Simulating the cross-linguistic pattern of Optional Infinitive errors in children's declaratives and Wh- questions

Daniel Freudenthal, Julian M. Pine

Department of Psychological Sciences

University of Liverpool

Gary Jones

Division of Psychology

Nottingham Trent University

Fernand Gobet

Department of Psychological Sciences

University of Liverpool

ABSTRACT

One of the most striking features of children's early multi-word speech is their tendency to produce non-finite verb forms in contexts in which a finite verb form is required (Optional Infinitive [OI] errors, Wexler, 1994). MOSAIC is a computational model of language learning that simulates developmental changes in the rate of OI errors across several different languages by learning compound finite constructions from the right edge of the utterance (Freudenthal, Pine & Gobet, 2006a; 2009; Freudenthal, Pine, Aguado-Orea & Gobet, 2007). However, MOSAIC currently only simulates the pattern of OI errors in declaratives, and there are important differences in the cross-linguistic patterning of OI errors in declaratives and Wh- questions. In the present study, we describe a new version of MOSAIC that learns from both the right and left edges of the utterance. Our simulations demonstrate that this new version of the model is able to capture the cross-linguistic patterning of OI errors in declaratives in English, Dutch, German and Spanish by learning from declarative input, and the cross-linguistic patterning of OI errors in Wh- questions in English, German and Spanish by learning from interrogative input. These results show that MOSAIC is able to provide an integrated account of the cross-linguistic patterning of OI errors in declaratives and Wh-questions, and provide further support for the view, instantiated in MOSAIC, that OI errors are compound-finite utterances with missing modals or auxiliaries.

Keywords: Optional Infinitive Errors; Computational Modelling; Cross-linguistic Analysis.

1. Introduction

One of the challenges facing computational approaches to language development is to develop models whose behaviour can be directly compared with that of language-learning children. Another is to show that these models can be extended beyond one particular language to simulate the cross-linguistic data. MOSAIC (Model of Syntax Acquisition in Children) is a computational model of language acquisition that attempts to meet these challenges by using the same learning mechanism to simulate the behaviour of language-learning children in several different languages. Previous work with MOSAIC has shown that the model is able to simulate cross-linguistic variation with respect to the Optional Infinitive (OI) phenomenon in terms of the interaction between a mechanism that learns from the right edge of the utterance and differences in the distributional properties of OI and non-OI languages (Freudenthal, Pine & Gobet, 2006a; Freudenthal, Pine, Aguado-Orea & Gobet, 2007). However, although MOSAIC provides a good fit to the developmental pattern of OI errors in four different languages (English, Dutch, German and Spanish), it currently only does so for declarative utterances.

In the present study, we describe a new version of MOSAIC that learns from both the right and left edges of the utterance, and investigate whether this version of the model is able to simulate the cross-linguistic pattern of OI errors in both declaratives and Wh-questions. It should be noted that requiring MOSAIC to simulate data on Wh- questions as well as declaratives is a very strong test of the model, since there are important differences in the cross-linguistic patterning of OI errors in Wh- questions and OI errors in declaratives. For example, although OI errors occur at high rates in both declaratives and Wh- questions in English, this is not the case in Dutch and German, where OI errors

occur at high rates in declaratives, but at very low rates in Wh- questions. In the present paper, we investigate whether it is possible to simulate these differences using the same learning mechanism. The aim is to provide a unified account of the pattern of OI errors in declaratives and Wh- questions in terms of the interaction between psychologically motivated constraints on learning and differences in the distributional properties of the language to which children are exposed.

1.1. The Optional Infinitive Phenomenon

One of the most striking features of children's early multi-word speech is the tendency to produce non-finite verb forms in contexts in which an adult would produce a finite verb form. For example, English-speaking children produce utterances such as 1a instead of the correct 1b; Dutch children produce utterances such as 2a instead of the correct 2b; German children produce utterances such as 3a instead of the correct 3b; and Spanish children (occasionally) produce utterances such as 4a instead of the correct 4b:

English

- (1a) That go there

  That go-INF there

  (1b) That goes there

  That go-FIN there

Dutch

- (2a) Mama ijs eten

  Mama ice-cream eat-INF

  (2b) Mama eet ijs

  Mama eat-FIN ice cream

German

- (3a) Papa Kaffee trinken

  Papa coffee drink-INF

  (3b) Papa trinkt Kaffee

  Papa drink-FIN coffee

Spanish

- (4a) Jugar al fútbol

  (He) play-INF football

  (4b) Juega al fútbol

  (He) play-FIN football

These errors involve the use of a non-finite verb form (zero-marked in English, but marked with the infinitival morpheme *–en* in Dutch and German and *–ar* in Spanish) in contexts in which a finite verb form is obligatory. Since they tend to occur at a stage when the child is also producing correctly marked finite forms, they have come to be known in the literature as Optional Infinitive (OI) errors (Wexler, 1994).

A number of theories have been proposed to account for the occurrence of OI errors in children's speech. For example, Hyams (1996) argues that children can leave functional heads such as I (Inflection) and D (Determiner) underspecified in the underlying representation of the sentence, which results in a lack of finiteness in the verbal domain and a lack of specificity in the nominal domain; and Rizzi (1994) argues that, rather than projecting a full CP (Complementizer Phrase) structure, children have the option of truncating lower down the clause, with a structure truncated below TP (Tense Phrase) resulting in a nonfinite clause. However, the most influential account of OI errors is

probably that of Wexler (1994, 1998; Schütze & Wexler, 1996).

Wexler's theory is designed to explain the occurrence of OI errors in obligatory subject languages such as English Dutch, and German (Wexler, 1994), and the near absence of such errors in INFL-licensed null subject languages such as Spanish and Italian (Wexler, 1998). According to Wexler (1998), children have correctly set all the inflectional and clause structure parameters of their language from a very early age, but are subject to a 'Unique Checking Constraint' (UCC) that prevents them from checking the D-feature of the subject DP against more than one D-feature. Obligatory subject languages such as English, Dutch and German require the checking of two D-features (Tense and Agreement). The UCC therefore results in the under-specification of Tense or Agreement in the underlying representation of the sentence and the occurrence of OI errors. INFL-licensed null subject languages such as Spanish require the checking of only one D-feature (Tense) on finite verbs. As a result, OI errors are rare in these languages.

Wexler's (1998) account provides a principled way of distinguishing between OI languages like English, Dutch and German and non-OI languages like Spanish. Moreover, it also has the advantage that it can explain the cross-linguistic patterning of OI errors in Wh- questions (Wexler, 1998). Interestingly, the pattern of errors in Wh-questions is different from the pattern of errors in declaratives. Thus, although children learning the OI language English make OI errors in both declaratives (*That go there*) and Wh- questions (*Where that go?*), and children learning non-OI languages such as Spanish rarely make OI errors in either declaratives or Wh- questions, children learning many OI languages (including Dutch and German) make OI errors in declaratives (e.g. Dutch: *Dat daar gaan,* German: *Das da gehen* (That there go-INF)), but do not make such errors in

Wh- questions (e.g. Dutch: *Waar dat gaan?* German: *Wo das gehen?* (Where that go-INF?)).

Table 1: Presence and absence of OI errors in declaratives in Wh- questions in English, Dutch, German and Spanish

Language	OI errors occur in Declaratives	OI errors occur in Wh-questions
English	YES	YES
Dutch	YES	NO
German	YES	NO
Spanish	NO	NO

This pattern of results (summarised in Table 1) is inconsistent with the predictions of other generativist accounts. For example, according to Rizzi's truncation account, OI errors reflect the truncation of structures below TP. Since truncation below TP necessarily implies the absence of CP (because CP is higher than TP in the sentence structure), and since Wh- questions require the presence of CP, OI errors are predicted to occur in declaratives, but not in Wh- questions. Rizzi's account can therefore explain the absence of OI errors in Dutch and German Wh- questions. However, it cannot explain the occurrence of OI errors in English Wh- questions. According to Wexler's Unique Checking Constraint account, on the other hand, OI errors reflect the under-specification of Tense and/or Agreement and are therefore predicted to occur in both declaratives and Wh- questions. However, OI errors in Wh- questions are specifically blocked in V2 languages like Dutch and German as a consequence of Very Early Parameter Setting.

Thus, it is a feature of V2 languages that, when the first position of a matrix clause is filled, the second position must be filled by a finite verb form (see examples 5 and 6).

Dutch

(5a) Ik lees een boek

I read-FIN a book

(5b) Gisteren las ik een boek

Yesterday read-FIN I a book

German

(6a) Ich lese ein Buch

I-read-FIN a book

(6b) Gestern las ich ein Buch

Yesterday read-FIN I a book

Since Wexler assumes that Dutch and German children know this fact from early in development, and since the first position of a Wh- question is always filled by a Wh-word, Dutch and German children are constrained to produce a finite main verb or a finite auxiliary in second position, and OI errors do not occur. Wexler's account can therefore explain both the occurrence of OI errors in English Wh- questions and the absence of OI errors in Dutch and German Wh- questions, and hence provides an integrated account of the cross-linguistic patterning of OI errors in both declaratives and Wh- questions.

1.2. An alternative account of Optional Infinitive errors.

Wexler's (1998) theory does a good job of explaining the cross-linguistic patterning of OI errors. However, in their recent work, Freudenthal et al. have shown that this

patterning can also be understood in terms of the interaction between a relatively simple learning mechanism (MOSAIC) and cross-linguistic differences in the distributional statistics of the input that children receive (Freudenthal et al., 2006a, 2007, 2009). MOSAIC is a computational model of language learning, with no built-in knowledge of syntactic categories or rules, which takes as input corpora of orthographically transcribed child-directed speech and learns to produce as output 'child-like' utterances that become progressively longer as learning proceeds. As a result of these characteristics, MOSAIC can be used to generate corpora of utterances at different stages of development, and hence to model the behaviour of children in different languages across a range of Mean Length of Utterance (MLU) values.

MOSAIC simulates OI errors because it has a strong utterance-final bias in learning. This bias results in the production of partial utterances that were present as utterance-final phrases in the input to which the model was exposed. The utterances in the input that give rise to OI errors are *compound finites*: utterances that contain both a finite verb (usually a modal or auxiliary) and a non-finite verb. Thus, MOSAIC learns to produce utterances resembling English OI errors such as *Go there* and *That go there* as truncated versions of utterances such as *(That can) go there* and *(Does) that go there?* Similarly, MOSAIC learns to produce utterances resembling Dutch OI errors such as *IJs eten* and *Mama ijs eten* as truncated versions of utterances such as *(Mama wil) ijs eten (Mummy wants to eat ice cream)* and *Wil Mama ijs eten? (Does Mummy want to eat ice cream?).*

MOSAIC simulates the developmental patterning of OI errors because it learns to produce progressively longer utterance-final phrases as a function of the amount of input to which it is exposed. Children start out producing OI errors at high rates, and produce

fewer OI errors as the length of their utterances increases. MOSAIC simulates this phenomenon because of the way that compound finites pattern in OI languages. In compound finites, the finite verb precedes the non-finite verb. Since MOSAIC produces increasingly long utterance-final phrases, the early (short) phrases it produces are likely to contain only non-finite verb forms. As the phrases MOSAIC produces become longer, finite modals and auxiliaries start to appear, and OI errors are slowly replaced by compound finites.

The mechanisms implemented in MOSAIC build knowledge of sentence structure by learning sequences from the input and generalising across those sequences. MOSAIC is therefore often regarded as a usage-based model of language learning (cf. Goldberg, 1995; 2006; Lieven, Pine & Baldwin, 1997; Pine, Lieven & Rowland, 1998; Tomasello, 2000; 2003, MacWhinney, 2004). However, although the mechanisms implemented in MOSAIC are certainly consistent with a usage-based analysis, it is important to recognise that MOSAIC is a relatively simple distributional analyser, with no access to semantic information, which is not sufficiently powerful to acquire many aspects of adult syntax. MOSAIC is therefore not itself a realistic model of the language acquisition process. What MOSAIC does provide, however, is a powerful means of testing hypotheses about the relation between cross-linguistic variation in children's early language and cross-linguistic differences in the language to which they are exposed. For example, Freudenthal et al. (2006a) showed that the hypothesis implemented in MOSAIC — that OI errors are truncated compound finites — is able to explain the developmental patterning of the OI phenomenon in two languages: English and Dutch; and Freudenthal et al. (2007) went on to show that this kind of account can explain both the apparently

qualitative difference in the rate of OI errors between Dutch/German and Spanish and the more subtle quantitative difference in the rate of OI errors between Dutch and German. MOSAIC has also been shown to simulate several other phenomena in children's multi-word speech, including the pattern of subject omission errors in English (Freudenthal, Pine & Gobet, 2007), the relation between subject provision and finiteness marking in English and Dutch (Freudenthal et al. 2006b), and differences in the referential properties of OI errors in English, Dutch and German (Freudenthal et al. 2009).

