

# Frequent Frames, Flexible Frames and the Noun-Verb Asymmetry

Daniel Freudenthal, Julian Pine

School of Psychology, University of Liverpool

Gary Jones

School of Psychology, Nottingham Trent University

Fernand Gobet

School of Psychology, University of Liverpool

## Abstract

In this paper we compare several mechanisms for using distributional statistics to derive word class information. We contrast three different ways of computing statistics for independent left and right neighbours with the notion of a frequent frame. We also investigate the role of utterance boundaries as context items and weighting of frequency information in terms of the successful simulation of the noun-verb asymmetry. It is argued that independent contexts can classify items with a higher degree of accuracy than frequent frames, a finding that is more pronounced for larger input sets. Frequent frames classify a larger number of items, but do so with lower accuracy. Utterance boundaries are useful for the development of a noun category, particularly at intermediate levels of frequency sensitivity.

**Keywords:** Word class derivation, independent contexts, frequent frames.

## Introduction

Several authors have shown that distributional statistics can provide powerful cues for acquiring syntactic categories; words that belong to the same syntactic category tend to be preceded and followed by the same words. Thus, nouns tend to be preceded by determiners and adjectives and followed by verbs. Redington, Chater and Finch (1998), building on work by Chater and Finch (1992), investigate several variants of the same basic principle: for any of a set of target words, a context vector was derived that contained (rank orders of) counts of the 150 most frequent words in the corpus, in positions preceding and following the target words. Redington et al. computed correlations between the context vectors of the target words, which were then used as input to a Hierarchical Cluster Analysis, and concluded that the resulting classes mapped closely onto broad syntactic classes. Redington et al. explore a number of variants of the basic mechanism, but get their best results by using a context of one preceding and one following word, and using a rank order correlation as their distance measure.

An alternative mechanism for acquiring syntactic categories has been proposed by Mintz (2003). Mintz introduces the notion of a frequent 'frame': two lexical items with one word intervening (e.g. *He X to*). Mintz argues that the (45) most frequent frames in the (English) corpora he analyses show high internal consistency in

terms of the grammatical category of the items that occur in the central position. The notion of a frequent frame is therefore thought to provide a powerful cue that children might employ in the acquisition of syntactic categories. More recent work has confirmed the utility of frequent frames for French (Chemla et al. 2009), but results have been less promising for languages with relatively free word order such as Dutch (Erkelens, 2009) and German (Stumper et al. 2011).

A major difference between the approaches of Redington et al. and Mintz is that the approach described by Redington is inherently graded and frequency sensitive in nature. Thus, in this approach, co-occurrence statistics are collected across all uses of a particular word. Depending on the exact implementation, the approach can also show varying degrees of frequency sensitivity with context vectors containing (rank orders of) word counts. Similarity is then expressed as a correlation-like measure across context vectors, which can be interpreted as a probability of two items being of the same class. This graded context-sensitivity is absent from Mintz's approach. Thus, while Mintz's analysis is restricted to the 45 most frequent frames, it clusters together all items that co-occur in one of these frames. The approach therefore ignores many contexts in which a word may occur, and instead clusters items on the basis of (potentially one) occurrence in specific high frequency contexts.

Typically, mechanisms for extracting grammatical categories are evaluated in terms of accuracy (the extent to which items that are clustered together belong to the same syntactic category) and completeness (the extent to which all items within one syntactic category are clustered together). St. Clair et al. (2010), as well as Monaghan (2004), compared frames and independent contexts as used by Redington et al. in the context of connectionist simulations, and found that frames were accurate but resulted in low completeness while independent contexts performed similarly in terms of accuracy but outperformed frames in terms of completeness.

However, while high accuracy is clearly a desirable property of a mechanism that derives syntactic categories (children for instance make very few word class errors), it is less clear if high completeness is desirable, particularly if one is interested in modeling children's early linguistic abilities. Thus, while children ultimately develop

linguistic abilities that suggest the presence of relatively abstract linguistic categories, their early multi-word speech has been characterized as lexically specific. Moreover, completeness is often measured across word classes, when there appear to be developmental discontinuities in children's productive use of grammatical categories. Thus, children appear to be more prepared to produce novel nouns than novel verbs in familiar contexts (Akhtar & Tomasello, 1997), a finding which has led to the suggestion that children may develop a productive noun category earlier than a productive verb category (Tomasello, 2000).

