. An obvious candidate is the model's utterance-final bias, which results in the model producing progressively longer utterance-final strings as a function of the amount of input to which it has been exposed. Because in Dutch nonfinite verb forms occur after finite verb forms and are restricted to sentence-final position, short utterances learned from utterance-final position are more likely to include a single nonfinite verb form than longer utterances and hence to be classified as OI errors. However, as MLU increases, more of the utterances that the model produces will contain a finite verb form, including simple-finite utterances (i.e., utterances that only include finite verb forms) and compound-finite utterances (i.e., utterances that include both a finite and a nonfinite verb form). Thus, as MLU increases, the proportion of simple-finite and compound-finite utterances will increase and the proportion of nonfinite utterances will decrease.

One obvious way of investigating the importance of MOSAIC's utterance-final bias in learning is to analyze the relation between finiteness marking and utterance position in the input sets on which the models were trained. This was done by applying the same coding procedure used to analyze the model's output both to the total input set for each child and to subsets of this input set consisting of utterance-final strings of length 1, 2, 3, 4, and 5. The results of this analysis are presented in Table 7 and show that, although the proportion of nonfinite utterances in the total input set for each child is less than .20, the proportion of utterance-final strings that would be coded as nonfinite utterances increases dramatically as the length of these strings decreases (from .32 on average in five-word utterance-final strings to .89 on average in one-word utterance-final strings).

These results suggest that the model's utterance-final bias in learning plays an important role in determining the proportion of nonfinite utterances produced by the model at different points in development. However, a second factor that might affect the proportion of finite and nonfinite utterances in the model's output is change in the proportion of generated as opposed

Table 7
The proportion of utterances and utterance-final strings including verbs in the Dutch input that would be classified as nonfinite, simple-finite, and compound-finite utterances

	One-Word Strings	Two-Word Strings	Three-Word Strings	Four-Word Strings	Five-Word Strings	Complete Utterances
Matthijs's input						
Nonfinite	.90	.77	.61	.46	.34	.17
Simple-finite	.10	.18	.29	.36	.40	.44
Compound-finite	.00	.04	.10	.18	.26	.39
Peter's input						
Nonfinite	.87	.72	.53	.40	.30	.19
Simple-finite	.13	.24	.36	.42	.45	.48
Compound-finite	.00	.04	.10	.18	.24	.34

to rote-learned utterances that the model produces as its MLU increases. Generated utterances are more likely to contain finite than nonfinite verb forms because finite verb forms are more likely to become linked to one another than nonfinite verb forms as they are more frequent in the input and tend to occur in a wider range of contexts. Thus, as the model becomes more generative, the proportion of nonfinite utterances in its output is likely to decrease.

To assess the relative importance of these two factors, a regression analysis was carried out on the proportion of nonfinite utterances in the model's output using MLU and the proportion of generated utterances as predictors.[7] This analysis revealed that both MLU and the proportion of generated utterances contributed significantly to the regression equation. MLU alone explained 88% of the variance. The proportion of generated utterances explained an additional 10% of the variance. These results show that the most important determinant of the decrease in the proportion of nonfinite utterances produced by the model is the model's ability to produce progressively longer utterance-final strings as a function of the amount of input to which it has been exposed. However, they also suggest that changes in the generativity of the model play a role over and above that of increasing utterance length. This suggests that, if it were possible to increase the generativity of the model during the later stages without increasing the MLU of its output, it might be possible to simulate the developmental data even more closely.

3.2. *Results for the English simulations*

Fig. 6 shows the data for the 2 English-speaking children Anne and Becky and their respective models. It is clear from Fig. 6 that, although the drop in the proportion of nonfinite utterances for both the children and their models is much less pronounced than it is in Dutch, the same version of the model that provides a good fit to the data from the 2 Dutch children also simulates the developmental decrease in the proportion of nonfinite utterances shown by the 2 English children.

