Simulating Developmental Changes in Noun Richness through Performance-limited Distributional Analysis

Daniel Freudenthal[1], Julian M. Pine[1], Gary Jones[2] and Fernand Gobet[1]
[1]Department of Psychological Sciences, University of Liverpool
[2]Division of Psychology, Nottingham Trent University

Abstract

In this paper we examine how a mechanism that learns word classes from distributional information can contribute to the simulation of child language. Using a novel measure of noun richness, it is shown that the ratio of nouns to verbs in young children's speech is considerably higher than in adult speech. Simulations with MOSAIC show that this effect can be partially (but not completely) explained by an utterance-final bias in learning. The remainder of the effect is explained by the early emergence of a productive noun category, which can be learned through distributional analysis.

Keywords: Language Acquisition; Learning biases; Productive noun use.

Introduction

A key question in the study of language acquisition is how children build word classes such as noun and verb. One source of information that children might employ in this process is the distributional properties of the language they are exposed to – the lexical environments in which words occur can act as a cue to a word's grammatical category. English words that are preceded by determiners such as *a* and *the* and followed by (auxiliary) verbs such as *can* and *will* tend to be nouns, while words that are preceded by (pro)nouns like *I* and *You* and followed by determiners such as *a* and *the* tend to be verbs.

Several approaches to this problem have been proposed, but they tend to focus on building large word classes with high accuracy rather than developing mechanisms that can be plausibly applied by, and fit developmental data from language-learning children. Thus, mechanisms for distributional analysis routinely collect data from large corpora and entire utterances, and make limited contact with the (developmental) child data. Recent work by Freudenthal et al. (this volume) has shown that it's possible to perform distributional analysis developmentally by embedding it in an existing model of language acquisition that learns to represent increasingly long utterances. In this paper we examine how the developmental emergence of the classes of verb and noun interacts with the model's learning biases to explain actual developmental child data – the decreasing ratio of nouns to verbs in child speech.

Two influential approaches to learning word classes from distributional information are those of Redington, Chater and Finch (1998), and Mintz (2003). Mintz (2003) introduces the notion of a *frequent frame* – a combination of two words with one word intervening (e.g. *You* X *A*). Mintz restricts his analysis to the 45 most frequent frames for a given corpus, and finds that the words that co-occur within a given frame tend to belong to the same category. The relative simplicity of Mintz's approach suggests it is well within the capabilities of language-learning children. However, it has been argued that it doesn't work very well in languages with relatively free word order, such as Dutch (Erkelens, 2008), or German (Stumper et al. 2011). It has also been argued that the approach is too sensitive to noise, and relatively poor at classifying nouns (Freudenthal et al. 2013).

In Redington et al.'s approach, the contexts for words that are to be classified (*target words*) are expressed as vectors containing counts for a number of high frequency *context words* in preceding and following position. Target and context words are typically restricted to the 1000 and 150 most frequent words for a given corpus. Similarity is expressed as the (rank-order) correlation between these vectors and the matrix of correlations is used as input for a cluster analysis. Redington et al.'s approach appears more robust. However, it has been criticized for having a high computational overhead (St. Clair et al. 2010). The fact that the approach requires the child to track frequencies of large numbers of words makes it less plausible as a mechanism that might be used by language-learning children.

While both approaches are intended as mechanisms that language-learning children use, neither approach incorporates a developmental component. That is, both approaches collect statistics from large corpora and complete utterances, and thus ignore the fact that children's early utterances are no more than one or two words long, and only gradually increase in length over a period spanning several years. While children may well attend to longer utterances than they produce, it seems unlikely that they process entire utterances from a young age. A mechanism that collects statistics across complete utterances may thus use information that is not available to language-learning children in the early stages of development.

The importance of development is further underscored by experimental evidence which suggests that children's productive use of words develops at different speeds for different word classes. In particular, data from production studies suggest that children develop a highly productive noun category before they develop a fully productive verb category (Akhtar & Tomasello, 1997; Olguin & Tomasello, 1999; Tomasello & Olguin, 1993).