On the basis of these considerations, it would seem that a mechanism that is plausibly employed by language learning children is one that favours accuracy over completeness and favours the linking of high numbers of nouns over the linking of high numbers of verbs. One factor that might impact on the relative likelihood of linking nouns and verbs is the weighting of frequency information. Nouns for instance have a relatively high likelihood of being preceded by one of a small set of determiners. A second factor that is likely to affect the relative linking of nouns and verbs is the availability of utterance boundaries as framing elements. Nouns have a relatively high likelihood of occurring in utterance final position, and the utterance boundary is thus a potentially powerful cue for a noun category. Freudenthal et al. (2008), in the context of connectionist simulations, provide some evidence in support of this suggestion.

The main aims of this paper then are as follows: 1. To compare frequent frames and measures similar to those used by Redington et al. (1998) in terms of their ability to simulate the word classes apparent in children's early speech, 2. To investigate how different levels of frequency sensitivity as well as the availability of utterance boundaries may impact on these mechanisms.

### Similarity measures used

In the current paper we compare 4 different measures of similarity. We consider the frequent frames approach described by Mintz (2003), as well as 3 different implementations of the independent contexts approach described by Redington et al. (1998). In line with Redington et al., we considered as target words (i.e. words to be classified) the 1000 most frequent words in the corpus. The frequent frames approach closely followed the implementation by Mintz: we considered the target words that co-occurred in the 45 most frequent lexical frames within a corpus. The implementations of independent contexts closely followed the implementation of Redington et al. The context for a given word was encoded as a vector of length 300 consisting of counts of the 150 most frequent words in the corpus in the position directly preceding and following the target word. The actual similarity measures based on these vectors were: 1. Spearman rank order correlation (as used by Redington et al); 2. Cosine similarity based on raw frequency counts;

and 3. Cosine similarity based on the square root of frequency counts. These three different measures differ with respect to the weighting of frequency information, which is highest for cosine similarity based on raw counts and lowest for rank orders. Weighting of frequency information is even lower for frequent frames: this measure only considers whether or not target words co-occur within a given frame, not how often they co-occur.

### Corpora

The analyses were performed on the child-directed speech of the 12 children in the Manchester corpus (Theakston et al., 2001). The child-directed speech in the Manchester corpus is typically in the range of 25,000 to 30,000 utterances per child. Corpora were cleaned up minimally, and only multi-word utterances were analysed. For all corpora the following statistics were collected: for every word in the corpus, counts were collected for the items that preceded and followed it, as well as the frames (A X B) that surrounded them. Frame counts were then tallied across words to determine overall counts for (frequent) frames.

One additional manipulation involved the merging of the corpora for the 12 individual children into one large corpus. This manipulation was included to determine whether the mechanisms under investigation are differentially affected by changes in corpus size.

For the purpose of determining accuracy of derived word pairings, words were assigned their respective Part of Speech (POS) tag as employed on the %mor coding tier of the corpus. POS tags in the Manchester corpus are relatively detailed and distinguish between main verbs and auxiliaries, as well as nouns, pronouns and proper nouns. The categories employed here are therefore similar to what Mintz terms 'expanded' labelling. Where multiple POS tags were used for one word form (e.g. forms that can be used as either noun or verb) the most frequently used POS tag was assigned. That is: words forms were assigned to the grammatical category in which they were used most frequently.

### Results

There are several ways in which one can evaluate the success of different similarity measures. Redington et al. (1998) performed a cluster analysis, and plotted accuracy and completeness as a function of the similarity level (or number of clusters extracted). While this is informative, it does raise a number of problems in interpreting the outcome. First, it is not immediately obvious at what similarity level one should compare different mechanisms, and second, the clustering process itself can be performed in different ways which have the potential to influence the results in terms of accuracy and completeness. For these reasons we opted to sidestep the clustering process, and perform a more direct evaluation of the similarity scores. This was done by extracting all possible word pairs, and computing the relevant similarity

Table 1: Accuracy and number of classified words for the four different distance measures.