The results of these simulations suggest that it is possible to explain a number of key features of children's early multi-word speech in terms of the interaction between a resource-limited distributional learning mechanism and the surface properties of the language to which young children are exposed. However, they are not necessarily incompatible with a generativist model of language acquisition. Indeed, the account of OI errors implemented in MOSAIC is similar in some respects to a class of generativist models (e.g., Boser, Lust, Santelmann & Whitman, 1992; Ferdinand, 1996; Ingram & Thompson, 1996; Josefsson, 2002), which treat OI errors as finite clauses that contain a null modal. A key problem for such models, however, is that, because they do not include a role for processing limitations in learning, they are unable to explain why OI errors occur so much more frequently in early Dutch and German than modal constructions occur in the input, nor why OI errors occur so much more frequently in early Dutch and German than they do in early Spanish and Italian, in which modal constructions occur at similar rates. The learning mechanism implemented in MOSAIC provides a simple and elegant explanation of both of these phenomena, which achieves a good fit to quantitative data on the rate at which children produce OI errors at different MLU levels in English,

Dutch, German and Spanish. It is unclear whether a generativist version of this mechanism, which modelled OI errors in terms of the interaction between an utterance-final bias in learning and the acquisition of more abstract structural representations, could achieve a similar level of fit to the data.

1.3. Limitations of the current version of MOSAIC

The results of previous studies using MOSAIC show that it is possible to simulate the cross-linguistic patterning of OI errors surprisingly well in terms of the interaction between processing limitations in learning and differences in the distributional properties of the language to which children are exposed. However, one obvious weakness of the version of MOSAIC used in these studies is that it only simulates the cross-linguistic patterning of OI errors in declaratives.

There are two reasons for this state of affairs. The first is that, like many distributional learning mechanisms (though see Frank, Goldwater & Keller, 2013), MOSAIC does not distinguish between declaratives and questions in its input. The current version of the model therefore has no way of distinguishing between representations learned from questions and representations learned from declaratives. This is an important weakness for two reasons. First, it prevents the model from being used to simulate the pattern of OI errors in questions. Second, it results in certain inconsistencies in the way that the model simulates OI errors in declaratives. Thus, although MOSAIC readily simulates OI errors with missing subjects as utterance-final phrases learned from declarative input (e.g. *Go home* from *He wants to go home*), the model simulates OI errors with subjects as

utterance-final phrases learned from questions (e.g. *He go home* from *Did he go home?*)[1]. Given the obvious prosodic, pragmatic and structural differences between declaratives and questions, this feature of the model might be regarded as somewhat implausible, particularly in view of recent evidence that even very young children can distinguish between declaratives and questions in the input (Seidl, Hollich & Jusczyk, 2003; Homae, Watanabe, Nakano, Asakawa & Taga, 2006; Soderstrom, Ko & Nevzorova, 2011, Geffen & Mintz, 2012; 2015; Frota, Butler & Vigário, 2014). For example, Geffen and Mintz (2015) show that by 12 months children can distinguish between declaratives and polar interrogatives even in the absence of prosodic cues, and argue that although infants initially use phonological information to distinguish between sentence types, they have already begun to learn generalisations about the corresponding word-order patterns before the onset of multi-word speech.

The second reason is that MOSAIC learns exclusively from the right edge of the utterance. That is to say, the model only encodes words or strings in the input when all of the material to the right of the target word or string has already been encoded in the network. This feature of the model has the advantage that it allows direct investigation of the effects of assuming an utterance-final bias in learning, uncomplicated by other factors. However, it also means that the model is only able to simulate errors that occur as utterance-final phrases in the input. Thus, MOSAIC is able to simulate truncation errors

[1] In fact, sequences such as 'Daddy do it' do occur in declaratives in certain types of double verb construction in English (e.g. 'We saw Daddy do it'). However, such constructions are too rare to explain the level of OI errors with subjects in English-speaking children's output.

(e.g. *Go there* from *He wants to go there*). However, the model is unable to simulate errors that involve the omission of sentence-internal elements (e.g. *Play train* from *Play with the train*). It is therefore unable to simulate OI errors in Wh- questions (e.g. by learning *Where he go?* from *Where did he go?).*

In view of these problems, the aim of the present paper is to describe a new version of MOSAIC that eliminates the weaknesses of the previous version by 1) distinguishing between declaratives and questions in its input, and 2) learning from both edges of the utterance. The new version of the model represents utterance-internal omission errors as concatenations of utterance-initial and utterance-final strings and hence has the potential to simulate both OI errors with subjects in declaratives (e.g. *He _ go there*) and OI errors in Wh- questions (*Where _ he go?*). As a result, it has several important advantages over the previous version of the model.

First, by incorporating some sensitivity to the left as well as the right edge of the utterance, it implements processing limitations in learning in a more realistic way. For example, it is consistent with a wealth of evidence that humans show primacy as well as recency effects in sequence and list learning (Murdock, 1962; Jahnke, 1965; Bellezza, Andrasik & Lewis, 1982; Hulme, Roodenrys, Schweickert, Brown, Martin & Stuart, 1997; Gupta, 2005; Gupta, Lipinski, Abbs & Lin, 2005)[2].

[2] This issue is not as straightforward as it might at first appear, since recency effects are generally taken to reflect capacity limitations in short-term memory (which *are* likely to be a factor in language learning), whereas primacy effects are often taken to reflect active elaboration processes such as rehearsal (which *are not* likely to be a factor, at least in younger children). However, primacy effects have often been found in humans in the

Second, it allows the model to represent a much wider range of error types. For example, it allows the model not only to represent truncated utterances (e.g. *(He wants to) go there*, but also utterances with missing utterance-internal elements (e.g. *Kick (the) ball* or *He (wants to) go there*). This is an important development since sentence-internal omission errors of this kind are a characteristic feature of children's early multi-word speech, which is often described as 'telegraphic' because of the high frequency with which such errors occur (Brown, 1973).

Third, and most important, it has the potential to provide an integrated account of the cross-linguistic patterning of OI errors in both declaratives and Wh- questions. Of course, it should be noted that a model that simulates OI errors through utterance-internal omission has the potential to produce OI errors in declaratives and Wh- questions not only in English, but also in Dutch, German and Spanish. It is therefore not at all clear that MOSAIC will be able to simulate both the high levels of OI errors in English Wh-questions and the low levels of OI errors in Dutch, German and Spanish Wh- questions. This issue will be addressed in the present study by running simulations of children's use of Wh- questions across the four languages using the same version of the model as that

---

absence of rehearsal (e.g. Neath, 1993; Sikström, 2006). Moreover, primacy (and recency) effects have also been found in non-human species (monkeys and pigeons), where active elaboration processes are unlikely to play a major role (Wright, Santiago, Sands, Kendrick & Rook, 1985). These findings can be explained in terms of the increased (temporal) distinctiveness of items occurring at the beginning and end of a list or sequence, and suggest that language-learning children are likely to be preferentially sensitive to both the beginning and the end of unfamiliar utterances.

used to simulate the pattern of OI errors in children's declarative utterances. The only difference between the declarative simulations and the Wh- question simulations is that the declarative simulations will focus on utterances learned from declarative input and the Wh- question simulations will focus on utterances learned from interrogative input.

To summarise, the aim of the present paper is to describe a new version of MOSAIC that learns from both the right and the left edge of the utterance, and to investigate the extent to which it is able to provide an integrated account of the cross-linguistic patterning of OI errors in declaratives and Wh- questions. The new version of MOSAIC is designed to learn structures with missing sentence-internal elements and hence to simulate a wider range of error types than previous versions of the model. It also differentiates between declaratives and questions in the input, learning declarative structures from declarative input and interrogative structures from interrogative input. This allows us to investigate whether it is possible to simulate differences in the cross-linguistic patterning of OI errors across declaratives and Wh- questions in terms of the interaction between the same learning mechanism and cross-linguistic differences in the distributional patterning of declarative and interrogative input.

The remainder of the paper is organised as follows. First, we describe the new version of MOSAIC, together with its mechanism for producing utterances with missing sentence-internal elements. Second, we report simulations aimed at determining whether this new version of MOSAIC is able to provide a good quantitative fit to the Optional Infinitive phenomenon in declarative utterances. Third, we report simulations of the cross-linguistic patterning of OI errors in Wh- questions, before concluding with a discussion of the implications of our results for the field as a whole.

2. MOSAIC

MOSAIC is an unsupervised learning mechanism that accepts as input corpora of orthographically transcribed child-directed speech, and produces as output 'child-like' utterances that become progressively longer as learning proceeds. Some of these utterances are produced by rote (i.e. have occurred as utterances or parts of utterances in the input). Others are produced generatively (i.e. by substituting words into frames that have occurred as utterances or parts of utterances in the input). Since the average length of MOSAIC's output increases with learning, MOSAIC can be used to simulate developmental changes in children's speech as a function of increasing MLU.

2.1. The MOSAIC Network

The basis of MOSAIC is an n-ary discrimination net that consists of nodes and arcs connecting those nodes. At the head of the network is an empty root node. Nodes directly beneath the root node are called primitive nodes and are used to store the words that MOSAIC encodes. Nodes at deeper levels in the network are used to encode the phrases that have been encoded in the model. The arcs or 'tests links' that connect nodes are used to store the difference between the nodes they connect. A MOSAIC network is slowly built up from exposure to the input it receives. As MOSAIC sees more input it creates more nodes, and nodes at deeper levels in the network. The model thus stores (and is hence able to produce) more and longer utterances as a function of learning.

2.2. Edge-first learning in MOSAIC

A central feature of MOSAIC is that it learns from the edge of the utterance. The version of MOSAIC used in the present study learns from both the right and left edges. However, previous versions of MOSAIC have learned exclusively from the right edge. For clarity

of exposition, we first describe right edge learning. We then describe left edge learning, and the way in which the products of right edge learning and left edge learning are combined within the model.

2.2.1. Right edge learning

To illustrate right edge learning, consider an empty MOSAIC network that is shown the utterance: *He goes home* a number of times. MOSAIC's right edge learning mechanism will only encode a word or phrase when everything that follows that phrase in the utterance has already been encoded in the network. This mechanism can be likened to a moving window or buffer. Whenever an unknown word or word transition is encountered, the contents of the buffer are emptied, and only the most recently encountered word is left as a target for encoding. Thus, on the first presentation of *He goes home*, the model will process the utterance in a left-to-right fashion. When it reaches the end of the utterance, the buffer will contain the word *home*, and a node for this word will be created. On the second presentation of the utterance, the buffer will contain the phrase *goes home* when reaching the end of the utterance. The model will now attempt to encode this phrase, but since no primitive node exists for the word *goes* this primitive node will be created first. On the third presentation, a non-primitive node containing the phrase *goes home* will be created. A fourth presentation will result in a primitive node for the word *he*, while a fifth presentation will result in the creation of a non-primitive node for the phrase *he goes home*. Fig. 1 shows the model after five presentations of the utterance *He goes home*.

Right edge learning has the effect of restricting the strings represented in MOSAIC to utterance-final sequences that have occurred in the input (or novel sequences based on

utterance-final frames that have occurred in the input — see section on generating output from MOSAIC below). Previous work with MOSAIC has shown that this mechanism allows the model to simulate variation in the developmental patterning of OI errors across several different languages (Freudenthal et al., 2006a; 2007; 2009; 2010).