A more detailed look at the data in Fig. 6 reveals that, as in Dutch, the model underestimates the proportion of nonfinite utterances at the earliest stage (.80 on average vs. .97 for the children) and overestimates the proportion of nonfinite utterances at the final stage (.69 on average

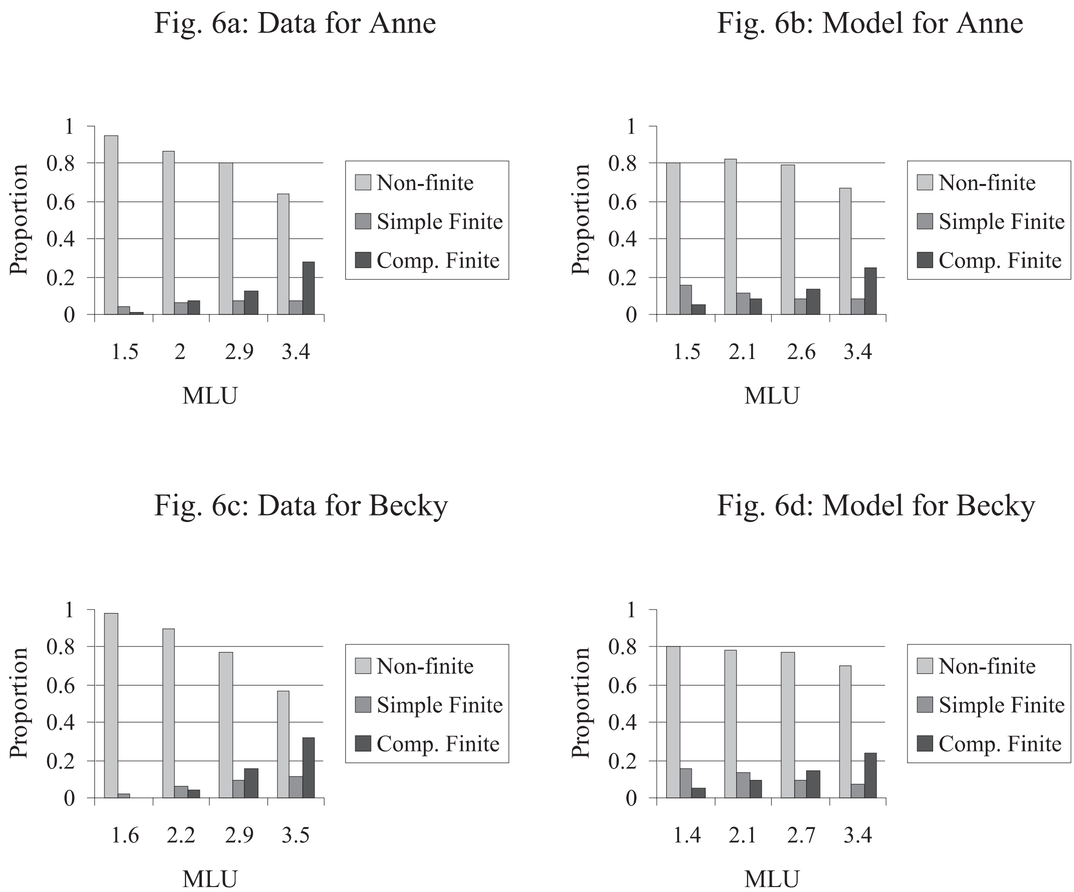

Fig. 6. Proportion of nonfinite, simple-finite, and compound-finite utterances for Anne, Becky, and their respective models as a function of MLU.

vs. .61). Nevertheless, the model seems to provide a relatively good fit to the data. Thus, the correlations between the children's and the models' data are .99 for Anne and .97 for Becky, and the root mean square errors for the four stages depicted in Fig. 6 are .11, .04, .01, and .02 for Anne's model and .13, .08, .01, and .09 for Becky's model.

The role of MOSAIC's utterance-final bias in simulating the English data was investigated in the same way as in the Dutch simulations (i.e., by applying the same coding procedure used to analyze the model's output both to the total input set for each child and to subsets of this input set consisting of utterance-final strings of Lengths 1, 2, 3, 4, and 5). The results of this analysis are presented in Table 8 and show that, although the proportion of utterances that would be classified as nonfinite in the total input set for each child is less than .50, the proportion of utterance-final strings that would be coded as nonfinite utterances changes dramatically as the length of these strings increases (from .90 on average in one-word utterance-final strings to .55 on average in five-word utterance-final strings).

