

Freudenthal, D., Pine, J. M., & Gobet, F. (2008). On the utility of conjoint and compositional frames and utterance boundaries as predictors of word categories. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

On the Utility of Conjoint and Compositional Frames and Utterance Boundaries as Predictors of Word Categories

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

Julian 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

This paper reports the results of a series of connectionist simulations aimed at establishing the value of different types of contexts as predictors of the grammatical categories of words. A comparison is made between 'compositional' frames (Monaghan & Christiansen, 2004), and non-compositional or 'conjoint' frames (Mintz, 2003). Attention is given to the role of utterance boundaries both as a category to be predicted and as a predictor. The role of developmental constraints is investigated by examining the effect of restricting the analysis to utterance-final frames. In line with results reported by Monaghan and Christiansen compositional frames are better predictors than conjoint frames, though the latter provide a small performance improvement when combined with compositional frames. Utterance boundaries are shown to be detrimental to performance when included as an item to be predicted while improving performance when included as a predictor. The utility of utterance boundaries is further supported by the finding that when the analysis is restricted to utterance-final frames (which are likely to be a particularly important source of information early in development) frames including utterance boundaries are far better predictors than lexical frames.

Introduction

Several authors have argued that co-occurrence statistics can serve as a powerful cue that children utilise in determining the grammatical category of words they encounter in the linguistic input they hear. For instance, following work by Finch and Chater (1994), Redington, Chater and Finch (1998) showed that words of the same grammatical category tend to have a high degree of overlap in terms of the context vectors that encode the words that precede and follow the target words. Thus, nouns tend to be preceded by determiners and adjectives, and followed by verbs. Similarly, verbs tend to be preceded by (pro)nouns and followed by determiners and (pro)nouns. One major question that has arisen from this line of work concerns how useful different types of contexts are for classifying target words.

While Redington et al. treated preceding and following contexts as independent, Mintz (2003) assessed the value of conjoint contexts or frames: a pair of words with one word intervening between them. The notion of a frame is intuitively appealing as frames are more constraining than independent contexts and can therefore be expected to result in grammatical classes that are of higher quality than

categories derived from independent contexts. Mintz extracted from corpora of child-directed speech the 45 most frequent frames and determined the overlap in terms of grammatical category between the words that occurred in these individual frames. Mintz concluded that these frames were good predictors for grammatical category in terms of accuracy but less so in terms of completeness. That is, while the words that co-occurred in particular frames had a high likelihood of belonging to the same category, words from the same category tended to occur in many different frames. While frames classified some 50% of the word tokens in the input file, completeness in terms of the percentage of word types categorized was also relatively low at approximately 15% (Monaghan & Christiansen 2004).

Monaghan and Christiansen (2004) provide a direct comparison of conjoint and independent contexts by training a neural net to predict the grammatical category of target words on the basis of several types of context derived from the corpus of maternal speech directed at Anne from the Manchester corpus (Theakston, Lieven Pine & Rowland, 2001). In contrast to Mintz, Monaghan and Christiansen did not restrict their contexts to the most frequent frames, but drew their frames from the whole input corpus.

Monaghan and Christiansen found that a model trained using independent contexts (or compositional frames) outperformed a model trained on conjoint frames. The model trained on conjoint frames performed no better than a base-line model that was trained on randomized frames. The model trained on conjoint frames, however, displayed a default effect. Monaghan and Christiansen included in their simulation frames that contained utterance boundaries as their middle element. That is, the utterance boundary was included as a category that the model learned to predict. The model that was trained on conjoint frames predicted the utterance boundary for all stimuli. While the performance of the model trained on compositional frames (which also had to learn to predict utterance boundaries) was significantly better, the default effect displayed by the model trained on conjoint frames raises questions about the role of the utterance boundary in these simulations. Frames containing the utterance boundary made up a significant proportion (~25%) of the stimuli. Given that the amount of variation in frames that straddle the utterance boundary is likely to be relatively high and the fact that the utterance boundary is not very meaningful as a grammatical category, one may

wonder how a model trained on frames would perform if frames that straddle the utterance boundary were excluded from the training set.

