

Simulating the Noun-Verb Asymmetry in the Productivity of Children's Speech

Daniel Freudenthal (D.Freudenthal@Liv.Ac.Uk)

Julian M. Pine (Julian.Pine@Liv.Ac.Uk)

School of Psychology, University of Liverpool

Fernand Gobet (Fernand.Gobet@Brunel.Ac.Uk)

School of Social Sciences, Brunel University

Abstract

Several authors propose that children may acquire syntactic categories on the basis of co-occurrence statistics of words in the input. This paper assesses the relative merits of two such accounts by assessing the type and amount of productive language that results from computing co-occurrence statistics over conjoint and independent preceding and following contexts. This is achieved through the implementation of these methods in MOSAIC, a computational model of syntax acquisition that produces utterances that can be directly compared to child speech, and has a developmental component (i.e. produces increasingly long utterances). It is shown that the computation of co-occurrence statistics over conjoint contexts or frames results in a pattern of productive speech that more closely resembles that displayed by language learning children. The simulation of the developmental patterning of children's productive speech furthermore suggests two refinements to this basic mechanism: inclusion of utterance boundaries, and the weighting of frames for their lexical content.

Introduction

Children acquiring their native language are faced with a task of considerable complexity. They need to acquire a system described by syntactic rules as well as the syntactic categories over which these rules are defined. This problem has been referred to as the 'bootstrapping problem'. Several solutions to the bootstrapping problem have been suggested. The *distributional* approach makes use of the fact that words that belong to the same word class tend to be preceded and followed by similar words. Thus, nouns tend to be preceded by determiners and adjectives, and followed by verbs. Similarly, verbs are preceded by (pro)nouns and followed by determiners and (pro)nouns. A system that tracks the overlap in the lexical items that precede and follow individual words may therefore be able to cluster these words into syntactic classes. These word classes could then potentially be used to infer phrasal categories such as noun phrase and verb phrase (see e.g. Finch & Chater, 1994).

The success of this second stage of the learning mechanism depends crucially on the quality of the syntactic classes that were derived in the first stage. For this reason, several researchers have explored different mechanisms for computing co-occurrence statistics and the effects these have on the quality of the derived classes. Finch & Chater (1994) analysed a 40,000,000 word corpus of USENET newsgroup data and used the rank order correlation between the (independent) sets of two word phrases that preceded

and followed target words to inform a hierarchical cluster analysis that derives word classes. Redington, Chater & Finch (1998) perform a similar analysis on a corpus of several million words of child directed speech obtained from the CHILDES data base. They compared the performance over contexts of length one and two and found that the quality of results was very similar.

Mintz (2003) uses a slightly different approach. Mintz introduces the concept of a *frame*; two jointly occurring words with one word intervening. Computing co-occurrence statistics over conjoined pairs rather than independent sets of preceding and following words has the desirable property that it is more constraining and is therefore likely to lead to grammatical categories that are of higher quality. Mintz restricts his analysis to frames that have a high frequency in the input, and finds that the items that co-occur in these frames have a high likelihood of belonging to the same word class. Mintz does not perform a cluster analysis but does suggest that more comprehensive classes can be obtained using a relatively simple unification procedure based on overlap in the words contained in the classes.

The approach taken by Mintz, Redington et al. and Finch & Chater clearly shows there is a considerable amount of information in the distributional characteristics of the input that could potentially be used by a child acquiring language. Freudenthal et al. (2005a), however, argue that these approaches suffer from an inherent difficulty as they fail to consider how derived categories can actually be used in the production of (novel) utterances. Freudenthal et al. report work on MOSAIC, a computational model of language acquisition that produces actual utterances as output and implements a mechanism for the production of novel utterances that is very similar to that implemented by Redington et al. MOSAIC links together words that are followed and preceded by similar words and substitutes these words when producing output from the model. The fact that MOSAIC produces actual output results in potential classification errors quickly becoming apparent. Freudenthal et al. also argue that such classification errors may not always be apparent using the standard evaluation measures of accuracy and completeness as these depend on researcher's intuitions regarding an item's grammatical class. Inspection of the (representative) verb classes reported by Mintz (2003), for example, shows that these include past tense, present tense and progressive verb forms, imperative and non-imperative verb forms, and transitive

and intransitive verbs. Substituting such items in production will quickly lead to errors that may not be apparent when using a metric that simply classes all items as verbs.