These results suggest that, as in Dutch, the most important factor in determining the proportion of nonfinite utterances produced by the English models at different points in development

Table 8
The proportion of utterances and utterance-final strings including verbs in the English input that would be classified as nonfinite, simple-finite, and compound-finite utterances

	One-Word Strings	Two-Word Strings	Three-Word Strings	Four-Word Strings	Five-Word Strings	Complete Utterances
Anne's input						
Nonfinite	.90	.82	.70	.62	.54	.35
Simple-finite	.10	.08	.08	.08	.08	.06
Compound-finite	.00	.08	.22	.30	.39	.59
Becky's input						
Nonfinite	.90	.83	.72	.63	.56	.43
Simple-finite	.10	.10	.08	.07	.06	.05
Compound-finite	.00	.08	.20	.33	.38	.51

is MOSAIC's utterance-final bias in learning. However, as in Dutch, the proportion of finite and nonfinite utterances in the model's output is also likely to be affected by changes in the proportion of generated as opposed to rote-learned utterances that the model produces as its MLU increases.

To assess the relative importance of these two factors, a regression analysis was carried out on the proportion of nonfinite utterances in the model's output, using MLU and the proportion of generated utterances as predictors. This analysis revealed that, as in Dutch, both MLU and the proportion of generated utterances contributed significantly to the regression equation. MLU alone explained 75% of the variance. The proportion of generated utterances explained an additional 15% of the variance.

These results are clearly very similar to those reported earlier for Dutch and suggest that the mechanisms that allow MOSAIC to simulate the English data are broadly the same as those that allow MOSAIC to simulate the Dutch data. However, as has already been pointed out, it could be argued that the level of ambiguity in English is so great that the decision to deal with this ambiguity by treating all zero-marked forms in the same way makes it much easier to fit the data on early child English than it is to fit the data on early child Dutch.

To deal with this problem, a set of subsidiary analyses were performed on the English data that focused exclusively on contexts with the third-person singular subjects: *he, she, it, this* (*one*), or *that* (*one*). The results of this analysis are presented in Fig. 7. It should be noted that this analysis is based on three MLU levels rather than four. This is because it requires utterances that contain at least two words (i.e., a subject and a verb form) and so could only be performed on data from the last three MLU stages (i.e., data in which the MLU was reasonably high). It should also be noted that the MLU figures for the children (but not their models) differ slightly from those in Fig. 6. This is because the more stringent inclusion criteria for this analysis meant that it was necessary to code a larger amount of child (but not model) data to extract a sufficient number of utterances to derive reliable measures of the relevant proportions.

It is clear from Fig. 7 that even in the case of this more conservative analysis there is an excellent fit between the children's speech and the output of the model. Thus, the correlations between the children's and the models' data (disregarding the proportion of ambiguous utter-