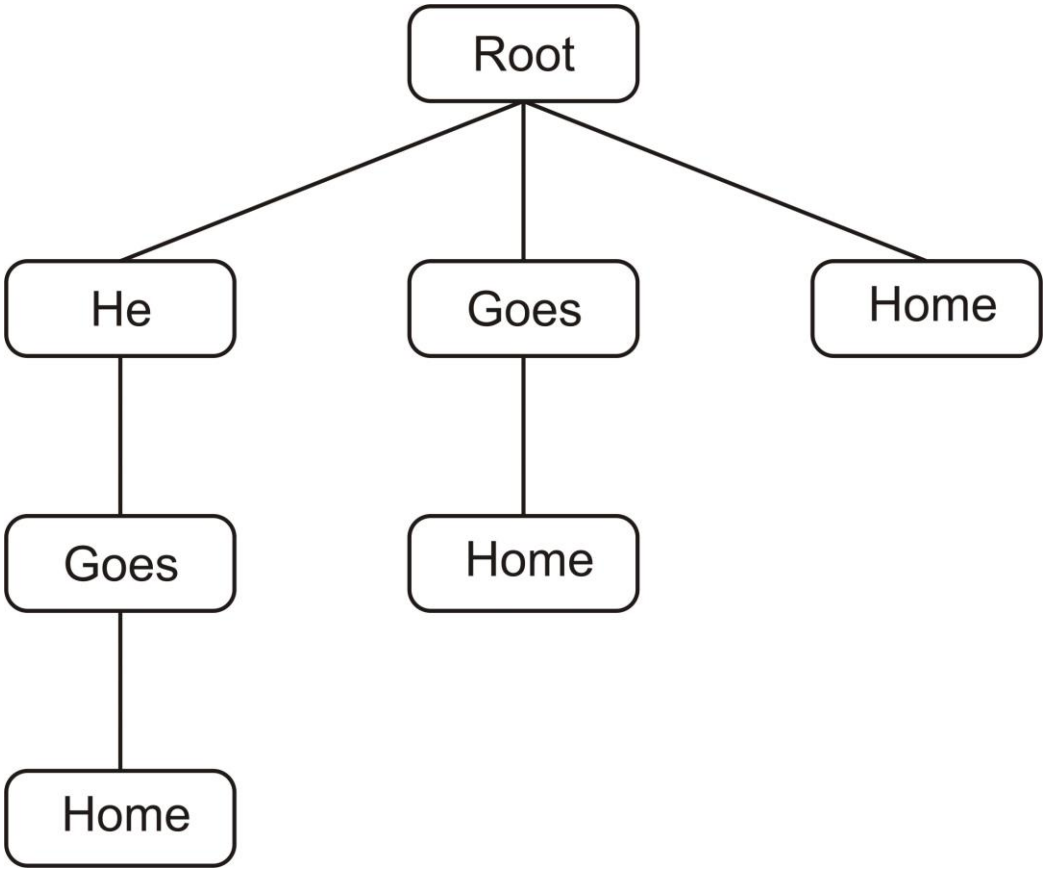


Fig. 1: A MOSAIC network after it has seen the phrase *He goes home* five times.

2.2.2. Left-edge learning

Left edge learning works in a similar way to right edge learning, except that it is anchored at the left edge of the utterance and restricted to a single word (or chunk — see section on chunking below). MOSAIC thus builds up its representation from both edges

of the utterance, displaying a (small) primacy and a (larger) recency effect in learning. However, MOSAIC also combines the products of right and left edge learning by associating utterance-initial and utterance-final elements on the basis of their co-occurrence in utterances in the input. This process is illustrated in Fig. 2. The network depicted in Fig. 2 has encoded the utterance-initial word *he* and the utterance-final phrases *go home* and *go away*. The word *he* and the phrase *go home* have been associated on the basis of their co-occurrence in the utterance *He can go home*, allowing the model to produce the utterance *he go home*. Since the difficulty of associating utterance-initial and utterance-final elements would be expected to increase as a function of the distance between the elements in the target utterance, the probability (p) of associating utterance-initial and utterance-final elements within the model is governed by the formula: $p = 1/d$, where d is the number of words intervening between the relevant utterance-initial and utterance-final elements.

The addition to the model of left edge learning and a mechanism for associating the products of right and left edge learning has the effect of expanding the range of strings represented in MOSAIC to include utterance-initial words and chunks and strings with missing utterance-internal elements. Restricting left edge learning to one word or chunk has the effect of ensuring that utterance-internal omission errors occur near the left edge of the target utterance. Together these mechanisms allow MOSAIC to learn both OI errors with subjects in declaratives (e.g. *He go home*) from declarative input, and OI errors in Wh- questions (e.g. *Where (does) that go?*) from interrogative input.

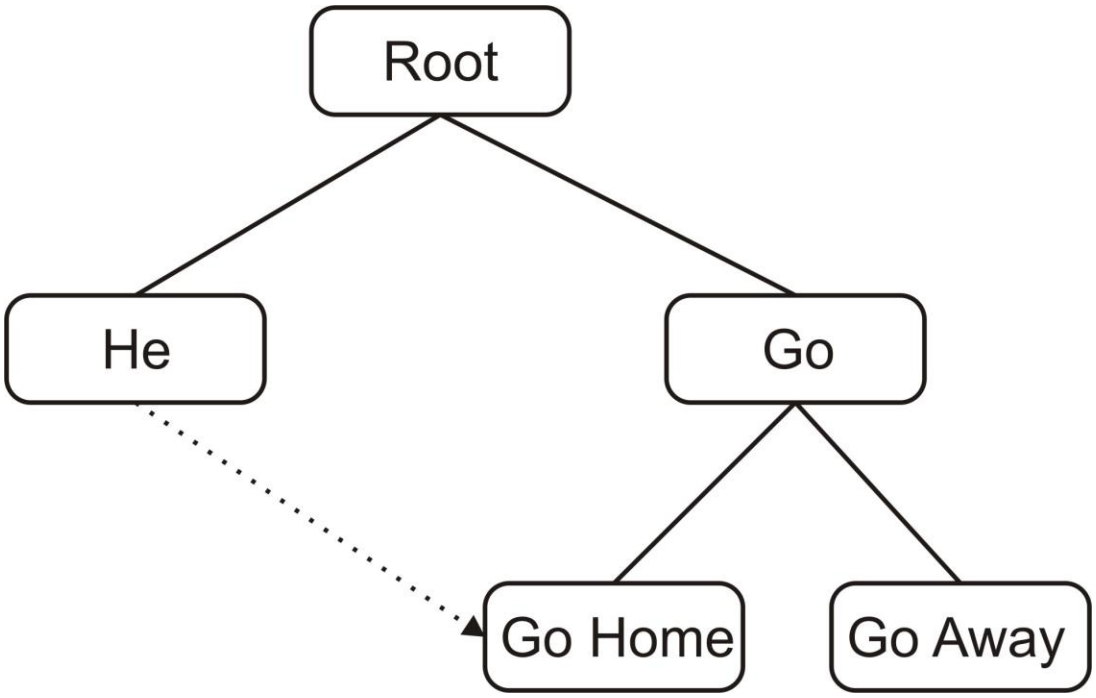


Fig. 2: A MOSAIC network that has associated the word *He* and the phrase *Go home.*

2.3. Learning Speed

In the examples presented so far, MOSAIC has been described as if nodes were created in the model whenever the opportunity arose. In fact, however, node creation in MOSAIC is slow and probabilistic, and words and phrases typically have to be encountered many times before they are encoded. Input corpora are fed through the model several times and the node creation probability (NCP) increases with every exposure to the input. Node-creation is governed by the following formula:

$$NCP = \left(\frac{1}{1 + e^{(m - u/c)/t}} \right)^{\sqrt{d}}$$

where: NCP = Node Creation Probability,

m = a constant, set to 40 for these simulations,

c = corpus size (number of utterances),

u = total number of utterances seen,

t = a constant, set to 3 for these simulations,

d = distance to the edge of the utterance.

The formula for Node Creation Probability and the role of its parameters is explored in more detail in Appendix A (Node Creation Probability).

2.4. Generating output from MOSAIC

There are two mechanisms for generating output from MOSAIC. The first is a rote production mechanism that is only able to produce utterances that were present in the model's input. The second is a generative production mechanism that allows the model to produce novel utterances by substituting items for each other on the basis of their distributional similarity. The first of these processes involves traversing all of the branches in the network. When a terminal node (or end-of-utterance marker) is encountered, the phrase encoded in that branch is produced together with concatenations of that phrase and all the utterance-initial words and chunks with which it has been associated. This process results in the production of all the (utterance-final) phrases and all the utterances with missing internal elements that are represented in the network. The only restriction applied to this mechanism is that utterance-initial and utterance-final elements can only be concatenated if they are anchored. That is to say, utterance-initial elements can only be concatenated if the last word in the element has occurred in utterance-final position, and utterance-final elements can only be concatenated if the first

word in the element has occurred in utterance-initial position. This restriction is designed to increase the plausibility of MOSAIC's output by ruling out sentence internal omission errors in which the beginning or the end of the utterance is not a semantically plausible unit. For example, it prevents the model from producing utterances such as *The (boy) wants a drink* on the assumption that, whatever the status of young children's grammatical knowledge, they would be unlikely to make such errors because including the word *the* without including the word that it modifies (in this case *boy*) would make neither semantic nor pragmatic sense.

The generative production mechanism works by tracking the contexts (preceding and following words) in which the words encoded in the model have been encountered in the input. Words that share a sufficient amount of overlap (20% for the present simulations) in both the words that precede and the words that follow them are considered equivalent and can be substituted when generating output from the model, resulting in phrases and concatenations that were not present in the input. Several authors have shown that words with high overlap in terms of shared sentential contexts are likely to be of the same syntactic class (e.g. Redington, Chater & Finch, 1998; Mintz 2003).

2.5. Chunking

The generativity mechanism described above generally results in substitutions of items from the same syntactic class. Substitutions that appear of good quality in one situation, however, may result in anomalous utterances in other situations. For example, *do* and *make* tend to occur in similar contexts in the input because they are both transitive main verbs. However, *do* is also an auxiliary (e.g. *Do you want a cookie?*), whereas *make* is not. Substitution of *make* for *do* in main verb contexts tends to result in grammatical

and/or child-like utterances. However, substitution of *make* for *do* in auxiliary contexts tends to result in utterances that are neither grammatical nor typical of child speech. MOSAIC avoids such unwanted substitutions by implementing a chunking mechanism. According to theories based on chunking (Chase & Simon, 1973; Gobet & Simon, 1998; Gobet et al., 2001), frequently encountered stimuli are grouped into larger structures that can be retrieved as one unit. Chunking in MOSAIC is implemented by registering the frequency with which nodes in the model are visited when processing the model's input. When the frequency for a multi-word phrase exceeds a predetermined threshold, a new node at the primitive level is created to encode that multi-word phrase. The multi-word phrase is subsequently treated as one unit by the generativity mechanism. That is, individual words that make up the multi-word phrase cannot be substituted in the context of the newly created chunk. The chunking mechanism, which is described in more detail in Freudenthal, Pine and Gobet (2005), also allows MOSAIC to learn utterances that contain frequent phrases more quickly than would otherwise be the case, and plays an important role in determining the type of utterances with missing internal elements that the model represents. Thus, the fact that left edge learning is restricted to one chunk rather than one word means that, although utterance-initial omission errors tend to be anchored at the left edge of the utterance (e.g. *That (can) go there* or *Where (does) that go?*), they can also occur further to the right of the utterance, provided the target utterance begins with a high frequency string (e.g. *That one (can) go there* or *Where is (he) going?*).