A further advantage of producing actual utterances is that it allows for a comparison between the output of one’s language learning model and the characteristics of actual child speech. Asymmetries in children’s tendency to generalize across words from different syntactic categories can then be used to inform the implementation of mechanisms that compute co-occurrence statistics. One important asymmetry that has been identified in the recent developmental literature is that children tend to be more conservative in their substitution of verbs than in their substitution of nouns. For example, several experimental studies have shown that children will readily substitute novel nouns in familiar verbal contexts, but tend to restrict their use of novel verbs to contexts in which they have heard them used in the input (see e.g. Akhtar & Tomasello, 1997; Tomasello, 2000). Moreover, Fisher (2002) points out that this pattern of generalization is precisely what one would expect given the nature of the system that the child is acquiring since restrictions on the argument structures in which different verbs can occur mean that generalizations across verbs tend to be ‘riskier’ than generalizations across nouns.

When taken together, these considerations place strong constraints on the development of mechanisms for extracting syntactic categories since they suggest the need for a mechanism that generalizes more readily across nouns than across verbs. They also suggest that simulating this asymmetry will not only increase the child-likeness of the model’s output, but also reduce the probability of generating ungrammatical utterances.

Further constraints on the feasibility of mechanisms for the extraction of syntactic categories become apparent when considering the fact that such mechanisms are likely to be used by children acquiring a language. The most notable way in which child speech differs from adult speech is that it is considerably shorter. Children initially produce utterances that are only one or two words long. The mean length of their utterances (MLU) slowly increases as they grow older. This restriction on the length of children’s speech suggests that the length of the phrases that children represent is considerably shorter than the length of the phrases they hear. This considerably reduces the number and type of contexts that could potentially feed into a system that computes co-occurrence statistics. Failure to consider this developmental component may therefore lead to researchers considering mechanisms that utilize information that is not necessarily available to a child.

The aim of this paper is to assess the relative virtues of using conjoined or independent sets of preceding and following items for the extraction of syntactic categories. This is done in the context of MOSAIC, an implemented model of language acquisition that has a developmental component. Specific attention will be given to how well the different mechanisms approximate the noun-verb

asymmetry apparent in children’s productive speech, as well as the constraints that result from simulating children’s increasing utterance length.

The remainder of this paper is organized as follows. First, we briefly describe MOSAIC, the model that is used as a test bed for mechanisms for the extraction of syntactic categories. Next, MOSAIC’s current mechanism for linking distributionally similar items is described. Finally, a number of substitution mechanisms are implemented and the output evaluated in terms of quantity and plausibility.

MOSAIC

MOSAIC (Model of Syntax Acquisition in Children) is a computational model that has mostly been applied to the cross-linguistic simulation of the development of finiteness marking in children acquiring their native language (Freudenthal, Pine and Gobet 2006). MOSAIC learns off realistic, child directed input and learns to produce progressively longer utterances that can be directly compared to child speech. The basis of MOSAIC is an n-ary discrimination net that is used to incrementally store (fragments of) the utterances that MOSAIC has seen as input. Learning in MOSAIC is anchored at the beginning and end of the utterances to which it is exposed. That is, MOSAIC only learns a phrase when everything preceding or following that phrase in the utterance has already been encoded in the network. MOSAIC produces two types of utterances: utterance-final phrases and concatenations of utterance-initial and utterance-final phrases: utterances with missing sentence-internal elements. MOSAIC’s mechanism for producing incomplete phrases has been shown to provide a good fit to the Optional Infinitive phenomenon across four languages: English, Dutch, German and French (Freudenthal et al, 2005b). Fig. 1 shows a sample MOSAIC network.