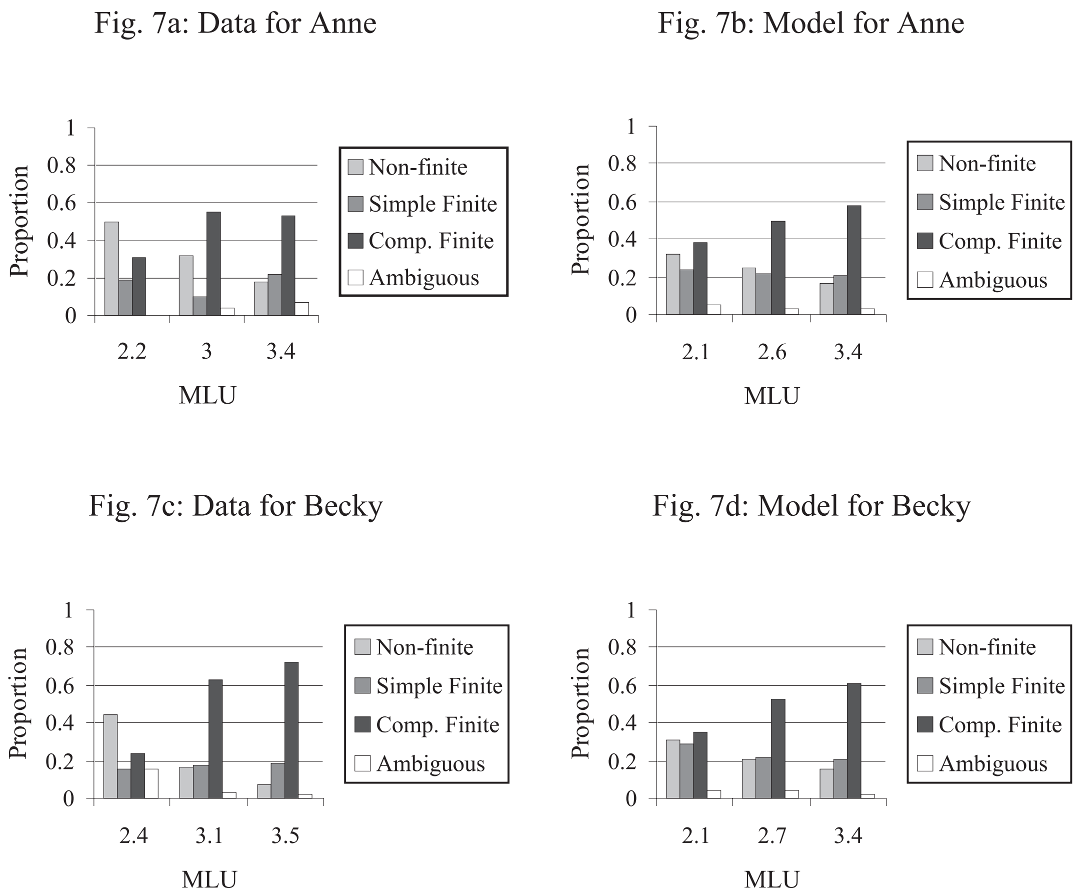

Fig 7. Proportion of nonfinite, simple-finite, compound-finite, and ambiguous utterances with *he, she, it, that* (*one*), or *this* (*one*) as subject for Anne, Becky, and their respective models as a function of MLU.

ances) are .84 for Anne and .89 for Becky, and the root mean square errors for the three stages depicted in Fig. 7 are .10, .07, and .03 for Anne's model, and .12, .06, and .07 for Becky's model.

These results show that the model's ability to simulate the data from English-speaking children is not simply a result of the much greater level of ambiguity that exists in English as opposed to Dutch data. They also reveal a striking difference in the proportion of compound finites in the English and Dutch data at comparable MLUs, with the English children and their models tending to produce a much higher proportion of compound finites than the Dutch children and their models. Although a direct comparison of these proportions is somewhat difficult because the English data are based on only a subset of those utterances that include a finite or nonfinite verb form, there is an obvious explanation for this difference. This is that, because nonfinites are restricted to sentence-final position in Dutch, far more material (e.g., object arguments, prepositional phrases and adverbs) can—and does—occur between the finite and the nonfinite verb form in Dutch compound finites than between the finite and the nonfinite verb form in English compound finites. This difference makes it easier for both the child and the model to learn structures including compound finites in English than in Dutch and, hence, results in a much lower proportion of compound finites in Dutch than in English, particularly at intermediate points in development.

4. Discussion

The central aim of this study was to investigate whether the OI phenomenon in Dutch and English could be explained in terms of a resource-limited distributional analysis of the speech addressed to Dutch- and English-speaking children. This was done by using MOSAIC to simulate data on the development of finiteness marking in Dutch and English. The results show, first, that MOSAIC is able to simulate the substantial drop in the proportion of nonfinite verb forms in finite clauses in children learning Dutch; second, that MOSAIC is able to simulate the fact that Dutch children's use of finite and nonfinite verb forms tends to respect the word-order pattern of the language; and, third, that the same version of MOSAIC is also able to simulate the data from children learning English. They also show that MOSAIC's ability to simulate the data from both Dutch and English is not simply a consequence of the high level of ambiguity in the data of English-speaking children. Thus, MOSAIC not only simulates the overall pattern of performance in children learning English, but also provides a good fit to the results of a much more targeted analysis of English focused specifically on contexts in which children's use of zero-marked verb forms is clearly incorrect.