Recent work by Freudenthal et al. (this volume) has shown that Redington et al.'s approach can be amended in a way that makes it far more plausible as a mechanism used by language-learning children. Specifically, Freudenthal et al. show that it is possible to add a developmental component to the mechanism by gradually expanding the contexts to which it is exposed in a developmentally plausible way. Freudenthal et al. did this in the context of MOSAIC (Freudenthal et al. 2007, 2010, 2015), an existing computational model that has been used to successfully simulate a number of key phenomena in language acquisition. MOSAIC is a simple learning model that takes as input orthographically transcribed child-directed speech. Training in MOSAIC takes place by feeding the input corpus through the model multiple times. A key feature of MOSAIC's learning mechanism is that it builds up its representation of the utterances to which it is exposed by starting at the right edge of the utterance and slowly working its way to the left. Thus, with each exposure to the input MOSAIC represents increasingly long utterance-final phrases that become increasingly adult-like. The utterance-final bias is MOSAIC's key mechanism for simulating cross-linguistic differences in children's early speech. MOSAIC's utterance-final bias is therefore independently motivated, and MOSAIC provides a natural setting for investigating developmental variants of mechanisms that build word classes on the basis of distributional information. A detailed description of MOSAIC and the way it is trained is provided in Freudenthal et al. (2015).

Freudenthal et al. (this volume) also show that it is possible to significantly simplify the mechanism of Redington et al. Redington et al. collect counts for a fixed number of context words in preceding and following position, and express similarity as the rank-order correlation between vectors containing these counts. Since the rank-order correlation is non-parametric in nature, the frequency information contained in the counts is largely ignored. Freudenthal et al. therefore test the performance of a mechanism that does away with token frequency information altogether and expresses the context for a word as a simple list of words (types) in preceding and following position. Similarity between words is then expressed as the amount of overlap between these lists of words. It is shown that, given the corpora employed, this simplified version performs as well as the rank-order correlation.

The fact that Redington et al.'s approach can be applied in a developmental setting and can be simplified considerably greatly increases its plausibility as a mechanism that can be used by language-learning children. This plausibility would be enhanced even further if the mechanism could be shown to provide a good fit to child data. Freudenthal et al. provide a first attempt at this by tracking the mechanism's developing ability to classify nouns and verbs. Using a measure of 'noun richness', they show that, when the mechanism is embedded within MOSAIC, it tends to link together nouns in the early stages, with verbs coming in in the later stages. This is due to the fact that nouns frequently appear in utterance-final position, and MOSAIC therefore registers contexts for (and thus classifies) nouns earlier than it registers contexts for verbs. The number of items that are classified in the early stages, however, is rather low, and Freudenthal et al. argue that the inclusion of utterance boundaries as framing elements could potentially increase the number of classified items.

The main aim of this paper is to further investigate the possibility that an utterance-final learner that builds word classes on the basis of distributional information can successfully simulate aspects of child speech. To this end, the noun richness score used by Freudenthal et al. is developed further to provide a descriptive measure of child speech, and thus a target for simulation. Our analyses are organised as follows.

A first analysis will focus on the relative numbers of nouns and verbs in the speech of children and their mothers. If, as hypothesized, children show highly productive use of nouns and less productive use of verbs in the early stages, one would expect the ratio of nouns to verbs in child speech to exceed that in adults.

A second analysis will focus on the extent to which the utterance-final bias instantiated in MOSAIC can account for a potential advantage for nouns in child speech. While a noun advantage is consistent with early productivity around nouns, it is also possible that it reflects the operation of an utterance-final bias in learning. This possibility will be investigated by taking the child-directed speech from the corpora used in the first analysis, and using them as input for MOSAIC models. Rote[1] output from the models will be generated at various MLU (Mean Length of Utterance) points, and the noun richness scores at various MLU points will be compared with those of the actual children.