|         | Cos-Sim, Raw Freq |      | Cos-Sim. Sqrt Freq |      | Spearman rank order |      | Frequent Frames |       |
|---------|-------------------|------|--------------------|------|---------------------|------|-----------------|-------|
|         | Acc.              | N    | Acc.               | N    | Acc.                | N    | Acc.            | N     |
| Anne    | .90               | 1555 | .92                | 2367 | .91                 | 1314 | .69             | 8688  |
| Aran    | .95               | 596  | .94                | 2046 | .92                 | 1772 | .66             | 23612 |
| Becky   | .83               | 565  | .82                | 1718 | .88                 | 1746 | .71             | 5423  |
| Carl    | .88               | 576  | .92                | 2113 | .92                 | 1894 | .70             | 6191  |
| Dominic | .81               | 409  | .84                | 1418 | .84                 | 1313 | .62             | 7904  |
| Gail    | .84               | 355  | .89                | 1223 | .91                 | 1152 | .59             | 8381  |
| Joel    | .76               | 329  | .86                | 1063 | .90                 | 1089 | .60             | 7182  |
| John    | .76               | 326  | .83                | 1938 | .85                 | 2570 | .73             | 9269  |
| Liz     | .78               | 340  | .83                | 1171 | .85                 | 1546 | .66             | 5354  |
| Nicola  | .78               | 356  | .89                | 1115 | .91                 | 1272 | .62             | 9949  |
| Ruth    | .92               | 276  | .91                | 910  | .89                 | 1024 | .61             | 8089  |
| Warren  | .79               | 403  | .84                | 1199 | .86                 | 1222 | .70             | 14405 |
| Average | .84               | 507  | .87                | 1523 | .89                 | 1492 | .66             | 9537  |
| Merged  | .93               | 1559 | .95                | 9561 | .93                 | 8636 | .55             | 76676 |

metrics for every word pair. Where a similarity metric exceeded a certain threshold the word pair was considered to belong to the same category.

This procedure obviously raises the question of what threshold should be chosen for the different similarity metrics. Generally, higher values for the threshold will result in higher levels of accuracy and lower numbers of classified items, but these numbers may differ across metrics for a specific value of the threshold. For this reason, we decided to choose a different value for the threshold across metrics such that the resulting accuracy was always relatively high (~ 90%) and comparable across the metrics<sup>1</sup>, thus allowing for a meaningful interpretation of differences in completeness. Table 1 shows percentage accuracy as well as number of classified items for the 12 individual children in the Manchester corpus, the average for these children as well as scores for the merged Manchester corpus.

The concept of a threshold for classification is irrelevant for frequent frames (thus making it impossible to peg accuracy at 90%), as the notion of a frame entails that two items that co-occur in one of the frequent frames are of the same word class. Table 1 therefore lists accuracy and number of classified items for all word pairs that co-occurred in one of the 45 most frequent frames in

the relevant corpus. Word pairs that co-occurred in multiple frames were counted only once.

Looking at the individual children and their average in Table 1, it is obvious that there are substantial differences between the different metrics. Frequent frames classify a large number of pairs, but do so at relatively low accuracy. Accuracy levels for frequent frames are lower than reported by Mintz (who reports a type accuracy of .91). This lower accuracy is at least partly caused by the fact that, for the current analyses, words were assigned to their most common category. While such a procedure makes sense for graded measures that collate statistics over different contexts, it may be less appropriate for the frame style analysis. Thus several words can be used as either a noun or a verb (e.g. *pull*, *paint*). In the corpora employed here, *pull* is overwhelmingly used as a verb, while *paint* is used as a noun more often than a verb (and as a consequence, is considered to belong to the noun class for the current analyses). The frequent frames analysis will classify these items together (resulting in a false alarm) because they co-occur in the frame *you X your*.

For this reason we performed a second, contextual accuracy analysis on the frames analysis: for every word pair that co-occurred in one of the frequent frames, we considered the actual category of the word within the (most frequent) frame. This analysis resulted in accuracy scores (.76 on average) that were higher than in the standard analysis, but still lower than those attained by the probabilistic measures.

A comparison of the probabilistic measures also reveals differences. Spearman rank-order and square root cosine similarity classify a similar number of items at similar levels of accuracy. Raw cosine similarity on the other

<sup>1</sup> The actual threshold levels were chosen to result in accuracy levels close to 90% on the basis of a pilot study and were set at .95 for raw cosine similarity, .70 for sqrt cosine similarity, and .50 for rank order correlation. The same thresholds were used throughout the analyses reported in this paper. Pilot work furthermore suggested that the accuracy of the different measures was similarly affected by proportional threshold variations.

hand only classifies around a third of the number of items that the other probabilistic measures classify.