These results suggest that it is possible to explain the key features of the OI phenomenon in Dutch and English in terms of the interaction between an utterance-final bias in learning and the distributional characteristics of child-directed speech in the two languages. They also illustrate the potential power of using computational modeling techniques to investigate the extent to which cross-linguistic similarities in the developmental data can be explained in terms of common processing constraints as opposed to innate knowledge of universal grammar. An obvious question for future research is whether the same constraints that allow MOSAIC to simulate the OI phenomenon in English and Dutch also allow MOSAIC to simulate the developmental patterning of finiteness marking in other languages, particularly languages such as Italian and Spanish in which children do not appear to go through an OI stage.

If we focus on the question of how MOSAIC simulates the data on Dutch and English, the results of this study suggest that the most important factor is MOSAIC's ability to reproduce and generate from progressively longer utterance-final sequences as a function of learning (although MOSAIC's tendency to produce a higher proportion of generated as opposed to rote-learned utterances later in development also appears to play a role). Thus, 88% of the variance in the proportion of nonfinite utterances in Dutch and 75% of the variance in the proportion of nonfinite utterances in English was explained by MLU alone (with an additional 10% and 15%, respectively, being explained by the proportion of generated utterances). These results are interesting for two reasons.

First, they show how building one relatively simple processing constraint into the language-learning mechanism can have a very profound effect on the developmental dynamics of the learning process. Thus, MOSAIC learns OI errors in Dutch from utterances that include compound finites. These utterances constitute only around 30% of the utterances including verbs in parents' child-directed speech. However, restricting MOSAIC to reproducing and generating from utterance-final sequences results in rates of nonfinite utterances in Dutch of approximately 70% during the early stages (at MLUs of approximately 1.5) that drop to around 30% during the later stages (at MLUs of approximately 4.0). We suspect that few would have predicted that such a simple constraint would have had such a large effect on the model's output before the simulations were run.

Second, they illustrate how a simple constraint of this kind can interact with the distributional characteristics of the input language to result in interesting patterns of cross-linguistic similarity and difference in the shape of the language learner's output. Thus, although Dutch and English are, in principle, quite different languages (Dutch is a subject–object–verb/verb second language in which verb position is dependent on finiteness, whereas English is a subject–verb–object language in which verb position is not dependent on finiteness), MOSAIC's utterance-final bias results in an OI stage in both Dutch and English because, in both Dutch and English, finite verb forms occur earlier in compound finites than nonfinite verb forms. On the other hand, MOSAIC's utterance-final bias also appears to interact with differences in the distributional characteristics of Dutch and English to result in a lower proportion of compound finites in Dutch than in English at intermediate points in development. Thus, the fact that nonfinite verb forms are restricted to sentence-final position in Dutch means that more lexical material tends to occur between the finite and nonfinite verb forms in Dutch compound finites than in English compound finites. This has the effect of making Dutch compound finites somewhat more difficult to learn than English compound finites.

Of course, although MOSAIC does a remarkably good job of simulating the OI phenomenon in Dutch and English, it is important to acknowledge that the fit between the output of the model and the output of the children on whose input the model has been trained is still far from perfect. Thus, in both Dutch and English, MOSAIC tends to underestimate the proportion of nonfinite utterances during the early stages and to overestimate the proportion of nonfinite utterances during the later stages.