2.6. Running Simulations in MOSAIC

Simulations are run in MOSAIC by feeding the input corpus through the model several times and generating an output file after each cycle through the input corpus. This output file consists of rote-learned utterance-final phrases together with internal-omission rote-learned phrases (i.e. concatenated utterance-initial and utterance-final phrases); and generated (novel) utterance-final phrases together with internal-omission generated phrases (i.e. concatenated utterance-initial and utterance-final phrases where one of the two phrases is novel). both utterance-final strings and concatenations of utterance-initial and utterance-final strings that have been learned directly from the input, and utterances generated by substituting words into utterance-final frames and concatenations of utterance-initial and utterance-final frames using the generativity mechanism described above. Because the size of the network grows as a function of the amount of input to which the model has been exposed, the average length of the model's output increases with every cycle through the input corpus. Output files can therefore be selected for analysis on the basis of their MLU and compared with data from children at the same stage of development. Because MOSAIC's output consists of sets of utterance types rather than utterance tokens, corpora of child utterances are also reduced to sets of utterance types before matching output files are selected.

3. The Simulations

3.1 Input corpora

The corpora used as input for the present simulations were those of Anne and Becky (English), Matthijs and Peter (Dutch), Juan (Spanish) and Leo (German). These are the same corpora used in Freudenthal et al. (2007). The data for the English children are part of the Manchester corpus (Theakston, Lieven, Pine & Rowland, 2001), and consist of

approximately 33,000 input utterances for Anne and 27,000 input utterances for Becky. The data for the Dutch children are part of the Groningen corpus (Bol, 1996; Wijnen, Kempen & Gillis, 2001) and consist of approximately 14,000 input utterances for Matthijs and 13,000 input utterances for Peter. The data for the Spanish child are part of the Madrid corpus (Aguado-Orea & Pine, 2015) and consist of approximately 25,000 input utterances. The data for the German child constitute a dense corpus (consisting of over 140,000 parental utterances) made available to us by the Max Planck institute for Evolutionary Anthropology in Leipzig (Behrens, 2006). Since this dataset is considerably larger than all of the other corpora, a random sample of 30,000 maternal utterances was extracted from the German corpus to serve as input to the model.

3.2. Preparation of the input corpora

Preparation of the input corpora involved extracting all parental speech from the relevant transcripts, and deleting all filler material, pauses, hesitations, false starts, and utterances where one or more words were unintelligible to the transcriber. This is the same procedure as that used in Freudenthal et al. (2007). However, preparation of the input for the present simulations departed from Freudenthal et al. in one important respect. Freudenthal et al. treated all input as if it was declarative. That is, question marks in the transcripts were ignored. For the present simulations, this information was retained in the utterances extracted from the transcripts, and all words in interrogative utterances where marked for being part of a question. Marking words in this way makes it possible to determine whether (partial) utterances have been learned from questions in the input, and thus allows the separation of MOSAIC's output into utterances learned from interrogative input and utterances learned from declarative input. It is thus a way of implementing the

assumption that children can distinguish between declaratives and questions before the onset of multi-word speech. Note, however, that treating words learned from declaratives and words learned from questions as separate entries is not intended as a realistic way of representing children's knowledge of the words themselves; it is simply a convenient way of allowing the model to distinguish between declaratives and interrogatives, which we know young children are able to do. Moreover, although it does have some effect on the speed with which the model learns particular words (because it changes the frequency statistics of words in the input), previous work with MOSAIC has shown that it actually has little effect on the output of the model. For example, Freudenthal et al. (2009) used the same strategy to keep track of information about whether infinitive forms had been learned from modal or non-modal contexts and found that representing infinitive forms that had and had not been learned from modal contexts as separate lexical entries had very little effect on the patterning of finite and non-finite forms in the model's output. We therefore decided to use this strategy again in the new version of the model.

3.3. Preparation of the child data

Preparation of the child data proceeded in a similar manner to the preparation of the input data. The transcripts of individual recording sessions were aggregated into batches reflecting different MLU points or developmental stages. For Dutch and German, 4 different stages between MLU 1.5 and 4.0 were selected. For English and Spanish the first MLU point was dropped, resulting in three developmental stages. For English, the first MLU point was dropped because the analysis was restricted to utterances with third person singular subjects, which are rare in the data until the child's MLU has reached a reasonably high level. For Spanish, the first MLU point was dropped because the Spanish

child did not produce a sufficiently large number of utterances at this point in development.

3.4. Running the simulations

Simulations were run by feeding the input corpora through the model several times and generating output (of increasing average length) after each exposure to the input set. The output files that most closely matched the MLU of the children at the different developmental stages were then selected for analysis. In order to establish the reliability of the results, 10 simulations were run for each child by feeding the same input corpus through models that were initialized with different random seeds. The results were then averaged across the 10 simulations and compared with the results for the corresponding child. For the simulations of OI errors in declaratives, Root Mean Square Error measures (RMSEs) were used to assess the fit between the child and the average model. RMSEs provide a measure of the extent to which the patterning of the data across several different categories (e.g., non-finite, simple finite and compound finite) diverges between two cases (e.g., child and model). Comparing the RMSEs for the simulations using the old and the new versions of the model is therefore a good way of identifying any substantive changes in the model's fit to the child data. For the simulations of OI errors in Wh- questions, RMSEs were not used, since, for these simulations, there were only two categories: Wh- questions that included an OI error and Wh-questions that did not (see below). For the simulations of OI errors in declaratives, analysis was restricted to declarative output that included at least one verb other than the copula. For the

simulations of OI errors in Wh- questions, analysis was restricted to non-subject question output that contained both a Wh- word and a main verb. Sample output from declaratives and Wh- questions for two English models is provided in Appendix B.

3.5. Coding and data analysis

3.5.1. OI errors in declaratives

The analysis of OI errors in declaratives employed the same analysis routines that were used in Freudenthal et al. (2007). Analysis was restricted to utterance types rather than tokens and each utterance was assigned to one of three categories: simple-finite utterances, compound-finite utterances or non-finite utterances.

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 or German, and utterances containing third person singular verb forms and irregular past tense verb forms in English).

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 a form matching the infinitive in Dutch or German, and utterances containing a modal and an infinitive or an auxiliary and a perfect or progressive participle in English).

Non-finite 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 German and utterances containing zero-marked verb forms in English).

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 are some finite verb forms in Dutch, German and English that are indistinguishable from the infinitive. Thus, although there is strong evidence that Dutch, German 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 as opposed to a finite plural present tense verb form (in Dutch and German) 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. As Freudenthal et al. (2007) point out, this does not affect the validity of any analysis of the closeness of the fit between the data of the children and their respective models because the child data are analyzed in exactly the same way. Nevertheless, it could be argued that the level of ambiguity in English is so high that treating all zero-marked forms in the same way makes it trivially easy to simulate the data. For this reason, the analysis of English was restricted to utterances that contained a third person singular subject (e.g. *He go(es)),* since the provision of a zero marked form in a third person singular context is clearly incorrect.

3.5.2. OI errors in Wh- questions

The analysis of OI errors in Wh- questions was restricted to non-subject questions that contained both a Wh- word and a main verb[3]. As with the analysis of OI errors in

[3] Note that this definition excludes Wh- questions from which the Wh- word has been omitted. Such questions occur in the speech of English, Dutch, German and Spanish

declaratives, the main focus of this analysis was on determining whether the utterances produced by the children and the models were marked for finiteness. For English, this means that the source of ambiguity that exists in declarative utterances is removed. In English Wh- questions, finiteness marking is carried by the modal or auxiliary. Thus, any utterance that contains a non-finite verb and lacks a modal or auxiliary (e.g. *Where* go?) can automatically be classified as an OI error. It is therefore unnecessary to restrict the analysis of Wh- questions to utterances containing a third person singular subject. For Dutch, German and Spanish, the analysis proceeded in a similar manner. Utterances that contained a finite main verb or modal/auxiliary were classified as finite, whereas utterances that contained only a non-finite main verb were classified as non-finite. However, since in Dutch and German Wh- questions, non-finite and plural finite verb forms can be distinguished on the basis of their position in the utterance (i.e. non-finite verb forms occur in utterance-final position whereas plural finite verb forms are always followed by a plural NP subject), plural verb forms were classified as finite if they were followed by a plural NP subject. Thus, the correctly formed Wh- question: *Wo spielen sie?* (Where play they?) was classified as finite, whereas the incorrect Wh- Questions *Wo sie spielen?* (Where they play?) and *Wo spielen?* (Where play?) were classified as non-finite.

---

children during the early stages, and are readily simulated by the model as utterance-final phrases learned from Wh- questions in the input. However, they are not relevant to Wexler's analysis of the cross-linguistic patterning of OI errors in Wh- questions since they do not include a Wh- word (which is the factor that Wexler assumes triggers the use of a finite form in Dutch and German)

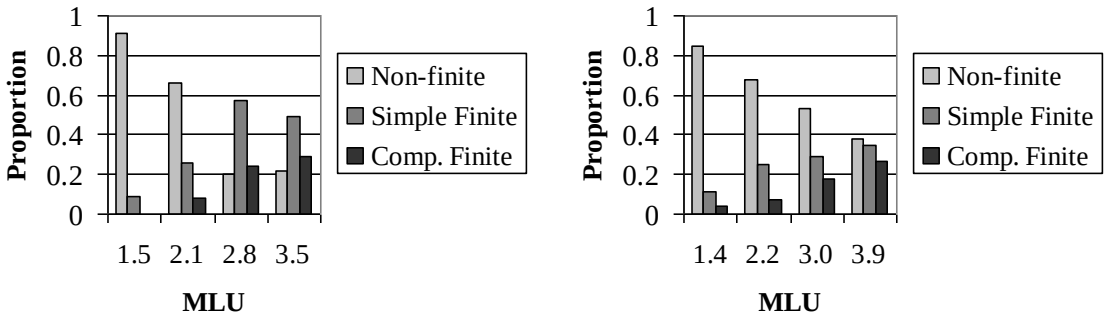
4. Results

4.1. Modelling the rate of OI errors in declaratives in Dutch, Spanish and German.

One of the key strengths of MOSAIC is that it provides a unified cross-linguistic account of the developmental patterning of OI errors in declaratives. In a first set of simulations we therefore investigate whether it is still possible to simulate this patterning using the new version of the model. Figures 3 and 4 show the results for the Dutch simulations with the new version of the model (c), as well as the simulations (b) and child data (a) reported in Freudenthal et al. (2007). The figures for the new simulations include error bars because they report results averaged across 10 runs of the model, each of which was initialised with a different random seed.

It is clear from these figures that the new simulations show the same pattern as the old simulations. Root mean squared errors (RMSEs) for the old simulations are .04, .01, .21 and .12 for Matthijs and .03, .07, .25 and .10 for Peter. RMSEs for the averages of the new simulations are .09, .02, .25 and .13 for Matthijs and .01, .02, .27 and .04 for Peter. Thus, although neither the old nor the new versions of the model capture all of the fine detail of the child data (particularly the pronounced drop in the proportion of OI errors between the second and third MLU points), both models simulate the very high rates of OI errors in Dutch during the early stages, and the substantial drop in the level of OI errors across the MLU range.

(a) Data for Matthijs (b) Old Model for Matthijs



(c) New Model for Matthijs

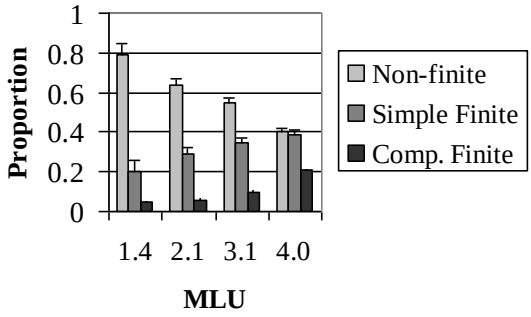
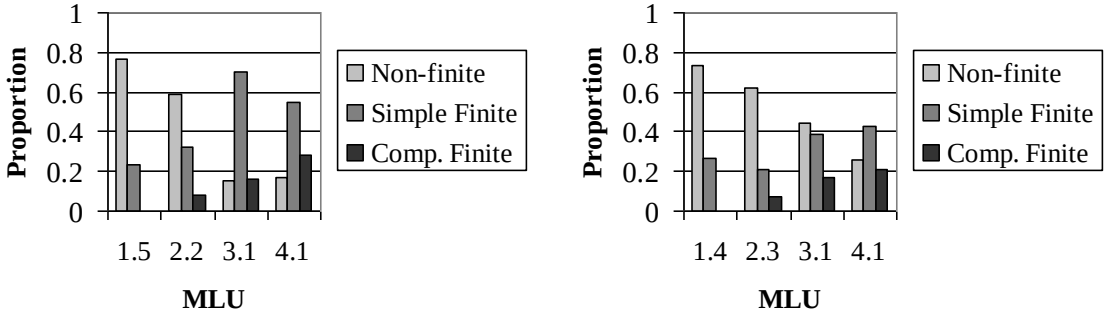


Fig. 3: Data, old and new model for Matthijs.