Looking at the results for the merged corpus, it becomes apparent that all four measures classify a larger number of items. The three probabilistic measures however, do so with slightly higher accuracy than for the individual children, while the frequent frames measure shows a decrease in accuracy (66% vs. 55%, and 76% vs. 69% for the contextual score). Thus, it appears that the merged corpus contains additional information that can be successfully employed by the probabilistic measures but not the frequent frames measure. The increased information in the merged corpus is actually detrimental to the accuracy score for frequent frames. This latter finding appears to be caused by the fact that the frequent frames approach is overly sensitive to the occurrence of ‘stray’ words within the frequent frames. The fact that a word needs to occur only once within a specific frame to be clustered with all other words within that frame means that infrequent words that are atypical of a particular frame can potentially exert undue influence on overall accuracy scores. This problem becomes more pronounced in larger corpora. Such effects are less of a problem for the probabilistic measures.

### The noun-verb asymmetry

It was argued earlier that children are more willing to use novel nouns in known contexts than they are to use verbs. This finding has been taken as evidence that children develop a productive noun category earlier than they develop a verb category. In this section, we examine to what extent the different metrics show a preference for the clustering of nouns and verbs. This was done by examining the ‘hits’ from the data in Table 1, and counting the number of noun-noun and verb-verb pairs. The resulting data (proportion of noun-noun pairs relative to noun-noun + verb-verb pairs) are displayed in Table 2.

It is clear from Table 2 that the measures that are most frequency sensitive cluster the highest proportion of nouns. Thus, cosine similarity based on raw frequencies clusters a relatively low number of items but these items consist almost exclusively of nouns. Square Root Cosine similarity and rank order correlation are equally productive in terms of the number of items they classify, with the more frequency sensitive Cosine Similarity linking more nouns. Frequent frames on the other hand overwhelmingly link verbs. It is also apparent from Table 2 that, for frames, there is considerable variation in the proportion of noun-noun pairings: Aran’s proportion is highest at 55%, but half the children show a proportion of noun-noun pairings under 5%.

Table 2: Proportion of noun-noun pairings relative to noun-noun plus verb-verb pairings (total N in parentheses), excluding utterance boundaries.

| NV-ratio | Cos-Raw       | Cos-Sqrt      | Spearman      | Frames         |
|----------|---------------|---------------|---------------|----------------|
| Anne     | .99<br>(1387) | .93<br>(2084) | .63<br>(1068) | .34<br>(5836)  |
| Aran     | .99<br>(560)  | .84<br>(1784) | .67<br>(1440) | .55<br>(15222) |
| Becky    | .99<br>(461)  | .91<br>(1349) | .66<br>(1421) | .03<br>(3685)  |
| Carl     | .99<br>(507)  | .89<br>(1885) | .72<br>(1652) | .01<br>(3964)  |
| Dominic  | .97<br>(316)  | .83<br>(1040) | .48<br>(937)  | .02<br>(4499)  |
| Gail     | .96<br>(284)  | .89<br>(975)  | .67<br>(888)  | .02<br>(4452)  |
| Joel     | .99<br>(237)  | .85<br>(831)  | .63<br>(853)  | .02<br>(3828)  |
| John     | .99<br>(246)  | .90<br>(1566) | .79<br>(2104) | .40<br>(6378)  |
| Liz      | .97<br>(250)  | .91<br>(894)  | .80<br>(1169) | .10<br>(3214)  |
| Nicola   | .96<br>(275)  | .74<br>(907)  | .56<br>(1049) | .16<br>(6056)  |
| Ruth     | .99<br>(244)  | .633<br>(746) | .85<br>(790)  | .04<br>(4645)  |
| Warren   | .98<br>(311)  | .76<br>(904)  | .47<br>(927)  | .52<br>(9565)  |
|          |               |               |               |                |
| Average  | .98<br>(432)  | .87<br>(1247) | .66<br>(1191) | .28<br>(5945)  |
|          |               |               |               |                |
| Merged   | .99<br>(1426) | .80<br>(8577) | .67<br>(7319) | .57<br>(63747) |

### The role of utterance boundaries

The analyses reported in Table 1 and 2 only considered ‘lexical contexts’. That is, only words were considered as context items. The following set of analyses included the beginnings and ends of utterances as context items. Redington et al. (1998) do consider utterance boundaries as context items in one of their analyses (and conclude that they are potentially useful), but the non-parametric nature of their distance metric (rank order correlations) may underestimate the potential utility of utterance boundaries.