One possible reason why MOSAIC tends to underestimate the proportion of nonfinite utterances at low MLUs is that differences in sentence position may not be the only factor responsible for children's increased sensitivity to nonfinite versus finite verb forms during the early stages. Thus, Wijnen et al. (2001). suggested that the high proportion of nonfinite forms in early child Dutch "may be related to an increased overall salience, due to their systematic appearance in sentence-final position *and* [italics added] their high conceptual transparency as compared to finite verbs" (Wijnen et al., p. 629). Some support for the idea that finite verb forms may be less conceptually transparent on average than nonfinite verbs comes from the fact that a high proportion of the simple-finite utterances produced by the model during the early stages consisted of utterances including isolated modals such as *mag* (*may*) or *can* that seemed to have been learned from elliptical utterances such as *Dat mag* (*That may/that is allowed*) or *He can.* Such utterances do not tend to be produced by children during the early stages of development, presumably because their function is to modulate rather than to express the central relational meaning of the utterance, so the meanings that they encode are inherently more complex and, therefore, less conceptually transparent than the meanings encoded by, for example, simple action verbs such as *jump* and *kick.* Because MOSAIC is a simple distributional analyzer with no means of representing semantic or conceptual information, it is obviously unable to simulate such conceptual transparency effects.

One possible reason why MOSAIC tends to overestimate the proportion of nonfinite utterances at higher MLUs is the model's limited generativity. Thus, one weakness of this version of the model is that it is constrained to traverse only one generative link per utterance. This constraint is obviously somewhat unrealistic. However, because generated utterances are more likely to include finite than nonfinite verb forms, and because the generative possibilities of language increase exponentially with increasing MLU, relaxing this constraint is likely to improve the fit between model and child by leading to a differential increase in the proportion of finite utterances at higher MLU levels. Future work will investigate this possibility by systematically relaxing this constraint on the generativity of the model and assessing the effect of these changes on the fit between model and child.

Another possible reason why MOSAIC tends to overestimate the proportion of nonfinites at higher MLUs is the model's limited ability to "unlearn." Thus, a second weakness of this version of the model is that, although it learns to produce progressively longer utterances as a function of development, it does not currently have any mechanism for "forgetting" or "learning not to" produce short incomplete utterances as its MLU increases. That is to say, because shorter and longer sequences learned from the same utterance in the input are represented separately in the network, the model continues to produce short incomplete utterances even after it has learned to produce the complete utterances from which these short utterances were learned. Because short incomplete utterances are more likely to be nonfinite than finite, developing MOSAIC's ability to unlearn is likely to lead to a better fit between the child and the model's output. One way in which this could be done would be by adding a mechanism to the model that links short incomplete utterances to the longer utterances of which they are a part and uses this information to prevent the model from producing or generating from utterances made obsolete by subsequent learning.

The extent to which either of the previously mentioned solutions is successful in improving the fit between model and child is obviously a question for future research. However, we would argue that the fact that MOSAIC allows us to ask questions about factors such as the interaction between generativity and increasing MLU illustrates its potential utility as a means of investigating the developmental patterning of children's early multiword speech.

Of course, despite this utility, MOSAIC is not, nor is it intended to be, a complete model of the language-learning process. On the contrary, MOSAIC is a relatively simple distributional analyzer, with no access to semantic information, which is clearly not powerful enough to acquire many aspects of adult syntax. On the other hand, the processing limitations built into MOSAIC, although inevitably somewhat idealized, do seem to approximate limitations that are likely to operate in children's language learning. That is to say, it seems reasonable to assume that there are limitations on the amount of distributional information that can be extracted by the child from any single utterance presentation, and that information from the end of the utterance may be more accessible than information from earlier in the utterance (although there may also be a processing advantage for information at the beginning of the utterance). The fact that these processing limitations allow MOSAIC to do such a good job of simulating the developmental data is therefore an important finding.

Finally, it is worth noting that MOSAIC is, to our knowledge, the only computational model of language development that simultaneously (a) learns from samples of naturalistic input that reflect the lexical statistics and distributional characteristics of the language being learned, (b) produces output that can be directly compared with that of real children, (c) simulates data from more than one language, and (d) takes explicit account of the potential role of factors such as increasing MLU and increasing generativity in determining the shape of the developmental data. As such, MOSAIC seems to us to provide a powerful means of investigating the relation between the distributional characteristics of children's early language and the distributional characteristics of the language to which they are exposed and, hence, of generating more process-oriented explanations of similarities and differences in the patterning of the cross-linguistic data. It also serves as a useful corrective to generativist accounts of these patterns (e.g., Hoekstra & Hyams, 1998; Wexler, 1994, 1998), which tend to assume a great deal of highly structured linguistic knowledge for which there is little positive evidence in children's early speech (e.g., Pine, Lieven, & Rowland, 1998; Tomasello, 2000; Wilson, 2003), and as a useful complement to more traditional computational approaches (e.g., Cartwright & Brent, 1997; Chang, 2002; Elman, 1993), which tend to focus on developing solutions to particular learnability problems rather than building process-oriented accounts of the developmental data.