(a) Data for Peter

(b) Old Model for Peter



(c) New Model for Peter

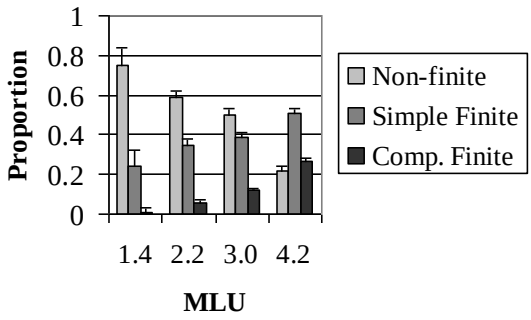
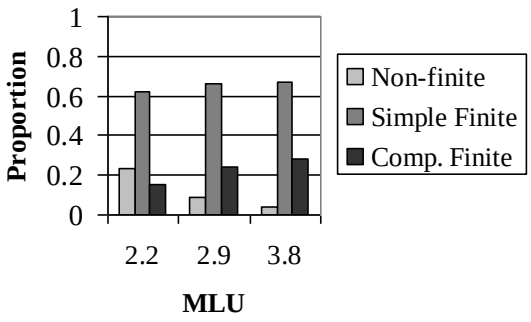


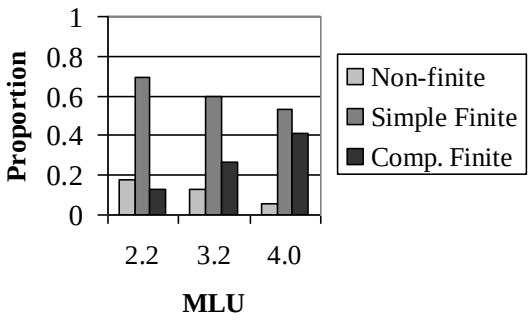
Fig. 4: Data, old and new model for Peter.

Figure 5 shows the results for the Spanish simulation with the new version of the model (c), as well as the simulation (b) and child data (a) reported in Freudenthal et al. (2007). Again it is clear that the new simulations show the same pattern as the old simulation. RMSEs for the old simulation are .06, .05 and .11. RMSEs for the average of the new simulations are .09 .06 and .09. Thus, although again there are some minor differences between the child and the model data, both models simulate the low level of OI errors in Spanish across the MLU range and the high level of simple-finite utterances.

(a) Data for Juan



(b) Old Model for Juan



(c) New Model for Juan

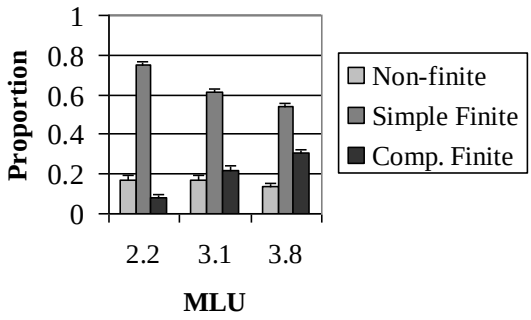
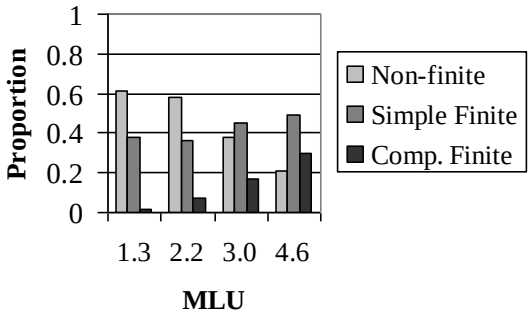


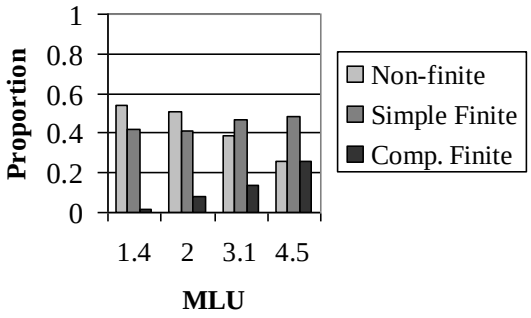
Fig. 5: Data, old and new model for Juan.

Figure 6 shows the results for the German simulation with the new version of the model (c), as well as the simulation (b) and the child data (a) reported in Freudenthal et al. (2007). These results are also very similar to those for the old simulation. RMSEs for the old simulation are .05, .05, .02 and .04. RMSEs for the average of the new simulations are .08, .06 .03 and .08. Thus, like the old model, the new model simulates the lower level of OI errors in German than in Dutch during the early stages, and the less pronounced drop in the level of OI errors across the MLU range.

(a) Data for Leo



(b) Old Model for Leo



(c) New Model for Leo

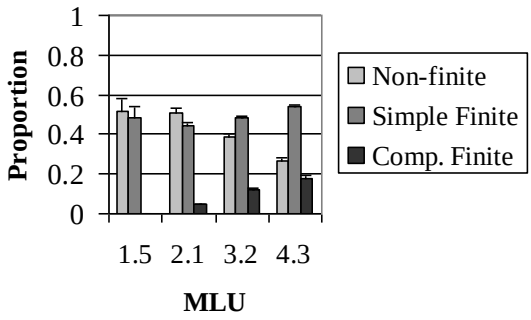


Fig. 6: Data, old and new model for Leo.

It is clear from these results that the new version of MOSAIC continues to simulate both the apparent qualitative difference in the rate of OI errors between Dutch/German and Spanish and the more subtle difference in the rate of OI errors between Dutch and German. These results show that neither the addition of left edge learning, nor the exclusive focus on declaratives in the model's input, seriously affect MOSAIC's ability to simulate the cross-linguistic data. It is therefore possible to simulate the developmental patterning of OI errors in Dutch, Spanish and German declaratives in terms of a mechanism that learns from both edges of the utterance and learns declaratives exclusively from declarative input.

4.2. Modelling the rate of OI errors in declaratives in English

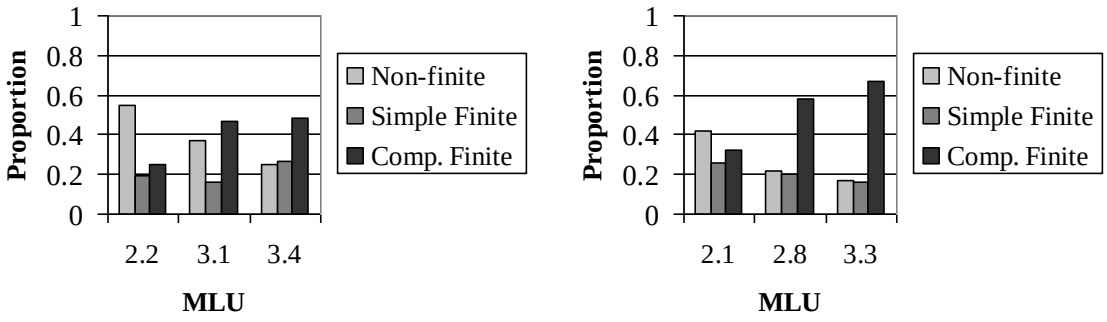
In addition to simulating the pattern of OI errors in Dutch, German and Spanish declaratives, MOSAIC is also able to simulate the developmental patterning of OI errors in English declaratives. However, simulating the pattern of OI errors in English is complicated by the fact that English is such a morphologically impoverished language that it is only possible to unambiguously identify OI errors in a restricted set of contexts.

For this reason, previous work with MOSAIC has focused on the pattern of OI errors in utterances with third person singular subjects (e.g. *He go out* instead of *He goes out* and *She do it* instead of *She does it*). Simulating this pattern is a particularly strong test of the new version of MOSAIC because strings such as *He go out* and *She do it* do not occur as sequences in English declarative input. The new version of the model is thus totally reliant on its mechanism for concatenating utterance-initial and utterance-final phrases when producing such errors.

Figures 7 and 8 show the results for the English simulations using the new version of the model (c), as well as the simulations (b) and child data (a) reported in Freudenthal et al. (2007). It is clear from these Figures that, despite the changes made to the underlying mechanism, the new version of the model continues to simulate the developmental data remarkably well. RMSEs for the old simulations are .09, .11 and .13 for Anne and .13, .07 and .05 for Becky. RMSEs for the averages of the new simulations are .12, .08 and .05 for Anne and .04, .09 and .08 for Becky.

(a) Data for Anne

(b) Old Model for Anne



(c) New Model for Anne

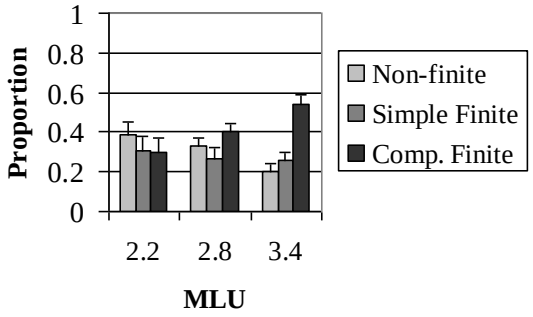
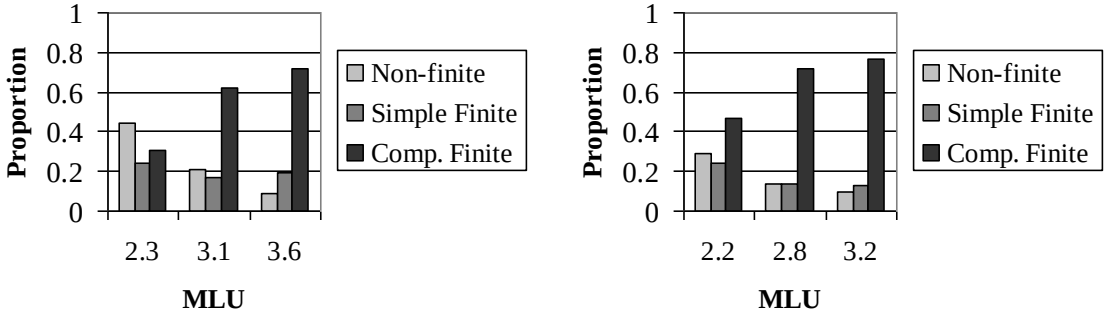


Fig. 7: Data, old and new simulations for Anne.

(a) Data for Becky (b) Old Model for Becky



(c) New Model for Becky

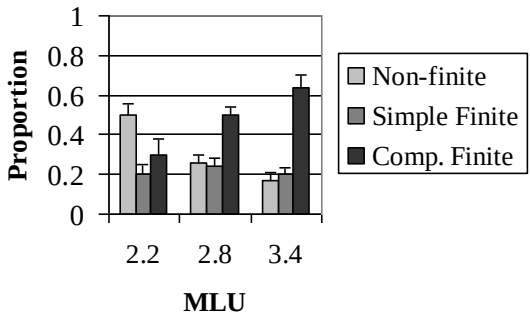


Fig. 8: Data, old and new models for Becky.

These results show that it is possible to replace MOSAIC's old mechanism for simulating OI errors with subjects with a new mechanism that concatenates utterance-initial words or chunks and utterance-final phrases. This mechanism allows MOSAIC to learn OI errors with subjects from subject + modal/auxiliary + verb structures rather than from questions in the input, and hence eliminates an important weakness of the earlier version of the model.

4.3. Simulating the pattern of OI errors in English, German and Spanish Wh- questions

The simulations reported above show that it is possible to develop a unified account of the development patterning of OI errors in English, Dutch, German and Spanish declaratives without relying on interrogative input. However, distinguishing between declaratives and questions in the input also has the advantage that it opens up the possibility of simulating the cross-linguistic patterning of OI errors in both declaratives and Wh- questions using exactly the same learning mechanism.