While the utterance boundary may not be very meaningful as a grammatical category, the value of the utterance boundary as a predictor has received relatively little attention in the literature. Mintz restricts his analysis to the 45 most frequent *lexical* frames: frames that contain a word in the two anchor positions. Monaghan and Christiansen drew frames from the entire input set but did not include frames with an utterance boundary at the anchor points. This relative lack of attention to the utterance boundary as a predictor is somewhat surprising given that the types of items that occur in utterance-initial and utterance-final position are clearly restricted, particularly when viewed in the context of frames. Frames containing utterance boundaries are also quite frequent: An analysis of the maternal speech directed at Anne reveals that frames containing an utterance boundary make up roughly 40% of all frames in terms of tokens and nearly 15% in terms of types. This suggests not only that there are many different frames with utterance boundaries, but also that some of these frames are actually very frequent. In fact, when allowing utterance boundaries in frames, it becomes apparent that frequent frames are predominantly frames with utterance boundaries: 44 of the 50 most frequent frames in the maternal speech directed at Anne contain an utterance boundary. This high frequency of frames containing utterance boundaries makes it unlikely that children would not be sensitive to them. This is even more apparent when one considers that some of these frames are very good predictors of the grammatical class of the items that appear in them. Thus, the most frequent frame in Anne's input is 'The X END' which contains 564 different items, the overwhelming majority of which are nouns. By comparison, the total number of words from Anne's corpus that was classified by Mintz is 405. Thus, one single frame that contains an utterance boundary classifies a larger number of words than the 45 frequent lexical frames selected by Mintz.

The role of the utterance boundary also raises a third issue about the usefulness of frames in the learning of grammatical categories: the potential role of developmental constraints in restricting the learner's access to distributional information in different parts of the utterance. The analysis of corpus statistics is a frequently used tool in studies aiming to determine the value of particular sources of distributional information. It is common practice in such studies to analyse the statistics of complete utterances. An implicit assumption here is that the statistics of complete utterances are available to children. Such an assumption may not be justified given the available child data. The first 'utterances' of children often consist of isolated words. As children grow older, their utterances gradually become longer until the mean length of their utterances (MLU) matches that of adults. The fact that young children's utterances are considerably shorter than adults' utterances raises the possibility that children may only represent partial

utterances, and hence may only track the statistics of partial utterances¹. Work with MOSAIC (Freudenthal et al. 2006, 2007a, 2007b) shows that the developmental patterning of a number of key phenomena in child speech can be successfully simulated using a learning mechanism that produces progressively longer utterance-final phrases. This finding suggests that children early in the acquisition process may be particularly sensitive to the material that occurs at the end of the utterance. Analyses of the distributional statistics of complete utterances may therefore examine information that is not available to the child. When one further considers that different locations in the sentence differ in terms of the types of items that are likely to occur there, it becomes apparent that developmental constraints may place important restrictions on the types of information that children may usefully employ in the acquisition of syntactic categories. Such constraints may further prove important in explaining developmental patterns in the data, such as children's greater willingness to use novel nouns than verbs in contexts in which they haven't been previously encountered (Tomasello, 2000).

The aims of this paper are to assess the relative virtues of conjoint and compositional frames as well as the role of the utterance boundary as a predictor the grammatical category of words. In order to allow a comparison with earlier work we followed the approach taken by Monaghan and Christiansen (2004). We trained a neural net with the same structure as that used by Monaghan and Christiansen to predict the category of target words based on different types of contexts. The presence of utterance boundaries as a predictor as well as a grammatical category was manipulated. In order to explore the potential role of developmental constraints, we additionally carried out simulations using only frames that occurred in utterance-final position. Previous work with MOSAIC has suggested that children are particularly sensitive to this position.