Notes

1. Finite forms are forms that are marked for agreement or tense or both (e.g., *goes, went*). Nonfinite forms are forms that are not marked for agreement or tense. Nonfinite forms include the infinitive (e.g., *go*), the past participle (e.g., *gone*), and the progressive participle (e.g., *going*).
2. This parameter is set to a number that allows a range of MLUs to be modeled with consecutive runs of the model. For larger input sizes, the parameter needs to be set to a larger value. Because the English input files are larger than the Dutch input files, a larger value was chosen for English.
3. If the encoded phrase is a sentence-final string (i.e., contains an *end marker*), .5 is subtracted from the length of the node contents. This is done to increase the likelihood of creating nodes encoding utterance-final phrases and constitutes an end-marker bias in learning, in addition to the constraint used in production (see Section 1.2.6). Note that the end marker is not counted when assessing the length of the utterance.
4. Strictly speaking, generative links are created between *nodes* encoding phrases that have the property mentioned previously. When the context is clear, we will use the simpler construction.
5. English irregular past tense forms were usually treated as finite verb forms because, unlike regular past tense verb forms (e.g., *kicked, dropped*), they can usually be distinguished from the nonfinite past participle form of the verb (e.g., *went* vs. *gone, flew* vs. *flown*). However, in keeping with the spirit of the coding scheme, irregular past tense

forms that could not be distinguished from the past participle the infinitival form of the verb or both (e.g., *brought, hit*) were treated as nonfinite verb forms.

6. Note that although only treating potentially nonfinite forms in third-person singular contexts as optional infinitive errors is clearly much more conservative than treating all potentially nonfinite forms in finite contexts as optional infinitive errors, it does not actually solve the ambiguity problem because it is still not possible to tell whether the relevant error is a true optional infinitive error or an agreement error (i.e., an error that reflects the child's use of a zero-marked finite verb form in a context requiring the use of finite verb form marked for third-person singular agreement).
7. Because the average length of generated utterances tends to be slightly greater than that of rote utterances, the effects of MLU and generativity are, in practice, somewhat confounded. To assess the effect of generativity over and above that of MLU, the MLU of rote-learned utterances was used in this analysis.

Acknowledgments

This research was funded by the Leverhulme Trust under grant number F/114/BK and by the ESRC under Grant No. R000223954.

References

Bates, E., & Carnavale, G. F. (1993). New directions in research on child development. *Developmental Review, 13,* 436–470.

Bloom, L., Lifter, K., & Hafitz, J. (1980). Semantics of verbs and the development of verb inflection in child language. *Language, 56,* 386–412.

Bloom, P. (1990). Subjectless sentences in child language. *Linguistic Inquiry, 21,* 491–504.

Bol, G. W. (1996). Optional subjects in Dutch child language. In C. Koster & F. Wijnen (Eds.), *Proceedings of the Groningen Assembly on Language Acquisition* (pp. 125–135). Groningen, The Netherlands: Centre for Language and Cognition Groningen.

Brown, R. (1973). *A first language: The early stages.* 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.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science, 26,* 609–651.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht, The Netherlands: Foris.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences, 5,* 82–88.

Croker, S., Pine, J. M., & Gobet, F. (2000). Modelling optional infinitive phenomena: A computational account. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modeling* (pp. 78–85). Veenendaal, The Netherlands: Universal Press.