Table 2: Proportions of OI errors in Wh- questions for Anne, Becky, Leo and Juan at three developmental stages (number of contributing utterances and MLU in parentheses).

	T1	T2	T3
Anne (English)	.70 (47, 2.49)	.42 (45, 3.29)	.07 (27, 3.64)
Becky (English)	.69 (42, 2.47)	.42 (24, 3.22)	.08 (178, 3.68)
Leo (German)	.17 (6, 2.40)	.00 (34, 3.57)	.00 (22, 3.76)
Juan (Spanish)	.00 (23, 2.46)	.03 (35, 3.45)	.00 (48, 3.96)

Table 2 provides data on the proportion of OI errors in Wh- questions at three MLU points for the English, German and Spanish children whose declarative data were simulated earlier. Data from the two Dutch children are not included since neither produced sufficiently high numbers of Wh- questions to allow a meaningful analysis to be performed. It can be seen from Table 2 that, consistent with the previous literature, OI errors occur at much higher rates in the English than in the German and Spanish children during the early stages (0.70 in Anne and 0.69 in Becky compared with .17 in Leo and .00 in Juan). However, it is also clear that the German child, Leo, and the Spanish child, Juan, produce few Wh- questions of any kind at T1. In order to obtain a more reliable estimate of the proportion of OI errors in Wh- questions in German and Spanish during the early stages, we therefore analysed the data from the 6 typically developing German-speaking children in the Szagun corpus (Szagun, 2001) and 5 Spanish-speaking children from the Aguirre, Irene, Marrero/Albalá and Ornat corpora (López Ornat, 1994), all of which are available in the CHILDES database (MacWhinney, 2000). The results of this analysis are presented in Table 3.

Table 3: Proportions of OI errors in Wh- questions in German and Spanish at three developmental stages (number of contributing utterances and MLU in parentheses).

German			
	T1	T2	T3
Anna	.14 (14, 2.41)	.00 (53, 3.09)	.00 (65, 3.49)
Emely	N/A (0, 2.34)	.00 (2, 2.91)	.00 (12, 3.83)
Falko	.25 (4, 2.56)	.13 (8, 3.30)	.00 (16, 3.76)
Lisa	.00 (4, 2.47)	.00 (12, 2.99)	.00 (29, 3.86)
Rahel	.22 (9, 2.56)	.00 (17, 3.87)	.00 (34, 4.50)
Soeren	.33 (3, 2.47)	.13 (8, 3.63)	.02 (86, 3.85)
Spanish			
	T1	T2	T3
Alfonso	.00 (8, 2.66)	.00 (3, 3.14)	N/A
Idaira	.00 (9, 2.45)	.00 (15, 3.38)	.00 (17, 3.98)
Irene	.00 (5, 2.54)	.00 (23, 3.11)	.00 (165, 3.93)
Magda	.00 (56, 2.52)	.00 (29, 3.07)	N/A
María	.00 (10, 2.44)	.00 (9, 3.26)	.00 (14, 3.94)

Although there are still few Wh- questions during the early stages in most of the children's speech, the data presented in Table 3 confirm that the rate of OI errors in Wh-questions is much lower in German and Spanish than it is in English. However, they also suggest that the rate of OI errors in Wh- questions is higher in German than it is in Spanish during the early stages. Thus, although none of the Spanish children produced OI errors in Wh- questions at T1, the average rate of OI errors in German at T1 is .23, with most of the German children producing at least one OI error (see Table 4 for a list of

examples). The implication is that there is an early stage during which German children do make OI errors in Wh- questions, but that such errors occur at much lower rates in German than they do in English.

Table 4: Examples of OI errors in German Wh-questions

Child	Error	Gloss
Leo	Wo stecken?	Where put-INF?
Anna	Wo mit machen? Wo Claudi die Tasse ausspuelen? Wo machen? Was mit machen?	Where with do-INF? Where Claudi the cup rinse-INF? Where do-INF? What with do-INF?
Falko	Wo der passen?	Where that fit-INF
Rahel	Wo malen? Wo das gemacht	Where paint-INF Where that make-PERF
Soeren	Wo da hingehen	Where there go-INF?

Table 5 presents results for the English, German and Spanish simulations. In each case, the proportion of OI errors is an average based on 10 runs of the model. It is clear from Table 5 that, although MOSAIC underestimates the level of OI errors in English at T1, it simulates both the higher rate of errors in English than in German and Spanish and the higher rate of errors in German than in Spanish. It also simulates the pronounced drop in the proportion of OI errors in Wh- questions in English across the MLU range. MOSAIC thus provides a reasonable fit to the cross-linguistic patterning of OI errors in Wh- questions.

Table 5: Average Proportions of OI errors for Anne, Becky, Leo and Juan's models at three developmental stages (SD and Average MLU in parentheses).

	T1	T2	T3
Anne (English)	.48 (.06, 2.59)	.30 (.06, 3.10),	.16 (.05, 3.80)
Becky (English)	.48 (.06, 2.58)	.33 (.05, 3.16)	.23 (.05, 3.69)
Leo (German)	.24 (.05, 2.51)	.11 (.04, 3.17)	.07 (.03, 3.78)
Juan (Spanish)	.04 (.01, 2.48)	.02 (.00, 3.13)	.01 (.01, 3.83)

In order to understand why MOSAIC simulates these cross-linguistic differences, it is necessary to analyse the distributional properties of Wh- questions in the three languages, and the way in which these properties interact with MOSAIC's learning mechanism. Thus, in English, (non-subject) Wh- questions are formed by a process involving subject/auxiliary inversion (e.g. *What can he do? Where does that go?*). As a result, all (non-subject) Wh- questions include both a finite auxiliary and a non-finite main verb, with *Do* filling the auxiliary slot in cases where there is no auxiliary in the corresponding declarative utterance (e.g. *It goes there*, *Where does it go?*). In German and Spanish, on the other hand, Wh- questions can involve either subject/auxiliary inversion (e.g. *Wo kann er spielen? ¿Dónde puede jugar?* (Where can (he) play-INF?)) or subject/main-verb inversion (e.g. Wo spielt er? *¿Dónde juega?* (Where play-FIN (he)?)). Thus, only a subset of Wh- questions includes both a finite auxiliary and a non-finite main verb, with the remainder including only a finite main verb. Since MOSAIC learns OI errors in Wh-questions by learning Wh- questions with missing auxiliaries from the input, one possible

explanation for the much lower rate of OI errors in German and Spanish is that the rate of compound-finite questions is also much lower in German and Spanish than in English.

This possibility was investigated by searching Leo's and Juan's corpora for Wh-questions and distinguishing between simple finites (i.e. those that included only a finite verb form) and compound finites (i.e. those that included both a finite and a non-finite verb form). This analysis revealed that the proportion of compound-finite Wh- questions in Leo's input was .31 and the proportion of compound-finite Wh- questions in Juan's input was .28. These proportions suggest that it is possible to explain the lower rate of OI errors in Wh- questions in German and Spanish as opposed to English in terms of the lower rate of compound-finite Wh- questions in the input (approximately 30% in German and Spanish as opposed to 100% in English). However, they also raise the question of why the rate of OI errors is so much lower in German Wh- questions than it is in German declaratives (where compound finites also occur at a rate of approximately 30%).

The reason why MOSAIC simulates this difference is that it is much easier for the model to learn OI errors in German declaratives than it is to learn OI errors in German Wh- questions. Thus, because OI errors in declaratives do not have to include utterance-initial material, they can be represented in the model either as utterance-internal omission errors (e.g. *Er (kann) es machen* (He (can-FIN) it do-INF) or as utterance-final truncation errors (e.g. *(Er kann) es machen*). Since the latter type of error can be learned without associating utterance-initial and utterance-final elements, OI errors occur frequently in the model's early declarative output, and more frequently than simple-finite verb forms, which are less likely to occur in utterance-final position in the input, and hence less likely to be learned in the early stages. The rate at which OI errors in declaratives occur in the

early stages therefore tends to be considerably higher than the rate at which compound finites occur in the input (i.e. 60% rather than 30%). OI errors in Wh- questions, on the other hand, can only be represented as utterance-internal omission errors and hence can only be learned by associating utterance-initial and utterance-final elements. Since the model's ability to associate utterance-initial and utterance-final elements is dependent on the distance between these elements in the target utterance, and since the distance between the utterance-initial Wh- word and the utterance-final non-finite verb in compound questions is relatively large, OI errors in Wh- questions do not tend to occur until the model has learned some relatively long utterance-final phrases. By this time the model also tends to have learned a relatively large number of correct finite questions from the right edge of the utterance. The rate at which OI errors in Wh- questions occur in the early stages therefore tends to be lower than the rate at which compound-finite Wh-questions occur (i.e. somewhere between 0 and 30%).

Interestingly, this explanation of the low rate of OI errors in Wh- questions in German can also be extended to account for the lower rate of OI errors in Wh- questions in Spanish. Thus, in Spanish, the rate of OI errors in Wh- questions is low for the same reason that it is low in German (i.e. because it is easier to learn correct simple-finite questions than it is to learn OI errors from compound-finite questions). However, the rate of OI errors in Wh- questions is even lower in Spanish than it is in German because Spanish simple-finite questions are shorter and therefore easier for MOSAIC to learn than German simple-finite questions. This latter difference reflects the fact that German is an obligatory subject language in which simple-finite Wh- questions always include a subject NP, and are therefore at least three words long (e.g. *Was willst du?* (What want-

FIN you?)). Spanish, on the other hand, is a pro-drop language in which simple-finite Wh- questions do not necessarily include a subject NP, and can therefore be only two words long (e.g. *¿Qué quieres?* (What want-FIN (you)?)). Since, in order to qualify as a Wh- question, an interrogative utterance must include an (utterance-initial) Wh- word, and since Wh- words tend to occur closer to the right edge of the utterance in Spanish simple-finite questions than in German simple-finite questions, MOSAIC tends to learn simple-finite Wh- questions earlier in Spanish than it does in German.

5. Discussion

The aim of the present paper was to describe a new version of MOSAIC that learns from both the right and the left edge of the utterance, and to investigate the extent to which it was able to provide an integrated account of the cross-linguistic patterning of OI errors in declaratives and Wh- questions. The new version of MOSAIC learns structures with missing sentence-internal elements and is hence able to simulate a wider range of error types than previous versions of the model. It also differentiates between declaratives and questions in the input, learning declarative structures from declarative input and interrogative structures from interrogative input. It therefore allows us to investigate the extent to which it is possible to simulate differences in the cross-linguistic patterning of OI errors in declaratives and Wh- questions in terms of the interaction between the same learning mechanism and cross-linguistic differences in the distributional patterning of declarative and interrogative input.

In a first set of simulations, it was shown that, despite the addition of left edge learning and the distinction drawn between declaratives and questions in the input, the model still provides a good fit to the data on OI errors in Dutch, Spanish and German

declaratives. More specifically, the model is able to simulate both the large, apparently qualitative, difference in the rate of OI errors in Dutch and Spanish during the early stages and the more subtle quantitative difference in the rate of OI errors between Dutch and German. Previous work with MOSAIC has shown that the model simulates these effects in terms of the interaction between right edge learning and differences in the proportion of utterance-final verb forms that are non-finite in the three languages. The present simulations suggest that adding left edge learning and restricting the model to learning declaratives from declarative input has little effect on this interaction. The implication is that the effects of right edge learning in MOSAIC are relatively robust, and hence unlikely to be affected by minor changes to the underlying mechanism.

In a second set of simulations, it was shown that, in addition to simulating the data on Dutch, Spanish and German, the new version of MOSAIC also provides a good fit to the developmental data on OI errors with third person singular subjects in English. These simulations constitute a stronger test of the new mechanisms within the model since, without these mechanisms, it would not have been possible to simulate the data on the basis of declarative input alone. They thus show that it is possible to replace MOSAIC's old mechanism for simulating OI errors with subjects with a new mechanism that concatenates utterance-initial and utterance-final phrases. This mechanism is considerably more plausible than the old mechanism since it no longer relies on the unrealistic assumption that OI errors with subjects are learned from questions in the input. However, it is worth emphasising that replacing the old mechanism in this way does not result in any substantive changes to the underlying theory of OI errors implemented in MOSAIC. Thus, MOSAIC still simulates OI errors as compound finite

utterances with missing finite modal/auxiliary verbs, and MOSAIC still simulates cross-linguistic variation in OI errors in terms of differences in the way in which compound finite verb forms pattern in different languages.