The Simulations

The simulations were run using LENS, with learning parameters set to their defaults. The model was a feed-forward network with the input units fully connected to a bank of 10 hidden units which was fully connected to an output layer. The number of output units was equal to the number of grammatical categories: 12 for simulations where the utterance boundary was included as a category and 11 where it was excluded. The number of inputs varied with the number and type of frames used in the simulations. Models that were trained on conjoint frames utilized one (large) bank of input units: one unit for every distinct frame. Models that were trained on compositional frames used two independent banks of input units that were fully connected to the hidden layer. The first bank of units represented the

¹ Of course, the fact that children initially only tend to produce short utterances does not necessarily mean that they only represent short utterances. However, it does at least raise the possibility that they may not represent all of the information in the utterances that they are analyzing.

first word in the frame while the second bank represented the last word in the frame. The number of units in these banks was equal to the number of distinct words making up the frames. Training the model with two independent banks of inputs allows the model to take into account the identity of the preceding and following word rather than the (dependent) frame. Training proceeded by exposing the model to a vector encoding the frame on the input layer and a vector encoding the category of the word in the frame on the output layer. All models were trained for 5 epochs where an epoch is one sweep through the entire training set. Testing took place on the training set.

Simulation 1: Conjoint vs. Compositional frames

The first simulation was aimed at replicating the results of Monaghan and Christiansen (2004). Like Monaghan and Christiansen, we used the maternal speech directed at Anne from the Manchester corpus (Theakston, Lieven, Pine & Rowland, 2001), available from the CHILDES data base (MacWhinney, 2000). All lexical conjoint and compositional frames (including those that straddled utterance boundaries, but excluding the boundary as a predictor) were selected and the category of the word appearing in the frame was extracted from the MOR-line contained in the CLAN transcripts. There was a total of 12 word categories (including the utterance boundary). Contracted forms that combine a (pro)noun and copula or modal verb (e.g. *He's*) were ignored as a grammatical category, but were included as predictors. This resulted in a total of 42,303 conjoint frames and a total of 93,212 stimuli. The input layer for the model trained on conjoint frames thus consisted of 42,303 units, and the individual frames were represented by 42,303 orthogonal input vectors.

There was a total of 3,324 different words in the input represented by 3,324 orthogonal vectors. The model trained on compositional frames thus used two input banks, each with 3,324 units. After training, the model was tested by determining if it predicted the correct word category given the frame as input. Table 1 gives the results for the different word categories. Overall performance was assessed through two measures: Accuracy and Coverage. Accuracy is simply the proportion of words correctly classified across all categories. Coverage is the average of the proportion correct for the different categories. This measure is not sensitive to differences in the number of stimuli in the different categories and thus provides a better measure of how well the model has learned the entire system.

As can be seen in Table 1 the model trained on compositional frames clearly outperforms the model trained on conjoint frames, both in terms of accuracy and in terms of coverage. Both models perform best on the utterance boundary, but the model trained on conjoint frames does not display the default effect reported by Monaghan and Christiansen (2004)². When excluding the utterance

boundary from the results the accuracy of the models drops to 38.7% for conjoint frames and 72.3% for compositional frames. Coverage also decreases, from 23% to 16% for conjoint frames and from 42% to 37% for compositional frames.

Thus, despite the model not showing the perfect default effect that was reported by Monaghan and Christiansen, it is clear that performance on the other categories is lower than on the utterance boundary. The utterance boundary, however, is not very meaningful as a syntactic category and, due to its high frequency, its inclusion has the potential to seriously degrade the model's performance on the other categories. This possibility was investigated in the second set of simulations.

Table 1: Percentage correctly classified in simulation 1.

CATEGORY	N	% CORRECT	
		Conjoint frames	Compositional frames
Prepositions	6699	9.2	65.2
Wh-words	699	0	0
Determiners	11901	8.9	86.7
Conjunctions	1281	0	0
Pronouns	11094	64.4	76.8
Numerals	117	0	0
Adverbs	1491	0	0
Interjections	306	0	0
Adjectives	2278	0	30.0
Nouns	5790	23.1	67.5
Verbs	18157	71.6	84.9
Boundary	33390	98.5	92.3
Total	93212	60.1	79.5
Coverage		23.0	42.0

Simulation 2: Excluding the boundary as a target

This set of simulations was similar to simulation 1, with the only difference being that the utterance boundary was removed as a target for prediction. This reduced the number of training items to 59,822. The number of distinct conjoint frames (and hence input units) was reduced to 25,235.