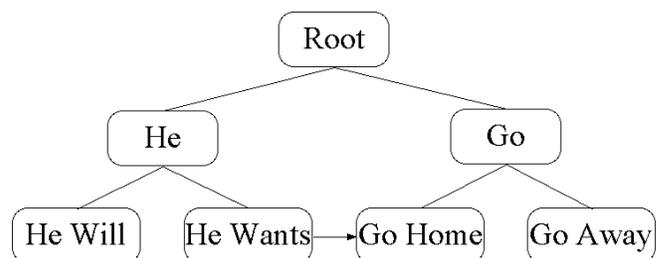


Figure 1: A partial MOSAIC network. The sentence-initial phrase *he wants*, and the sentence-final phrase *go home* have been associated, allowing the model to produce the utterance *He wants go home*. The model is also capable of producing the phrases *go home* and *go away*.

MOSAIC is capable of producing output with an increasing MLU. This is because learning is generally slow. Input is fed though MOSAIC and output is generated after every presentation of the input. The amount and length of phrases encoded in a MOSAIC network increases with every exposure to the input. Thus, developmental change can be

simulated by analyzing increasingly mature models and matching the MLU of the respective models with that of children at different developmental stages.

Productivity in MOSAIC

MOSAIC's mechanism for producing novel utterances is very similar to that described by Redington, Chater & Finch. For all nodes in the network, the preceding and following context (when encoded in the model) is stored. These contexts take the form of two independent lists of words that preceded and followed the target item. Thus, MOSAIC does not implement the notion of a frame, but assesses the preceding and following context independently. MOSAIC then considers the overlap between the contexts for pairs of words. If, for two words, the overlap in the words in both the preceding and following context exceeds a predetermined threshold, they are considered equivalent, and are connected through a *generative link*. Two words that are connected through a generative link can be substituted for each other in production. Thus, if the model has encoded the phrase *the red ball* and the words *red* and *blue* share sufficient overlap, the model is capable of producing the novel phrase *the blue ball*. MOSAIC is capable of substituting several words in an utterance. Thus, MOSAIC would be able to produce the phrase *a blue ball* if the words *a* and *the* also share a generative link.

The proportion of novel utterances in MOSAIC's output varies as a function of the mean length of the output. Typically, productivity increases from around 5% novel utterances at an MLU of 2 to around 50% productivity at an MLU of around 4. MOSAIC thus does not produce many novel utterances in the early stages of development. Productivity in MOSAIC also tends to revolve around verbs rather than nouns. Thus, MOSAIC tends to link together verbs more frequently than nouns. This runs counter to the notion that children are more conservative in substituting verbs than nouns. It also increases the risk of MOSAIC generating a high proportion of utterances that are ungrammatical, as verb-verb substitutions are more likely to be ungrammatical than noun-noun substitutions.

Implementing Frames in MOSAIC

Mintz performed an analysis of a number of corpora of child directed speech, and analyzed the contents of the 50 most frequent frames in each corpus. Items that co-occurred in one of these frames were considered equivalent. While this approach shows the potential value of using frames, it needs to be developed into a more dynamic and probabilistic mechanism in order to be suited for a model that encodes and produces progressively longer utterances.

We implemented the mechanism in a similar way to MOSAIC's current generativity mechanism: two items are considered equivalent if there is sufficient overlap in the frames they occur in. An example may serve to illustrate how the dynamics of the frames and independent contexts may differ. Suppose the model has encoded the phrases 'A man eats', and 'The man drinks'. The preceding

independent context for the word 'man' is now (A, The) while the following context is (eats, drinks). The frame context for 'man' is ((A - drinks), (The - eats)). If the model now encodes the phrases 'A woman eats' and 'The woman drinks', the independent preceding context would be (A, The), and the independent following context would be (eats, drinks) while the frame context would be ((A - eats), (The - drinks)). The overlap for the independent contexts would thus be 100%. The overlap in frames however, is 0%.