Mintz (2003) does not consider utterance boundaries, and it could be argued that there is little reason to consider them. Intuitively, the appealing feature of frames is that (because of their lexical nature) they are highly constraining and hence likely to result in relatively high accuracy. Allowing utterance boundaries in frames limits their constraining nature and may thus reduce accuracy levels. At the same time, however, frames including

utterance boundaries have the potential to capture large numbers of nouns (e.g. The X end\$) and thus might serve to counteract the verb bias apparent in Table 2. Table 3 shows the accuracy scores and number of word pairings for the analysis that include the utterance boundary as a framing element. The proportion of noun-noun pairings relative to noun-noun plus verb-verb pairings are shown in Table 4. For reasons of brevity, Tables 3 and 4 do not present data for the individual children in the Manchester corpus, but only the average and merged data across the 12 children.

Table 3: Proportion correct and number of word pairings including utterance boundaries.

|         | Cos-raw       | Cos-Sqrt       | Spearman       | Frames          |
|---------|---------------|----------------|----------------|-----------------|
| Average | .82<br>(862)  | .90<br>(5739)  | .90<br>(3163)  | .49<br>(81351)  |
| Merged  | .85<br>(4098) | .91<br>(23271) | .92<br>(11271) | .28<br>(316651) |

It is evident from Table 3 that the probabilistic measures deal well with the addition of the utterance boundary as a framing element. Accuracy levels are comparable to those shown in Table 1, while the number of word pairings has increased by a factor 2 to 3. Inspection of Table 4 furthermore indicates that the increase in completeness is largely the result of increased linking of noun pairs. This is particularly noticeable in the square root cosine similarity model, which links twice as many words as the rank order model. Thus, the average square root cosine similarity model links over 4700 noun-noun pairs, compared to ~1900 for the rank order model. This difference reflects the greater frequency sensitivity of the cosine model, and suggests that square root of raw frequency represents an optimum level of frequency sensitivity.

Table 4: Number of noun-noun pairings relative to noun-noun plus verb-verb pairings, including utterance boundaries.

| NV-ratio | Cos-raw       | Cos-Sqrt       | Spearman      | Frames         |
|----------|---------------|----------------|---------------|----------------|
| Average  | .99<br>(677)  | .95<br>(4957)  | .77<br>(2501) | .83<br>(38097) |
| Merged   | .99<br>(3256) | .90<br>(19924) | .67<br>(9558) | .72<br>(82924) |

## Conclusions

The main aim of this paper was to compare a mechanism for extracting syntactic categories based on independent contexts with Mintz style frequent frames in terms of their accuracy and ability to cluster nouns and verbs. A secondary aim was to investigate the role of frequency sensitivity and availability of utterance boundaries as framing elements.

The analyses presented here suggest that independent contexts result in better predictions than frequent frames. Frequent frames classify a larger number of words, but do so with lower overall accuracy.

Apart from being more accurate, the mechanisms based on independent contexts also cluster more nouns than verbs. This appears to be consistent with the suggestion that children form a productive noun category earlier than they form a verb category. The reverse is true of frequent frames: across the corpora frequent frames overwhelmingly cluster verbs rather than nouns, with noun-noun pairings making up under 5% of pairings for half the corpora.

When considering large input sets (i.e. the merged Manchester corpus), it becomes obvious that the mechanisms employing independent contexts are able to utilize the additional information contained in larger data sets to classify a larger number of items with similar levels of accuracy. The frequent frames mechanism also classifies a larger number of items when employed on a larger data set, but does so with lower accuracy. This result suggests that one of the strengths of frequent frames — its ability to quickly categorize a relatively large number of items on the basis of limited data — becomes a weakness when faced with larger datasets. Thus, the fact that the approach does not consider the frequency with which items occur in the target frames, results in it being relatively brittle and sensitive to noise and infrequent items in the input.