As can be seen in Table 2, excluding the utterance boundary as a target for prediction increases the accuracy for the other categories. For the model trained on conjoint frames overall accuracy on the lexical categories has increased from 38.7% to 58.5%. Coverage, however, is still relatively low at 23%. Accuracy and coverage on the lexical categories for the model trained on compositional models have increased slightly as well. These results suggest that, while the inclusion of the utterance boundary as a target

² There are a number of potential reasons for this difference between our results and those of Monaghan and Christiansen. First, there are differences in the simulations in terms of the

parametrisation of the neural net used. Second, Monaghan and Christiansen obtained word categories from the CELEX data base while we used categories obtained from the MOR-line in the CLAN transcripts. Third, differences in the preparation of the input (cleaning up and filtering of the transcripts) may have lead to differences in the training materials.

does not have a particularly large effect, it does lead to decreased performance. It is also clear from Table 2 that, as in the previous simulations, the model trained on compositional frames outperforms the model trained on conjoint frames, in particular on Prepositions, Adjectives and Nouns.

Table 2: Results for conjoint and compositional frames excluding the utterance boundary as a target for prediction.

CATEGORY	N	% CORRECT	
		Conjoint frames	Compositional frames
Prepositions	6699	7.4	80.6
Wh-words	699	0	0.3
Determiners	11901	69.4	87.3
Conjunctions	1281	0.0	5.4
Pronouns	11094	67.6	82.6
Numerals	117	0	0
Adverbs	1491	0	0
Interjections	306	0	0
Adjectives	2278	0	33.3
Nouns	5790	13.3	73.4
Verbs	18157	99.0	91.5
Total	59822	58.5	78.0
Coverage		23.3	41.3

Simulation 3: Using the boundary as a predictor

While the utterance boundary is not very meaningful as a lexical category, it was argued earlier that it can serve as a powerful predictor when included in a frame. The next set of simulations, reported in Table 3, tested this possibility. The utterance boundary as a target for prediction was excluded in this (and all following) simulations.

Table 3: Results for conjoint and compositional frames including the utterance boundary as a predictor.

CATEGORY	N	% CORRECT	
		Conjoint frames	Compositional frames
Prepositions	9101	18.9	65.2
Wh-words	2716	38.7	51.4
Determiners	13532	82.2	84.9
Conjunctions	2506	0	21.3
Pronouns	19676	79.8	74.0
Numerals	246	0	0
Adverbs	4233	0	33.4
Interjections	1689	0	8.4
Adjectives	3882	0	39.8
Nouns	15135	72.0	76.1
Verbs	28779	93.5	91.8
Total	101495	66.4	73.8
Coverage		35.0	49.7

As can be seen in Table 3, performance for the model trained on conjoint frames has increased most for Nouns and Wh- words. For compositional frames performance gains are seen for Wh- words, Conjunctions, and Adverbs. Inclusion of the utterance boundary thus leads to better performance for the models, in particular in terms of Coverage which increases by around 10 percentage points. It is also worth noting that the accuracy in these models is obtained over a much larger set of stimuli: approximately 100,000 items compared to approximately 60,000 items for the previous set of simulations.

Using Wh- words as an example, it is easy to see why inclusion of the utterance boundary as a predictor results in improved performance. While Wh- words can occur after lexical items (e.g. So/And what do you want?) they overwhelmingly occur in sentence-initial position. What's more, many of these utterance-initial frames (for instance 'BEG X Do') are highly predictive of Wh- words.

Simulation 4: Extended frames

While the previous simulations confirmed that compositional frames are better predictors than conjoint frames, it is possible that sensitivity to both conjoint and compositional frames is superior to sensitivity to just compositional frames. This was examined in the next set of simulations. In these simulations the network utilized three banks of input units, corresponding to the two independent banks used for compositional frames as well as the large bank used for conjoint frames. For completeness, these simulations were run with and without utterance boundaries as predictors. The results of these simulations are shown in Table 4. Both simulations show slightly higher levels of accuracy and coverage than the previous simulations.

Table 4: Results for extended frames with and without boundaries as predictors.