While, in the above example, the notion of a frame is clearly more constraining, it should be noted that the list of independent contexts may actually grow more quickly than the list of frames. This is because the model only needs to encode a two-word phrase to add an item to the independent preceding or following context. Thus, if the model encodes the (incomplete) phrase 'rich man', the word 'rich' is added to the independent preceding context, thus lowering the preceding overlap without affecting the context in terms of frames, as no following context has been encoded. Thus, the dynamics of the overlap between two items will be different for frames and independent contexts. As these dynamics are also affected by the frequency and variety of the contexts in which items occur, they may well affect verbs and nouns differently.

A final note concerns the status of sentence boundaries. Mintz defines a frame as two conjoint lexical items (words) with one word intervening. Thus, frames that contain sentence boundaries are excluded. While on the face of it frames including sentence boundaries are less restrictive than lexical frames it should be noted that they can actually be quite informative with respect to a word's grammatical class. Thus, the frames 'THE - END' and 'A - END' are very frequent and contain a large number of nouns. For the present simulations both lexical frames and frames with sentence boundaries were encoded in the model. Separate analyses were run utilizing all frames or lexical frames only to investigate the impact of including frames containing sentence boundaries.

The simulations

The simulations reported here were run using the corpora of child directed Speech for two English children (Anne and Becky). The corpora, which contain approximately 33,000 and 25,000 utterances respectively were fed through MOSAIC several times and output of increasing length was generated after every exposure to the input. The version of MOSAIC used for these simulation tracks both the frames and the independent preceding and following contexts for the words it encodes. Separate analyses were run using substitution of individual words on the basis of either the frames or the independent contexts.

Results

Simulations with MOSAIC's standard generativity mechanism were run first. Words were substituted when the overlap in terms of independent preceding and following contexts exceeded a threshold of 25%. Output was

generated at three different MLU points and analysed in terms of the percentage novel items and number of noun-noun and verb-verb substitutions.

As can be seen in Table 1 the model only starts to produce substantial amounts of novel utterances in the later stages of development. It is also apparent that, except during the earliest stage, the model is more productive around verbs than around nouns.

Table 1: Descriptive statistics for MOSAIC’s output using substitution on the basis of independent contexts.

Child	MLU	Proportion novel	Noun-sub	Verb-sub
Anne	2.08	.05	83	11
	3.16	.27	1873	2755
	4.71	.50	12026	18897
Becky	1.93	.02	20	9
	2.95	.22	1144	1223
	4.25	.49	4558	13984

Next, the same analysis (at 25% overlap) was performed using the frame-based generativity mechanism. In this first analysis of frame-based substitution only lexical frames were used (thus, frames containing sentence boundaries were ignored). Table 2 gives the results of this analysis.

Table 2: Descriptive statistics for MOSAIC’s output using frame-based substitution.

Child	MLU	Proportion novel	Noun-sub	Verb-sub
Anne	2.08	.05	15	0
	3.11	.29	6547	130
	4.05	.29	13111	740
Becky	1.93	.01	24	0
	2.86	.18	1102	324
	3.58	.19	3273	307

As can be seen in table 2 the restriction to frames results in MOSAIC showing a clear bias towards noun-noun substitutions. It is also apparent that, for the last developmental stage, the models are considerably less generative. In order to investigate if this decreased generativity accounts for the noun bias, we also ran the frame-based simulations with an overlap percentage of 20. This increased the generativity in the final developmental stage to .47 for Anne’s model and .34 for Becky’s model. Noun substitutions still outnumbered verb substitutions by about 15 to 1 for Anne, and 5 to 1 for Becky’s model. Thus, the frame-based generativity mechanism is genuinely more productive around nouns.