Croker, S., Pine, J. M., & Gobet, F. (2001). Modelling children's case-marking errors with MOSAIC. In E. M. Altmann, A. Cleeremans, C. D. Schunn, & W. D. Gray (Eds.), *Proceedings of the Fourth International Conference on Cognitive Modeling* (pp. 55–60). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

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

Freudenthal, D., Pine, J., & Gobet, F. (2001). Modeling the optional infinitive stage in MOSAIC: A generalisation to Dutch. In E. M. Altmann, A. Cleeremans, C. D. Schunn, & W. D. Gray (Eds.), *Proceedings of the Fourth International Conference on Cognitive Modeling* (pp. 79–84). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Freudenthal, D., Pine, J., & Gobet, F. (2002a). Modelling the development of Dutch optional infinitives in MOSAIC. In W. Gray & C. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Meeting of the Cognitive Science Society* (pp. 322–327). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Freudenthal, D., Pine, J., & Gobet, F. (2002b). Subject omission in children's language: The case for performance limitations in learning. In W. Gray & C. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Meeting of the Cognitive Science Society* (pp. 328–333). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5,* 236–243.

Gobet, F., & Oliver, I. (2002). *A simulation of memory for computer programs* (CREDIT Tech. Rep. 74). Nottingham, England: University of Nottingham, Centre for Research in Development.

Gobet, F., & Pine, J. (1997). Modelling the acquisition of syntactic categories. In M. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Society* (pp. 265–270). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

Harris, T., & Wexler, K. (1996). The optional-infinitive stage in child English: Evidence from negation. In H. Clahsen (Ed.), *Generative perspectives in language acquisition* (pp. 1–42). Philadelphia: Benjamins.

Hoekstra, T., & Hyams, N. (1998). Aspects of root infinitives. *Lingua, 106,* 81–112.

Ingram, D., & Thompson, W. (1996). Early syntactic acquisition in German: Evidence for the modal hypothesis. *Language, 72,* 97–120.

Jones, G., Gobet, F., & Pine, J. M. (2000a). Learning novel sound patterns. In N. Taatgen & J. Aasman (Eds.), *Proceedings of the Third International Conference on Cognitive Modelling* (pp.169–176). Veenendaal, The Netherlands: Universal Press.

Jones, G., Gobet, F., & Pine, J. M. (2000b). A process model of children's early verb use. In L. R. Gleitman & A. K. Joshu (Eds.), *Proceedings of the Twenty-Second Annual Meeting of the Cognitive Science Society* (pp. 723–728). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Jordens, P. (1990). The acquisition of verb placement in Dutch and German. *Linguistics, 28,* 1407–1448.

Joseph, K. L., & Pine, J. M. (2002). Does error-free use of French negation constitute evidence for Very Early Parameter Setting? *Journal of Child Language, 29,* 71–86.

Lane, P. C. R., Cheng, P. C.-H., & Gobet, F. (1999). Learning perceptual schemas to avoid the utility problem. In M. Bramer, A. Macintosh, & F. Coenen (Eds.), *Proceedings of the Nineteenth SGES International Conference on Knowledge Based Systems and Applied Artificial Intelligence* (pp. 72–82). Cambridge, England: Springer.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Naigles, L., & Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language, 25,* 95–120.

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.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA: Harvard University Press.

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

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Santelmann, L., & Jusczyk, P. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69,* 105–134.

Shady, M., & Gerken, L. (1999). Grammatical and caregiver cue in early sentence comprehension. *Journal of Child Language, 26,* 163–176.

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2003). Unsupervised efficient learning and representation of language structure. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the Twenty-Fifth Annual Meeting of the Cognitive Science Society* (pp. 1106–1111). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28,* 127–152.

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

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition, 40,* 21–81.

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

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua, 106,* 23–79.

Wijnen, F. (1998). The temporal interpretation of Dutch children's root infinitivals: The effect of eventivity. *First Language, 18,* 379–402.

Wijnen, F., Kempen, M., & Gillis, S. (2001). Root infinitives in Dutch early child language. *Journal of Child Language, 28,* 629–660.

Wijnen, F., & Verrips, M. (1998). The acquisition of Dutch syntax. In S. Gillis & A. De Houwer (Eds.), *The acquisition of Dutch* (pp. 223–299). Amsterdam: Benjamins.

Wilson, S. (2003). Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language, 30,* 75–115.