In a final set of simulations, it was shown that, in addition to simulating the data on OI errors in declaratives, the new version of MOSAIC also provides a reasonable fit to the developmental data on OI errors in Wh- questions. These simulations show that it is possible to simulate OI errors in Wh- questions as utterance-internal omission errors and hence to extend MOSAIC's account of the cross-linguistic patterning of OI errors in declaratives to explain the cross-linguistic patterning of OI errors in Wh- questions. More specifically, they show that the higher proportion of OI errors in English than in German or Spanish Wh- questions can be explained in terms of the absence of simple-finite Wh-questions in English, which means that all English Wh- questions are potential sources of OI errors. This explanation of the cross-linguistic patterning of OI errors in Wh-questions can be contrasted with that of Wexler (1998), which assumes highly abstract knowledge of the structural properties of the input language on the part of the child. Thus, according to Wexler (1998), OI errors are predicted in Wh- questions in OI languages like English, but not in non-OI languages like Spanish. However, they do not occur in Wh- questions in German because, although German is an OI language, German is also a V2 language, which requires a finite verb in second position when first position is filled by a Wh- word. The implication is that the difference between English and German reflects the fact that English is an SVO language whereas German is an SOV-V2 language. According to MOSAIC, on the other hand, OI errors occur in Wh- questions in English because English does not allow subject/main-verb inversion, but are rare in

Spanish and German because the majority of Wh- questions in these languages are simple-finite questions formed by subject/main-verb inversion. The implication is that the difference between English and German reflects the fact that German allows subject/main-verb inversion, whereas English does not. One way of distinguishing between these two accounts is to look at the rate at which OI errors occur in Wh-questions in an OI language like French, which is an SVO language (like English), but also a language that allows subject/main-verb inversion (like German). Interestingly, OI errors appear to be rare in Wh- questions in French (Crisma, 1992), which is consistent with the view, implemented in MOSAIC, that the absence of subject/main-verb inversion in English is the critical factor.

When taken as a whole, the results of the present study show that MOSAIC is able to provide a unified account of the cross-linguistic patterning of OI errors in declaratives and Wh- questions. However, these results also have a number of more general implications for the field as a whole. First, they show that imposing psychologically realistic constraints on a distributional learning mechanism such as MOSAIC can be a surprisingly powerful means of simulating patterns of error in young children's speech. Thus, although OI errors have typically been interpreted in the generativist literature as evidence for highly structured linguistic knowledge on the part of the child, the combination of right- and left-edge learning implemented in MOSAIC allows the model not only to simulate OI errors in both declaratives and Wh- questions as utterance-internal omission errors, but also to simulate developmental changes in the rate at which these errors occur. Moreover, in each case, this is done using exactly the same version of the model. The only difference between the simulations is that OI errors in declaratives

are learned from declarative input and OI errors in Wh- questions are learned from interrogative input. The implication is that previous research may have underestimated the extent to which it is possible to explain the often rather complex pattern of errors in children's speech in terms of processing limitations in learning.

Second, these results illustrate how psychologically motivated constraints on learning can interact with the distributional properties of the input language to result in striking (and in some cases unanticipated) cross-linguistic effects. Thus, right edge learning in MOSAIC interacts with the proportion of utterance-final verbs that are non-finite to result in a large difference in the rate of OI errors in Dutch and Spanish declaratives and a small difference in the rate of OI errors in Dutch and German declaratives. On the other hand, because of its sensitivity to the distance between elements, MOSAIC's mechanism for associating utterance-initial words and utterance-final phrases interacts with the proportion of Wh- questions that contain simple finites, and the length of these simple-finite questions, to result in a large difference in the rate of OI errors in English and German Wh- questions and a smaller difference in the rate of OI errors in German and Spanish Wh- questions. It is worth emphasising that neither of these patterns of effect would have been anticipated on the basis of a straightforward analysis of the input languages since the rate of compound finites in German declaratives is considerably lower than the rate of OI errors in German declaratives during the early stages, and the rate of compound finites in German Wh- questions is considerably higher than the rate of OI errors in German Wh- questions, and roughly equivalent to the rate of compound finites in Spanish Wh- questions. In each case, the pattern of effects can only be properly

understood in terms of the interaction between the processing constraints built into MOSAIC's learning mechanism and the distributional properties of the input language.

Finally, our results illustrate the value of an approach that uses computational modelling techniques to understand the relation between the distributional properties of the language to which children are exposed and the speech that children actually produce. This approach has allowed us to use the constraints inherent in developmental and cross-linguistic data to build an integrated account of the pattern of errors in children's speech, and to show that treating OI errors as compound finites with missing modals and auxiliaries can explain several important features of these data. These include the apparently qualitative difference in the rate of OI errors in OI and non-OI languages (Freudenthal et al. 2007); more subtle quantitative differences in the rate at which OI errors occur in different OI languages (Freudenthal et al., 2007; Freudenthal et al., 2010); the fact that OI errors in Dutch and German tend to have a modal reading (Freudenthal et al., 2009); and the fact that the rate at which OI errors occur with particular verbs is positively related to the rate at which those verbs occur in compound structures in the input (Freudenthal et al., 2010). The results of the present study show that this kind of account is also able to provide a unified explanation of the cross-linguistic patterning of OI errors in declaratives and Wh- questions. The present study therefore provides further support for the view that, contrary to the claims of many generativist researchers, OI errors are not non-finite verb forms used in place of simple-finite verb forms, but compound-finite verb forms with missing modals/auxiliaries.

Acknowledgements

Julian Pine is a Co-Director and Fernand Gobet is a Professor in the International Centre for Language and Communicative Development (LuCiD) at the University of Liverpool, for which the support of the Economic and Social Research Council (ES/L008955/1) is gratefully acknowledged. This research was funded by the Economic and Social Research Council under grant number RES000230211.

References

Aguado-Orea, J. & Pine, J. M. (2015). Comparing different accounts of the development of verb inflection in early child Spanish. *PloS One*, *10(3)*, e0119613.

Behrens, H. (2006). The input-output relationship in first language acquisition. *Language and Cognitive Processes*, *21*, 2-24.

Bellezza, F. S., Andrasik, F. & Lewis, R. P. (1982). The primacy effect and automatic processing in free recall. *Journal of General Psychology*, *106*, 175-189.

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: Centre for Language and Cognition.

Boser, C., Lust, B., Santelmann, L., & Whitman, J. (1992). The syntax of CP and V2 in early child German (ECG): the strong continuity hypothesis. *NELS 22: Proceedings of the Twenty-Second Annual Meeting of the North East Linguistic Society* (pp. 51-66). Amherst, MA: GLSA.

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

Chase, W. G. & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81.

Crisma, P. (1992). On the acquisition of *Wh*- questions in French. *Geneva Generative Papers*, *2*, 115–122.

Frank, S., Goldwater, S. & Keller, F. (2013). Adding sentence types to a model of syntactic category acquisition. *Topics in Cognitive Science*, *5*, 495-521.

Freudenthal, D., Pine, J. M., Aguado-Orea, J. & Gobet, F. (2007). Modelling the developmental patterning of finiteness marking in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science, 31*, 311-341.

Freudenthal, D., Pine, J. M. & Gobet, F. (2005). On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research, 6*, 17-25.

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

Freudenthal, D., Pine, J. M. & Gobet, F. (2006). Unifying cross-linguistic and within-language patterns of finiteness marking in MOSAIC. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 232-236). Mahwah, NJ: Erlbaum.

Freudenthal, D., Pine, J. M. & Gobet, F. (2007). Understanding the developmental dynamics of subject omission: The role of processing limitations in learning. *Journal of Child Language, 34*, 83-110.

Freudenthal, D., Pine, J. M. & Gobet, F. (2009). Simulating the referential properties of Dutch, German and English root infinitives in MOSAIC. *Language Learning and Development, 5*, 1-29.

Freudenthal, D., Pine, J. M. & Gobet, F. (2010). Explaining quantitative variation in the rate of Optional Infinitive errors across languages: A comparison of MOSAIC and the Variational Learning Model. *Journal of Child Language, 37*, 643-669.

Frota, S., Butler, J. & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question. *Infancy, 19*, 194-213.

Geffen, S. & Mintz, T. H. (2012). Seven-month-olds' discrimination of statements and questions. In A. Biller, E. Chung & A. Kimball (Eds.), *Supplement to Proceedings of the 36th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press.

Geffen, S. & Mintz, T. H. (2015). Can you believe it? 12-month-olds use word order to distinguish between declaratives and polar interrogatives. *Language Learning and Development, 11*, 270-284.

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

Gobet, F., & Simon, H. A. (1998). Expert chess memory: Revisiting the chunking hypothesis. *Memory, 6*, 225-255.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory, 13*, 318-324.

Gupta, P., Lipinski, J., Abbs, B. & Lin, P. H. (2005) Serial position effects in nonword repetition. *Journal of Memory and Language, 53*, 141-162.

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

Homae, F., Watanabe, H., Nakano, T., Asakawa, K. & Taga, G. (2006) The right hemisphere of sleeping infant perceives sentential prosody. *Neuroscience Research*, *54*, 276-280.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S. & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 1217-1232.

Hyams, N. (1996). The underspecification of functional categories in early grammar. In H. Clahsen (Ed.), *Generative perspectives in language acquisition* (pp. 1-42). Philadelphia: John Benjamins.

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

Jahnke, J. C. (1965). Primacy and recency effects in serial-position curves of immediate recall. *Journal of Experimental Psychology*, *70*, 130-132.

Josefsson, G. (2002). The use and function of nonfinite root clauses in Swedish child language. *Language Acquisition*, *10*, 273-320.

Lieven, E. V. M., Pine, J. M. & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, *24*(1), 187-219.

López Ornat, S. (1994) *La adquisición de la lengua Española*. Madrid: Siglo XXI.

MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk (3rd Edition).* Mahwah, NJ: Erlbaum.

MacWhinney, B. (2004). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, *31*(4), 883-914.

Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child-directed speech. *Cognition, 90,* 91-117.

Murdock, B. B. (1962). The serial position effect in free recall. *Journal of Experimental Psychology, 64,* 482-488.

Neath, I. (1993). Contextual and distinctive processes and the serial position functions. *Journal of Memory and Language, 32,* 820-840.

Phillips, C. (1995). Syntax at age two: Cross-linguistic differences. In C. Schütze, J. Ganger & K. Broihier (eds), *Papers on Language Processing and Acquisition*. MIT Working Papers in Linguistics, 26, 225-282.

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.

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

Rizzi, L. (1994). Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition,* 3, 371-393.

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

Seidl, A., Hollich, G. & Jusczyk, P. W. (2003). Early understanding of subject and object Wh-questions. *Infancy, 4*(3), 423-436.

Sikström, S. (2006). The isolation, primacy and recency effects predicted by an adaptive LTD/LTP threshold in postsynaptic cells. *Cognitive Science, 30,* 1-33.

Soderstrom, M., Ko, E-S. & Nevzorova, U. It's a question? Infants attend differently to yes/no questions and declaratives. *Infant Behavior and Development, 34,* 107-110.

Szagun, G. (2001). Learning different regularities: The acquisition of noun plurals by German-speaking children. *First Language, 21,* 109-141.

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). The item-based nature of children's early syntactic development. *Trends in Cognitive Sciences, 4,* 156-163.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.

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

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F., & Rook, R. G. (1985). Memory processing of serial lists by pigeons, monkeys and people. *Science,* 229, 287-289.