As an illustration, consider the frame *You X to*, which is the most frequent frame for the corpus of Carl as well as the merged Manchester corpus. Within Carl's corpus this frame contains 22 unique words, of which 20 (or 91%) are verbs. Across the Manchester corpus, the same frame contains a total of 89 unique words of which 67 (or 75%) are verbs. This increase in non-verbs has a marked impact on the accuracy for the frame which drops from .74 to .49. Many of the non-verbs occurring in the frame are legitimate (but infrequent) fillers for the frame (e.g. *back, off, down, happy, ready, anything, something, just, how, over, not, one*), while others are slightly more exotic: *that* (from *we brought you that to help you*), *tomorrow* (from *who's taking you tomorrow to playgroup*) to somewhat bizarre: *card* (from *a thank you card to give you*).

Since noisiness is an inherent property of the speech signal, which contains frequent repetitions, retracings and restarts, this finding suggests that frequent frames may not be a suitable source of information for category extraction unless combined with some sort of 'clean-up mechanism' or probabilistic element. Such an addition, however, would considerably weaken the great strength of this approach: its ability to rapidly classify items on the basis of little information.

All three mechanisms that computed statistics over independent contexts were able to attain higher accuracy levels, though they classified fewer items. It was argued that the probabilistic nature of these mechanisms allows

them to utilize the additional information in larger corpora without suffering from the brittleness associated with frequent frames.

A similar pattern emerges when including utterance boundaries as context items. For independent contexts, the inclusion of utterance boundaries results in comparable levels of accuracy, coupled with greater levels of completeness. The inclusion of utterance boundaries in frequent frames results in a drop in performance, in particular when considering the merged corpus. Such a finding may not be surprising (and indeed may not be in the 'spirit' of frequent frames), since the inclusion of utterance boundaries leads to the measure being less constrained than the lexical frequent frames that were originally proposed by Mintz. It does, however, provide an additional indication that independent contexts are less brittle and better able to incorporate additional, potentially noisy information.

The increased flexibility of independent contexts is further underscored by the analyses relating to the noun-verb asymmetry. Empirical work has suggested that children are more likely to substitute novel nouns in familiar contexts than they are to substitute novel verbs. (Akhtar & Tomasello, 1997; Tomasello, 2000). If the number of classified nouns versus verbs is an indication of such a tendency, frequent frames would appear to result in levels of verb-richness that are overly high. Thus, when excluding utterance boundaries, noun pairs make up on average only 28% of noun and verb pairs for frequent frames, compared to approximately 80% for independent contexts. While the inclusion of utterance boundaries leads to higher levels of noun pairs in the frequent frames approach, the results from Table 2 suggest that this is achieved at the expense of accuracy. For independent contexts, accuracy and noun richness remain high, whilst completeness is improved relative to the condition without utterance boundaries.

Some differences were also apparent within the different implementations of independent contexts. The present paper compared three different measures that differed primarily in terms of the weighting of frequency information. Within the constraints employed (which included a threshold value that results in an average accuracy level of approximately 90%), it was apparent that a similarity measure based on raw frequency counts results in relatively low completeness, while completeness for square root cosine similarity and rank order correlations perform at similar levels of completeness, with the square root cosine similarity measure showing more of a noun advantage than the less frequency sensitive rank order correlation. Overall, the inclusion of utterance boundaries leads to higher levels of noun-richness, suggesting it is a useful source of information. The size of this effect of noun-richness however was dependent upon frequency sensitivity: while noun-richness increased for all probabilistic measures, this was most pronounced for the square root cosine

similarity, suggesting that this represents an optimum level of frequency sensitivity.

## Acknowledgements

This research was supported by the Economic and Social Research Council under grant number ES/J011436/1.

## References

- Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology*, 33, 952-965.
- Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using 'frequent frames': What cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science*, 12, 396-406.
- Erkelens, M. A. (2009). *Learning to categorize verbs and nouns*. Unpublished PhD Thesis, Universiteit van Amsterdam, Amsterdam.
- Finch, S. P. & Chater, N. (1992). Bootstrapping syntactic categories. *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 820-825).
- Freudenthal, D., Pine, J. & Gobet, F. (2008). On the utility of conjoint and compositional frames and utterance boundaries as predictors of word categories. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1947-1952). Austin, TX: Cognitive Science Society.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analysing talk (3<sup>rd</sup> Edition)*. Mahwah, NJ: Erlbaum.
- Monaghan, P. & Christiansen, M. (2004). What information is useful and usable in language acquisition? *Proceedings of the 26<sup>th</sup> Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.
- 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 structures.
- 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.
- 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.
- 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.