A third analysis will determine whether the mechanism for building word classes as described by Freudenthal et al. (this volume) can improve MOSAIC's fit to the children's noun richness data. This will be done by taking the MOSAIC models at the different MLU points and linking together words that are distributionally similar based on the contexts that are represented in the models. The development of noun and verb classes will be tracked by counting the relative numbers of nouns and verbs that are classified together. Next, the rote output files from analysis 2 will be used to generate novel utterances by substituting words that have been linked on the basis of the distributional analysis, and the resulting noun richness scores will be compared to the relevant child data.

To preview the results, the analysis of child speech will show that, early in development, children's noun richness scores are more than 30 percentage points higher than those of adults. Across the MLU range studied here, children's noun richness scores drop substantially until they approximate those of adults. Simulations with rote output from MOSAIC (analysis 2) show that an utterance-final bias can account for roughly half of this decrease in noun richness (a drop of around 15%).

Substitution of distributionally similar words (analysis 3) can increase initial noun richness scores by an additional 10 percentage points, without affecting the fit to the later stages. However, this requires a substantial early noun class,

[1] Rote output contains no substitutions, and is learned directly from the input.

something that can be achieved by including utterance endings as framing elements.

Taken together, these findings suggest that children's early noun richness is jointly determined by utterance-final learning and increased productivity around nouns relative to verbs. They also confirm the viability of distributional analysis for building word classes in a developmental setting. However, they suggest that a plausible distributional analysis mechanism must show an early advantage for nouns, and thus cast doubt on mechanisms like that of Mintz (2003), which has been shown to be biased towards classifying verbs (Freudenthal et al. 2013).

Child Noun Richness

A first analysis was aimed at determining the relative rates of verb and noun use by children and their mothers. If children show productivity around nouns earlier than around verbs, one would expect children to show high rates of noun use in the early stages of development. The analysis was carried out on all 12 children from the Manchester corpus (Theakston et al. 2001). Corpora for individual children consist of 34 fortnightly recordings of interactions between child and mother, starting at a child age of approximately 2 years. Recordings are transcribed and include a MOR tier, which contains morphological coding, including Part of Speech (PoS) information for individual words.

For the current analysis, the MOR line was used to identify nouns and main verbs in the child and maternal speech. Noun richness was then expressed as the number of nouns divided by the number of nouns plus verbs. The analysis thus disregards function words and modal/auxiliary verbs (which are frequently omitted by children), and focuses on the two main classes of content words. Figure 1 presents the children's and mothers' noun richness scores. Data points represent averages for the 34 individual tapes contained in each corpus, with MLU computed for the child speech. Data reported in Fig. 1 was computed on the basis of utterance types – duplicate utterances were removed on a tape-by-tape basis. Fig. 1 shows a clear decrease in the children's noun richness scores (from .78 to .45), while the mothers' scores range between .42 and .46. Fig. 1 thus shows a clear advantage for nouns in the children's (early) speech, a finding that is consistent with the notion that young children show greater productivity around nouns than they do around verbs.

The data in Fig. 1 are consistent with greater early productivity around nouns. However, it is also possible that they reflect a processing bias for utterance-final phrases. Nouns tend to occur in utterance-final position, and the children's high early noun richness scores could potentially be explained by an utterance-final bias in learning. This possibility was examined by measuring noun richness in the output of MOSAIC models trained on the maternal speech for the 12 individual children in the Manchester corpus. The input was fed through the model multiple times, and output of increasing average length was generated after each exposure to the input. Average noun richness and MLU for the models over a range of runs are shown in Table 1. The input was fed through the models a total of 50 times, and (declarative) output was generated for runs 30 through 50. Since there is no morphology tier in the model output, words in MOSAIC's output were assigned to their most frequent PoS tag based on the MOR-lines across the Manchester corpus. As with the child analysis, duplicate utterances were removed from the analysis. As can be seen in Table 1, the model output shows a decrease in noun richness of approximately 15 percentage points between MLUs 1.5 and 4.7. The data in Table 2 therefore indicate that, while the utterance-final bias instantiated in MOSAIC can account for some of the early advantage for nouns, it is not sufficient to explain the size of the effect found in children.