CATEGORY	Lexical extended Frames		All extended Frames	
	N	% CORRECT	N	% CORRECT
Prepositions	6699	83.6	9101	76.7
Wh-words	699	4.3	2716	52.0
Determiners	11910	91.3	13532	88.7
Conjunctions	1281	16.9	2506	28.5
Pronouns	11094	88.3	19676	81.4
Numerals	117	0	246	0
Adverbs	1491	0	4233	35.2
Interjections	306	0	1689	16.3
Adjectives	2278	47.9	3882	38.5
Nouns	5790	71.3	15135	80.4
Verbs	18157	95.7	28779	89.7
Total	59822	81.8	101495	77.2
Coverage		45.3		53.4

Simulation 5: The role of development

The final set of simulations concerned the role of development. It was argued earlier that computation of distributional statistics over a full corpus may lead researchers to represent information that is not available to the developing child. In line with the results from simulations run with MOSAIC (Freudenthal et al. 2006, 2007a, 2007b) the early stages of development were simulated by extracting from Anne’s corpus all utterance-final frames (both lexical frames and frames containing an utterance boundary)³. Two simulations were run. The first simulation was trained on all lexical frames. The second simulation was trained on the lexical frames plus the frames containing an utterance boundary. Given the relatively poor performance of conjoint frames in the earlier simulations, these simulations were run using compositional frames only. The results of the simulations are shown in Table 5.

Table 5: Percentage correctly classified for utterance-final compositional frames with and without boundaries.

CATEGORY	Lexical, compositional Frames		All compositional Frames	
	N	% CORRECT	N	% CORRECT
Prepositions	2001	0	2774	0
Wh-words	25	0	368	0
Determiners	6060	100.0	6889	84.9
Conjunctions	185	0	267	0
Pronouns	3434	0	8132	42.9
Numerals	54	0	151	0
Adverbs	439	0	2845	0
Interjections	125	0	1160	0
Adjectives	1163	0	2627	0
Nouns	1923	0	10941	84.1
Verbs	3183	0	8316	29.6
Total	18592	32.6	45470	46.2
Coverage		9.0		22.0

As can be seen in table 5 the model that was trained on just lexical frames displays a clear default effect: the model predicts the determiner for each and every stimulus. This default effect appears to be caused by the fact that frames containing the determiner are the most frequent of the different categories, making up almost a third of all stimuli.

The model that was trained on frames including the utterance boundary performs considerably better, correctly classifying 46% of all stimuli and obtaining a coverage of 22%. The fact that these numbers are considerably lower than those reported in the simulations that were trained on the frames from the entire corpus is not surprising as utterance-final frames represent a subset (both in number and type) of the frames contained in the whole of the input. It is also worth noting that high accuracy is not necessarily

desirable in this particular case, as it seems unlikely that children early in development will classify words of all categories with (equally) high accuracy. Moreover, what is interesting about the model trained on all utterance-final frames is that it displays a clear advantage for the prediction of Nouns (84.1%) over the other categories (including Verbs; 29.6%). This finding corresponds well to the results of Akhtar and Tomasello (1997) who found that children are more likely to use novel nouns than novel verbs in contexts in which they have not been encountered, a finding which suggests that children form a productive noun category before a productive verb category (Tomasello, 2000). Taken together, these results suggest that children may be particularly sensitive to the distributional statistics of the endings of utterances and thus provide converging evidence for the constraints on the learning mechanism in MOSAIC which employs a strong utterance final bias. These results furthermore suggest that care should be taken in evaluating the utility of cues on the basis of full corpus analyses. While lexical compositional frames appear good predictors when taking the entire corpus into account (cf. Simulation 2), their value is extremely limited when the analysis is limited to frames that are likely to be available to language learning children in early stages of development.

Conclusions

The simulations reported in this paper were aimed at answering four main questions: First, we wanted to assess the relative virtue of using dependent contexts or conjoint frames versus independent contexts or compositional frames as predictors of the grammatical categories of the items contained in them. Second, we wanted to establish if the inclusion of the utterance boundary as a grammatical category may have been a factor in the default effect reported by Monaghan and Christiansen (2004). Third, we wanted to investigate the effect of including the utterance boundary as a predictor. Fourth, we were interested in how development might impact on the model’s accuracy in the prediction of different grammatical categories.