Inspection of the number and types of links that are created by the models provides some insight into why the frame based generativity mechanism is more productive around nouns than verbs. Anne’s model in the second developmental stage has encoded 649 verbs and 733 nouns. The frame based mechanism has created 308 noun-noun links and 18 verb-verb links. The mechanism that links

items on the basis of independent preceding and following context creates 124 noun-noun links and 80 verb-verb links. The frame-based generativity mechanism thus creates more noun-noun links, and fewer verb-verb links. Inspection of the verbs that get linked also reveals that the frame based mechanism does not simply link fewer verbs: it links different verbs, in particular verbs with a lower average frequency. The average frequency in the input corpus for verbs linked on the basis of frames is 8.89. For the independent contexts mechanism the equivalent number is 35.54. The frame-based mechanism thus appears more likely to link low-frequency items than the independent contexts mechanism. A bias towards linking low-frequency items will naturally favour the linking of nouns over verbs, as nouns are, on average, less frequent than verbs.

Some of the reasons why a frame-based mechanism is biased towards linking low-frequency items become apparent when inspecting the frames and independent contexts that the model has encoded for particular words. For the verb ‘put’ the model has encoded 23 preceding and 24 following contexts. These independent contexts combine to give a total of 57 frames. For the verb ‘see’ 10 preceding and 20 following contexts and 41 frames have been encoded. The overlap in independent contexts is 33%, yet the overlap in terms of frames is a mere 2%. For the nouns ‘table’ and ‘door’, 4 and 8 preceding and 6 and 4 following contexts have been encoded, which give rise to 4 and 5 frames respectively. The overlap in these frames is 29%, the overlap in independent contexts is 17%. It thus appears to be the case that, for frequent items, many (varied) contexts are encoded, which can potentially combine to give many different frames. Infrequent items occur in a small number of potentially more typical contexts, which do not combine to give a large number of frames. This results in the overlap in terms of frames being higher than for independent contexts for infrequent items. For frequent items, the overlap in terms of frames tends to be lower than for independent contexts.

In order to establish if this pattern holds more generally, we divided the words encoded in MOSAIC into a low, medium and high frequency group¹, and calculated the average number of preceding and following contexts as well as frames. As can be seen in Table 3, the ratio of the number of frames over the (average of) the independent contexts increases linearly with frequency: approximately 1, 1.5 and 2. While this increase is not surprising in itself (as the maximum number of frames for any word is the product of the number of preceding and following contexts), it does explain why a frame-based mechanism favours the linking of infrequent items. While, intuitively, frames are more constraining than independent contexts, they appear to be especially constraining for frequent items that appear in

¹ Frequency was measured as the number of times the node encoding a word was traversed when processing the input. Low, medium and high frequency words were defined as having a frequency count between 20 and 500, between 500 and 1000, and over 1000 respectively.

many contexts. While the high-low frequency distinction does not cut across verbs and nouns, verbs (in particular, regular present tense verbs) are, on average, more frequent, and occur in more frames. In fact, the average frequency of the main verbs² encoded in MOSAIC is twice that of the nouns encoded in MOSAIC. Verbs occur, on average, in 5.3 frames, compared to 3.69 for nouns. The frames in which verbs occur appear to be more varied as well. The total number of verb frames encoded in Anne’s model is 1,966, which comprises 1,229 unique frames. The total number of noun frames is 2,679, which comprises 864 unique frames. Thus, on average, every unique verb frame occurs 1.6 times. Every unique noun frame occurs 3.1 times. The decreased generativity around verbs for a frame-based substitution mechanism therefore appears to be (at least partially) caused by verbs occurring in more and more varied frames compared to independent contexts.

Table 3: Average number of preceding, following contexts and frames for words of low, medium and high frequency.