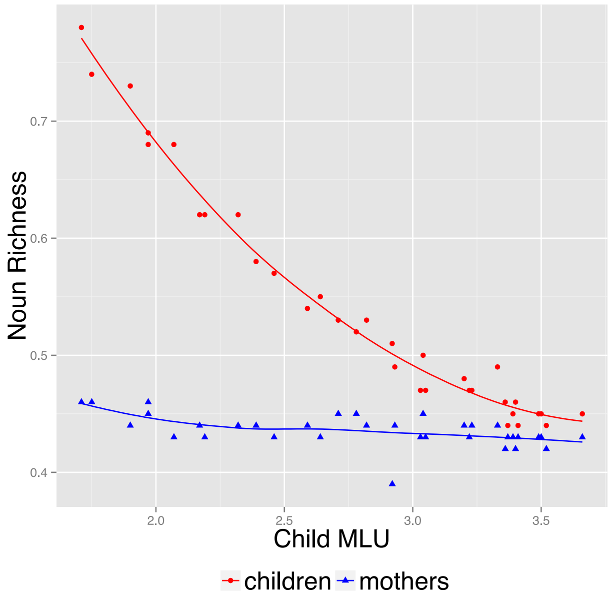

Fig. 1: Average noun richness scores for children and mothers in the Manchester corpus.

Table 1: Summary data and noun richness for rote output from MOSAIC models ranging from 30 to 50 exposures to the input. Data are averaged over 12 models

Run	MLU	Lines	Nouns	Verbs	Noun Richness
30	1.49	173	34	18	0.65
32	1.67	429	108	56	0.65
34	1.81	960	304	168	0.64
36	2.01	2,081	799	455	0.64
38	2.23	4,158	1,812	1,070	0.63
40	2.51	7,551	3,596	2,356	0.61
44	3.34	16,052	9,070	7,438	0.55
50	4.69	18,702	12,976	13,696	0.49

Building a Noun Category

In this section we investigate if early productivity around nouns can provide an additional source of noun richness. This will be done by linking together items based on the similarity of the contexts in which they occur, and substituting distributionally similar words in MOSAIC's output. Early work by Redington et al. has shown that such a distributional analysis can result in accurate word classes. However, Redington et al.'s method, has been criticized for carrying a high computational overhead (St. Clair et al. 2010). Moreover, it has not been applied in a developmental setting. Freudenthal et al. (this volume) show that a substantially pared down version of Redington et al.'s method can perform as well as the original mechanism. Specifically, the mechanism used by Freudenthal et al. does away with token frequency information. Rather than expressing the context for a given word as token counts for (150) context words in preceding and following position, context is presented as a simple list of words (types) that occurred in preceding and following position. Similarity is then expressed as a measure formally known as the *Jaccard* distance: the length of the intersection of the contexts divided over the length of the union. Two words that were preceded by {*a*, *the*} and {*a*, *green*} respectively thus have a similarity of 1/3.

Freudenthal et al. also show that the mechanism can work developmentally by applying it to the contexts represented in MOSAIC models at different points in training. More specifically, using the input corpora from the Manchester corpus (as used above), they find that the models tend to link together nouns in the early stages of development, with a verb category emerging in the later stages: Noun richness, defined as number of noun-noun links over noun-noun plus verb-verb links decreases over the model's development. The number of items that is classified, however, is relatively small, particularly early in development. Over runs 36 to 50 the average number of linked items ranges from under 30 to around 700 (see Table 2).