Regarding the first two questions, the simulations reported here show that inclusion of the utterance boundary as an item to be predicted does hinder the model’s ability to predict other grammatical categories. While the adverse effects of including the utterance boundary are not particularly large, the utterance boundary is not a very meaningful grammatical category. These results therefore suggest that it is preferable to exclude the utterance boundary as a grammatical category in future studies.

Regarding the relative virtue of conjoint or compositional frames, our results are broadly in line with those reported by Monaghan and Christiansen (2004). The performance of models trained on compositional frames is substantially better than that of models trained on conjoint frames, though the combination of the two (in ‘extended’ frames) does result in a slight performance improvement, both in terms of coverage and accuracy.

³ The utterance ‘he goes home’ thus contributed the frames ‘he X home’ and ‘goes X END’.

The reason why compositional frames perform better becomes apparent when considering the task faced by the network. In the simulations reported here the network learns to predict the grammatical category of an item based on the frame in which it occurred. A disadvantage of frames in this task is that individual frames may not occur with meaningful frequencies. Simulation 2 employed a total of approximately 60,000 stimuli made up of approximately 25,000 distinct frames. This means that many frames will have only occurred once in the entire stimulus set, thus making them hard to learn. While the same frequency distribution applies for compositional frames, a model trained on compositional frames is able to generalise from the statistics on the individual preceding and following items. More frequent frames that contain items from multiple categories suffer from a similar problem. When faced with a conjoint frame that contains items from multiple categories, the model is likely to respond by predicting the category that occurs within that frame most frequently. Unlike a model trained on compositional frames, it is unable to use the information from the items that make up the frame to override the default category for that frame.

While compositional frames are clearly superior within the task employed here, conjoint frames may still have advantages in other tasks. Freudenthal et al. (2007c), for instance, compared conjoint and compositional frames in a substitution task. In this task the model compared pairs of words and determined if they could be considered equivalent (and subsequently substituted in the production of output) on the basis of the amount of overlap in the (dependent or independent) contexts in which they had occurred. Since a word is likely to occur in multiple contexts, this task employs a notion of variability for both dependent and independent contexts. Freudenthal et al. (2007c) concluded that for the effect they simulated (greater substitution of nouns than verbs) conjoint frames provide a better fit to the data.

A third aim of this paper was to examine the role of the utterance boundary as a predictor rather than grammatical category. The inclusion of the utterance boundary resulted in better performance in terms of Coverage (but not necessarily in terms of Accuracy). It was argued that the improved performance was the result of the expansion of the training set with a number of frames that allowed the model to predict word types that tend to occur in utterance-initial or utterance-final position.

A fourth aim of the simulations was to determine how increased sensitivity to utterance-final position arising from developmental constraints may impact on the types of categories that can be learnt. It was shown that a model trained on lexical utterance-final frames defaulted to predicting the determiner. Inclusion of utterance boundaries in the frames resulted in a more plausible pattern of results with the model showing superior performance on nouns compared to verbs. It was argued that this latter finding, in association with children's greater productivity around nouns, constitutes converging evidence for the utterance-

final bias employed in MOSAIC. It also suggests that systems that track the distributional statistics of the entire input regardless of location in the utterance may utilise information that is not available to the language learning child in the early stages. This may result in a failure to capture the patterns in the data that are typical of the language learning child and lead researchers to overstate the utility of certain cues for the prediction of grammatical categories.

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.
- 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., Aguado-Orea, J. & Gobet, F. (2007a). 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. (2007b). 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. & Gobet, F. (2007c). Simulating the Noun-Verb asymmetry in children's productive speech. *Proceedings of the 8th International Conference on Cognitive Modelling* (pp. 115-120). New York: Psychology Press.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk (3rd Edition)*. Mahwah, NJ: Erlbaum.
- Mintz, T. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91-117.
- Monaghan, P. & Christiansen, M. (2004). What information is useful and usable in language acquisition? *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.
- Redington, M., Chater, N. & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425-469.
- Theakston, A.L., Lieven, E.V.M., Pine, J.M. & Rowland, C.F. (2001). The role of performance limitations in the acquisition of Verb-Argument structure: An alternative account. *Journal of Child Language*, 28, 127-152.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, 209-253.