Freq.	Number of words	Preceding contexts	Following contexts	frames
Low	2299	2.32	2.09	2.35
Medium	217	8.10	6.35	10.1
High	374	22.80	28.67	50.06

Introducing utterance-boundaries

While the frame-based generativity mechanism is successful in simulating the noun-verb asymmetry, it is also apparent that the model is still not very generative (particularly in the early stages). Several reasons can be put forward for why this is the case. First, the overlap threshold of 25% may be too high. Additional simulations with an overlap parameter of 10% showed that generativity is increased, but only for the later stages. Thus, even at 10% overlap Becky’s early model produces 1% novel utterances while Anne’s model produces 7% novel utterances. The reason why the early models remain less generative is that they actually encode relatively few frames. This is because a (lexical) frame is actually relatively long: 3 words. Particularly in the earlier developmental phases, MOSAIC encodes relatively short utterances. It is therefore unlikely that many phrases of three words are encoded. As a result, few frames are available. One possible way to increase the number of frames used for the decision to link two items is to include the frames that contain sentence boundaries. Frames that include sentence boundaries are quite frequent and are a potentially useful source of information.

The high frequency of frames that contain sentence boundaries is illustrated by an analysis of the frames encoded in Anne’s model in the earliest developmental phase. This model has encoded a total of 4,293 words. For these words a total number of 23,244 frames have been encoded. The number of frames that contain only lexical items (i.e. no sentence boundaries) is only 500. Thus, lexical

frames make up approximately 2% of the frames encoded in the model’s early stages. For Becky’s early model lexical frames make up approximately 3% of all the frames. Given the high frequency of frames that contain sentence boundaries, it appears unlikely that children would not be sensitive to such frames. Indeed, when analyzing the frames that occur in the child directed speech for Anne and Becky, it becomes apparent that the 50 most frequent frames all contain a sentence boundary. What’s more, the 2 most frequent frames (‘The - END’, ‘A - END’) are highly informative frames that each contain around 40 nouns.

Initial analyses with substitution based on all frames were aimed at establishing a suitable value for the overlap parameter. It became apparent that even at relatively high levels of overlap the model quickly became very generative and no longer showed a linear increase in the proportion of novel utterances. Instead, the model had relatively high levels of generativity at early stages of development. These values peaked at intermediate levels of development to subsequently decrease. This is not a characteristic of child speech, as children tend to become more productive with increasing MLU. It also became apparent that generativity around verbs became almost non-existent. Analysis of the developmental changes to the frames encoded in the model revealed that, over development, a larger proportion of the frames becomes lexical. Since lexical frames are more constraining than frames that contain a sentence boundary, it becomes less likely that two words have occurred in that particular frame. For instance, two nouns are more likely to share the frame ‘THE - END’ than the frame ‘THE - KICKS’. Thus, the overlap between two items tends to decrease as the number of lexical frames that these items have occurred in increases. This effect is more pronounced for verbs, as they tend to occur in more varied (lexical) contexts. The increase in lexical frames over the three developmental stages that were simulated is quite considerable. For both models the proportion of lexical frames is around 3% at the first MLU point, 15% at the second MLU point and 32% at the last MLU point.

In order to control for this increasing ‘informativeness’ of the frames (and obtain a more linear development of the model’s ability to generate novel utterances), it was decided to weight the lexical content of frames when calculating the overlap. In the previous simulations the overlap between two words was calculated as the number of overlapping frames divided by the union of the frames that either word has occurred in. For the weighted calculations, a lexical frame contributed 4 to the numerator, while a non-lexical frame contributed 1.

Table 4 gives the results of an analysis of MOSAIC’s output when including sentence-boundaries in frames. For these simulations the overlap threshold was set to 50% as the new definition of a frame is less restrictive than before. As can be seen in Table 4, the model now shows reasonable levels of generativity even at early stages of development. The model also shows a clear asymmetry between noun and verb substitutions. This asymmetry becomes less

² Excluding progressives and (regular) past tense.

pronounced during the later stages of development as the relative generativity around verbs increases.