Table 2: Number of links and accuracy scores for Jaccard distance at different points in development. Scores are averaged over all children in the Manchester corpus. Two words are linked together if the Jaccard distance exceeds 0.2, both for preceding and following contexts. The analysis is limited to the 1000 most frequent words of each corpus

Run	# of links	Acc.	Noun Acc.	Verb Acc.	Noun Richness
36	27	0.78	0.83	0.50	0.72
38	140	0.83	0.84	0.55	0.80
40	370	0.87	0.88	0.68	0.80
44	648	0.89	0.86	0.82	0.64
50	717	0.91	0.88	0.87	0.57

However, Freudenthal et al. only consider lexical items as contexts. Earlier work has shown that the inclusion of utterance boundaries can greatly increase the number of items that are classified together. For the current simulations it was therefore decided to include utterance endings as context elements. Initial simulations showed this results in increased linking of words, but also in a decrease in accuracy scores, particularly for verbs (the highest verb accuracy achieved was .65 for run 50). Given the non-parametric nature of the Jaccard distance, this is not surprising, since an utterance boundary in the shared context can override several non-matching lexical items. It was therefore decided to give precedence to lexical context items by disregarding utterance boundaries if they were the only shared context despite lexical items being available. This essentially gives greater weight to lexical context items, if present, but still allows linking on the basis of utterance endings if no lexical items are present.

Table 3: Average number of links and accuracy scores for Jaccard distance at a threshold of 0.2. Utterance endings are included unless lexical items are present in following context but these show no overlap.

Run	# of links	Noun links	Acc.	Noun Acc.	Verb Acc.	Noun Richness
30	1	0	0.24	0.04	0.13	0.08
32	11	2	0.39	0.51	0.08	0.70
34	95	72	0.76	0.84	0.21	0.96
36	554	464	0.86	0.90	0.21	0.98
38	1,278	1,035	0.84	0.86	0.24	0.98
40	1,475	1,126	0.83	0.85	0.43	0.94
44	1,305	926	0.89	0.88	0.74	0.81
50	1,452	976	0.91	0.89	0.81	0.75

As can be seen in Table 3, inclusion of the utterance boundary greatly increases the number of linked items, particularly for the early stages (by a factor of 20 for run 36). Noun accuracy scores are comparable to those in Table 2, though verb accuracy scores are lower and only exceed .6 from run 42 onwards. However, the data in Table 3 show clear emergence of an early noun category, and a steady decrease in noun richness scores from run 36 onwards. They therefore confirm that distributional analysis can be meaningfully applied in a developmental setting and can result in reasonable accuracy scores, even when including utterance endings as framing elements.

Simulating Child Noun Richness

A final analysis was aimed at investigating whether the word classes that were built using distributional analysis are sufficient to explain the pattern of noun richness displayed by the children. This was done by taking the output files from the MOSAIC models (reported in Table 1) and using the links reported in Table 3 to substitute words that are distributionally similar. Substitution of distributionally similar words allows the model to produce novel utterances that were not present in its rote output. The early emergence of a large noun class is likely to result in many substitutions of nouns in novel contexts[2], and therefore in higher noun richness scores than for the rote output. Table 4 gives noun richness scores for the models in runs 34 through 50 with

[2] Substitution can also lead to duplicate utterances. However, since analyses are conducted on utterance types, duplicates are removed from the analysis.

substitutions based on the relevant links for a given run. Substitution can increase the number of utterances that are generated considerably. For output files containing more than 3,000 utterances, substitution was therefore carried out on a random sample of 3,000 output utterances.

It is apparent from Table 4 that (early) noun richness scores are increased relative to those in Table 2. Across the range of runs, noun richness decreases from .73 to .45, a decrease of nearly 30 percentage points, and provides a better fit to the child data. It is also apparent, however, that MLU scores are increased relative to the rote output, particularly for the later stages. This increase in MLU is caused partly by more words being linked in the later stages of development. However, it also reflects the fact that longer utterances provide more opportunities for substitution, and thus contribute more novel utterances than short utterances[3].