Appendix A: Node Creation Probability (NCP)

Since developmental functions tend to be non-linear, the formula for Node Creation Probability is designed to result in a basic sigmoid curve when NCP is plotted against the number of utterances to which the model has been exposed. One important feature of the formula is that it includes the term $(m\text{-}u/c)/t$ where m and t are constants, c is corpus size and u is the total number of utterances seen by the model so far. This term has the effect of ensuring that NCP is equal for corpora of different sizes, given an equal number of exposures to the input set. The rationale for controlling for differences in corpus size in this way is that the major determinant of such differences within the CHILDES database is variation in sampling regime (i.e. how often and for how long recordings were made) rather than, for example, variation in the number of utterances to which children are exposed per hour of recording time. Setting m to 40 has the effect of ensuring that the NCP for an utterance-final word is 0.5 after the input corpus has been processed 40 times. Reducing or increasing the value of m within a range of 20 to 60 does not affect the shape of the sigmoid curve (see Fig. 9), and hence does not affect the characteristics of the network being built. It simply reduces or increases the number of times the model needs to be exposed to the input corpus for the NCP to reach a particular value. Dividing the term (m-u/c) by a constant t has the effect of reducing the rate at which the NCP changes as a function of the amount of input seen, with each increase in t resulting in a decrease in the rate at which the NCP changes (see Fig. 10). Setting the constant t to 3 is therefore simply a convenient way of slowing down learning such that the size of the network, and hence the MLU of the model's output, does not increase so quickly with each exposure to

the corpus that it would be necessary to stop the model in mid-run in order to match its MLU to the MLU points of interest in the child data.

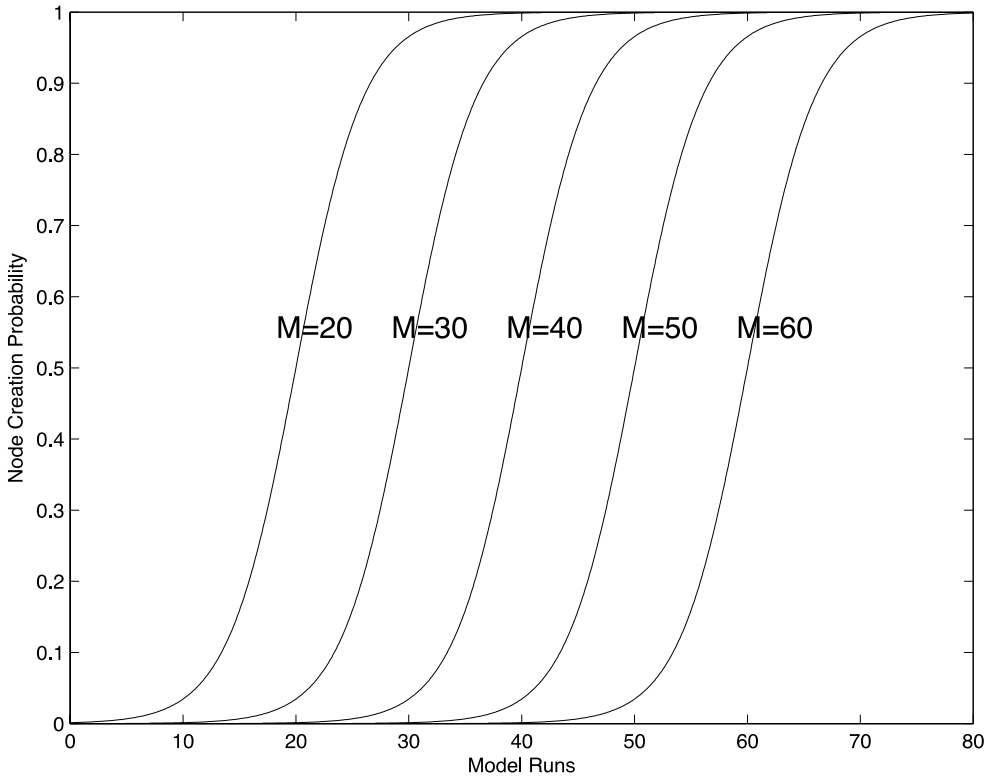


Fig. 9: Differences in the pattern of rate of change in NCP as a function of changes in *m*

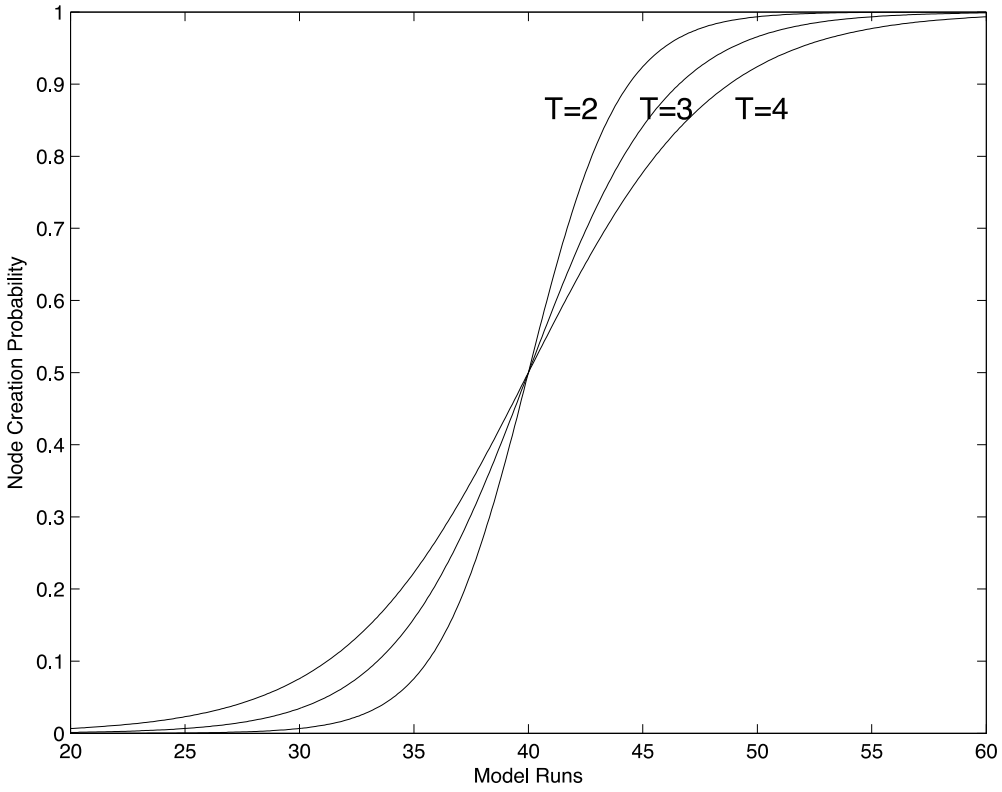


Fig. 10: Differences in the pattern of rate of change in NCP as a function of changes in *t*

A second important feature of the formula is that the base number is raised to $\sqrt{d}$, where *d* is the distance to the edge of the utterance (or length of the phrase being encoded). This feature of the formula has the effect of decreasing the probability of learning words or sequences as a function of their distance from the edge of the utterance. Raising the base number to $\sqrt{d}$ rather than *d* simply prevents the *d* parameter having so great an effect on NCP that the model is never able to fully learn longer utterances. The term *d* is also used to implement a bias for right edge learning over left edge learning. When considering a left edge phrase for encoding, the term *d* is incremented by 2 before computing the NCP. This results in left edge learning being slower than right edge learning. Unlike the term *(m-u/c)/t*, the terms $\sqrt{d}$ and *d+2* do play an important role in

determining the shape of the developing network. However, this is because these terms directly implement the theoretical claims instantiated in MOSAIC that children are subject to a strong recency effect and a weaker primacy effect in learning. It is thus the inclusion of these terms in the NCP formula that allows us to test the idea that it is possible to simulate the cross-linguistic patterning of OIs in declaratives and Wh-questions in terms of the interaction between edge-first learning and differences in the distributional properties of the input language.

Appendix B: Sample output from English models

Table 6 provides randomly-generated 50-utterance samples of declarative multi-word output for each of two English models at an MLU of approximately 2.5. Table 7 provides corresponding randomly-generated 50-utterance samples of Wh- question output. It can be seen from Table 6 that, although MOSAIC produces some declarative utterances (marked in bold) that are clearly anomalous (approximately 10%), the majority of utterances in the declarative samples consist either of grammatically acceptable strings (e.g., *I won't fit*, *He tastes beautiful*, *It doesn't go here*, *He's hiding*) or of strings that include plausible omission errors (e.g., *Open cupboard*, *That wake you up*, *Want pizza*, *He sit on here*). This also appears to be true of the Wh- question samples presented in Table 7, which include a small number of anomalous utterances (approximately 10% in each case), with the majority being made up of correct questions (e.g., *What do you want?*, *How does it go?*, *Where're you going?*, *Who do you mean?*) and questions from which either the auxiliary or both the subject and the auxiliary have been omitted (e.g., *What I eat?*, *Where you going?*, *What look like?*, *What found?*).

Table 6. Example declarative output

Anne's Model		
Push it up	Just having a sleep	Turn door
Go on the road	Think she is	Jump still
Like cake	You can eat it	You stop that
Better stop then	He wants more fence	You get it then

Play with me	Go on then	**Come not fishes**
So to see you	**Eat out of your mouth please**	He tastes beautiful
Slipped over	Have her hair	It's not to read
Fit in that box	Opening the settee	Do them that way
That wake you up	You do them	I won't fit
It fits	If you hide	Hide them
Open cupboard	Anne's broke it	Push them that table
Looks pretty	Forgot your hair	Told him
And see something	We could do that	I choose
Let's see space off	Do them	Mended it
In he fits	Can't have some toast	Do it in here
Can't get out	**Eat crying**	Gone too
Shut the lid	Get Mummy's back	

Table 6: Example declarative output (continued)

Becky's Model		
In go	Try hard	Having ice cream
Then eat	They don't do it	Throw one
I did take it	**Take say rock a bye baby**	Like there
Like ice cream	You pull that	Go in

Have a key	Might like it	Wonder why
Take of it	Took them off	He's hiding
It on want it	Don't want black	Get other red one
It doesn't go here	You can't find	Take the bag
They want that	Becky wants upside down	Turn it
Doesn't go in there	He sit on here	I don't know it
Disturbed him now	Pull it away	They leave that in
Gone again	**I want you want that**	Yes you know
Do on the people	I take it	Throw up then
You know	That doesn't fit in	You do take it
Go now then	Know you need that	Done that one
Want pizza	I don't want it	Eat here
To put that	Push that on here	

Table 7: Example Wh- question output

Anne's Model		
What hold?	What do you want then?	What we do now then?
What won't he want?	What're you doing hair?	What's he found?
What you have to go?	What's that monkey doing?	How does it go?
What say her name was?	What you eat that for?	What found there?

What eat now?	What's she find?	What I eat then?
What can you see?	Who's he got to do?	What you find?
How do you want?	What's he found to go?	Where are you going?
What found?	What're you gonna have?	Who's he find to go?
Where need me?	What we play with today?	**What she have to go?**
Where taking the shop?	What're you doing now?	Where you read them?
What else can we see?	Where's the horse gone?	Who you see?
What I eat?	**What're you eat about that?**	**What're you tell me?**
What we doing there?	Who's he gonna do?	What eat then?
What I find?	What's that baby doing?	What didn't we see?
What didn't they called?	What're you doing there?	What won't he need?
What get?	**Where didn't we going?**	What she eat?
What go with then?	What do about that then?	

Table 7: Example Wh- question output (continued)

Becky's Model		
What else do you know?	**What did you say it colour's?**	What you remember?
Where you going?	Who do you think that is?	What do you eat?
What do you like?	Where goes?	Where're you going?
What we looking for?	What do you think?	Where you want it?

When we going?	What look like?	What I done here?
What do you are?	What have we done here?	Where are you going?
Who do you mean?	What have you got there?	Where that goes?
What we done here?	**What do you need to Sukie?**	What see?
What is he doing?	Where that one go?	**What buy have?**
Where that go?	What did you say that about?	Where you want that**?**
Where does it hurt?	Where go?	Where's it gone?
Where's he going?	Who you see?	What we done?
What can you hear?	Who do you think it is?	What we buy?
What I done?	What did you say it about?	What were you saying?
What are I missing?	What have you bought there?	When going?
What make?	What's this doing in here?	What we do?
Where's want that?	What you done then?	