Table 4: Descriptive statistics for MOSAIC’s output using frame-based substitution with utterance boundaries.

Child	MLU	Proportion novel	Noun-sub	Verb-sub
Anne	2.01	.41	2232	14
	3.38	.48	15993	762
	4.07	.58	36000	3807
Becky	2.11	.26	609	5
	3.26	.45	5082	678
	4.12	.55	12799	3435

Conclusions

This paper set out to establish the relative merits of using frames or independent contexts as the basis for a substitution mechanism in the simulation of child speech. Particular emphasis was placed on the model’s ability to simulate the verb-noun asymmetry that is apparent in child speech whilst incorporating the constraints that derive from simulating children’s increasing average utterance length. The analyses reported here show that a generativity mechanism that uses independent contexts is biased towards substituting high frequency items, while a mechanism based on frames favours the substitution of lower frequency items that feature in less varied contexts. Since nouns tend to fit the latter category, and verbs fit the former category the frame based mechanism provides a better fit to the noun-verb asymmetry. It was furthermore shown that, while a frame-based generativity mechanism provides a better fit, lexical frames do not occur in meaningful numbers in a model that has only encoded short utterances. This results in low levels of generativity in early stages of development.

The inclusion of utterance boundaries in frames drastically increases the number of frames that are available to the model whilst at the same time including some frames that are potentially very informative for the formation of a noun class. The inclusion of utterance boundaries therefore results in increased generativity around nouns, particularly during the early stages of development. This increased generativity comes at a price however, as it decreases the generativity around verbs when a simple overlap threshold is used. One possible solution to this is to weight the overlap for the lexical content of frames. This results in a model which shows relatively high levels of generativity throughout development without compromising generativity around verbs.

On a more general level, the analyses reported in this paper show that the simulation of child data through the production of actual utterances and the inclusion of a developmental component highlights the fact that relatively subtle differences in the implementation of a generativity mechanism can have rather profound effects on the type of generativity that a model displays. Thus, while intuitively frames are more constraining than independent contexts, the analyses reported here show that they are particularly

constraining for frequent items that appear in varied contexts. The use of frames therefore decreases generativity around verbs, while increasing generativity around nouns. Such effects may have quite profound implications for a model’s ability to simulate the child data. They are, however, likely to remain hidden in approaches that simply assess the quality of derived grammatical classes without producing actual utterances.

The inclusion of a developmental component furthermore highlights the fact that, while frame-based substitution does a better job of capturing the noun-verb asymmetry, there are only a small number of lexical frames available to a system that encodes and produces short utterances. Lexical frames may therefore be of limited utility to children in early stages of development. Such effects are likely to remain hidden in approaches that track statistics across all of the input, and may result in overestimating the importance of lexical frames at the expense of far more frequent frames that include utterance boundaries.

Acknowledgements

This research was funded by the Economic and Social Research Council under grant number RES000230211.

References

- Akhtar, N., & Tomasello, M. (1997). Young children’s productivity with word order and verb morphology. *Developmental Psychology*, 33, 952-965.
- Finch, S. & Chater, N. (1994). Distributional bootstrapping: From word class to proto-sentence. In A. Ram and K. Eiselt (Eds.), *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 301-306). Hillsdale, NJ: Erlbaum.
- Fisher, C. (2002). The role of abstract syntactic knowledge in language acquisition: a reply to Tomasello (2000). *Cognition*, 82, 259-278.
- Freudenthal, D., Pine, J.M. & Gobet, F. (2006). Modelling the development of children’s use of optional infinitives in English and Dutch using MOSAIC. *Cognitive Science*, 30, 277-310.
- Freudenthal, D., Pine, J.M. & Gobet, F. (2005a). 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. (2005b). Simulating the cross-linguistic development of Optional Infinitive errors in MOSAIC In B.G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Mahwah NJ: LEA.
- Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91-117.
- Redington, M., Chater, N. & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425-469.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, 209-253.