Table 4: Average noun richness scores for MOSAIC output with substitutions

Run	MLU	Utterances	Nouns	Verbs	Noun Richness
34	2.03	1,435	612	267	0.67
36	2.37	4,503	2,798	971	0.73
38	2.58	8,169	5,679	2,191	0.72
40	2.98	11,346	8,933	4,342	0.67
44	4.26	25,571	23,939	19,299	0.55
50	5.44	35,578	32,628	38,854	0.45

Table 5: Noun richness scores for MOSAIC output with substitutions sampled back to rote MLU distribution

Run	MLU	Utterances	Nouns	Verbs	Noun Richness
34	1.81	960	387	169	0.68
36	2.01	2,081	1,179	405	0.74
38	2.23	2,979	1,953	686	0.74
40	2.52	3,000	2,135	945	0.69
44	3.34	3,000	2,305	1,612	0.59
50	4.70	3,000	2,330	2,540	0.47

Table 5 presents the results of a noun richness analysis that controls for this increase in MLU by sampling from the generated output to match the MLU profile of the rote output. That is, for every utterance of length N in the original output, 1 utterance of length N is sampled from the output with substitutions. As can be seen in Table 5, this procedure has minimal effects on the noun richness scores, indicating that the substitution procedure results in a genuine increase in (early) noun richness scores.

[3] In fact, allowing all substitution quickly makes the output file too large to manage. The likelihood of individual substitutions is therefore gradually decreased across runs. Additionally, to prevent them from dominating the output, no substitutions are made in utterances longer than 6 words.

Conclusions

Several authors have argued that children build word classes on the basis of distributional information. The main proposed mechanisms have either been computationally expensive or too sensitive to noise, and fail to incorporate a developmental component. Evaluation has also tended to focus on the standard mechanisms of accuracy and completeness rather than empirical child data.

Freudenthal et al. (this volume) have shown that it is possible to take Redington et al.'s (1998) mechanism and apply it in a developmental setting. They have also shown that it is possible to substantially reduce its computational complexity without affecting its performance. In terms of evaluation, the revised mechanism shows a pattern of noun and verb linkage that appears qualitatively plausible, but may not build classes that are sufficiently large.

The current work builds on the work of Freudenthal et al. (this volume) by introducing a novel evaluation metric and target for simulation. The analysis of noun richness scores confirms that children have an early preference for the production of nouns relative to verbs. To our knowledge, this is the first demonstration of such a preference in corpus analyses. This finding is consistent with the notion that children show earlier emergence of a noun than a verb category, but may also reflect the operation of an utterance-final bias in learning.

The analysis of MOSAIC's rote output suggests that approximately half of children's early noun preference can be explained through utterance-final learning. This finding is encouraging as it provides independent support for the utterance-final bias instantiated in MOSAIC. However, it is also apparent that the utterance-final bias as instantiated in MOSAIC is insufficient to explain the size of children's early noun preference, and therefore suggests that some of the early noun advantage may be explained by differences in children's early productivity around nouns and verbs.

This suggestion was investigated by performing a distributional analysis on the basis of the contexts encoded in MOSAIC in different stages of development. In line with suggestions by Freudenthal et al. (this volume), utterance endings were included as framing elements. This was shown to increase the number of classified items, but had a detrimental effect on classification accuracy, in particular for verbs. It was argued that this effect could be countered by disregarding utterance boundaries when lexical context items are present but show no overlap, effectively weighting contexts for their lexical content. Doing so greatly increased the number of links for the early stages of development, whilst maintaining reasonable accuracy scores, particularly for the later stages of development.

Finally, it was shown that, with the exception of the very early stages, substitution of linked items resulted in higher noun richness scores and a better fit to the child data. Thus, the distributional analysis implemented in MOSAIC was able to build an early noun class that was sufficient to raise early noun richness to levels close to those displayed by young children.

Taken together, the results described here provide converging evidence that the productivity of children's early noun category develops more quickly than the productivity of their early verb category (Tomasello, 1992). They also suggest that distributional analysis is a viable mechanism for building word classes, even at early stages of development, when relatively few contexts may be available. The analyses reported here suggest that, provided utterance endings are included as framing elements, a variant of Redington et al.'s mechanism can form an initial noun class that is large enough to simulate children's early noun richness scores.

The findings reported here also cast doubt on the feasibility of the frequent frames approach advocated by Mintz (2003). Freudenthal et al. (2013), show that, unlike Redington et al.'s mechanism, frequent frames tend to classify together verbs, rather than nouns. Mintz's approach is thus unlikely to be successful in simulating (early) child noun richness scores. It is also difficult to see how Mintz's approach might be modified to incorporate a developmental component. Freudenthal et al. (2013) show that, while inclusion of utterance boundaries increases the numbers of nouns classified by frequent frames, this has dramatic negative effects on classification accuracy. While the analyses reported here suggest that inclusion of utterance endings may negatively affect accuracy for Redington et al.'s mechanism, they also show that this can be remedied by weighting framing elements for lexical content.

The analyses reported here further illustrate the strength of our approach, which embeds a mechanism for learning word classes in an existing model of language acquisition to simulate developmental variation in children's production of verbs and nouns. By comparing rote and productive output we were able to show that children's early noun richness is jointly determined by an utterance-final bias in learning and early productivity around nouns. We were also able to show that the developmental biases (i.e. contexts encoded) in the model are sufficient for a variant of Redington et al.'s mechanism to provide such an advantage for nouns, thus providing evidence for both these developmental biases and the feasibility of distributional analysis.

Acknowledgements

Daniel Freudenthal, Julian Pine, and Fernand Gobet are members of the International Centre for Language and Communicative Development (LuCiD) at the University of Liverpool, for which support of the Economic and Social Research Council [ES/L008955/1] is gratefully acknowledged. This research was supported by ESRC Grant ES/J011436/1.

References

Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology, 33*, 952-965.

Erkelens, M. A. (2009). *Learning to categorize verbs and nouns.* Unpublished PhD Thesis, Universiteit van Amsterdam, Amsterdam.

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. (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.

Freudenthal, D., Pine, J.M., Jones, G. & Gobet, F. (2013): Frequent frames, flexible frames and the noun-verb asymmetry. In: M. Knauf, M. Pauen, N. Sebanz E I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the Cognitive Science Society*. (pp. 2327-2332). Austin, TX: Cognitive Science Society.

Freudenthal, D., Pine, J.M., Jones, G. & Gobet. F. (2015). Simulating the cross-linguistic pattern of Optional Infinitive errors in children's declaratives and Wh-questions. *Cognition*, *143*, 61-76.

Freudenthal, D., Pine, J.M., Jones, G. & Gobet, F. (this volume). Developmentally plausible learning of word categories from distributional statistics. *Proceedings of the 38th annual meeting of the Cognitive Science Society.*

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

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

Olguin, R., & Tomasello, M. (1993). Twenty-five-month-old children do not have a grammatical category of verb. *Cognitive Development*, 8, 245-272.

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

Stumper, B., Bannard, C., Lieven, E., & Tomasello, M. (2011). Frequent frames in German child-directed speech: A limited cue to grammatical categories. *Cognitive Science, 35*, 1190-1205.

St. Clair, M.C. Monaghan, P., & Christiansen, M.H. (2010). Learning grammatical categories from distributional cues. Flexible frames for language acquisition. *Cognition*, *116*, 341-360.

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. (1992). First verbs: A case study of early grammatical development. Cambridge: CUP.

Tomasello, M., & Olguin, R. (1993). Twenty-three-month-old children have a grammatical category of noun. *Cognitive Development, 8*, 451-464.