Frequent Frames as Cues to Part-of-Speech in Dutch: Why Filler Frequency Matters

Richard Eduard Leibbrandt (Richard.Leibbrandt@flinders.edu.au)

David Martin Ward Powers (David.Powers@flinders.edu.au)
School of Computer Science, Engineering and Mathematics,
Flinders University, Adelaide, Australia

Abstract

The Frequent Frames model (Mintz, 2003) attempts to assign words to word categories based on their distributional patterns of usage. This model is highly successful in categorizing words in child-directed speech in English, but has been shown by Erkelens (2008) to be less effective with Dutch material. We show that extending the amount of contextual information in a frame by making use of the full utterance context does not improve categorization performance, but that constraining the fillers of Frequent Frames to be relatively less frequently occurring words does improve categorization significantly. We connect the latter result to a basic dichotomy in some languages between function words and content words, and conclude that, at least for English and Dutch, paying attention to this dichotomy is of greater importance for distributional bootstrapping proposals than the specific distributional contexts that are used to categorize words.

Keywords: Language learning; Distributional bootstrapping; Parts-of-speech; Function words; Frequent frames

Introduction

The parts-of-speech of a language (word classes such as nouns, verbs and adjectives) are of crucial importance in describing the grammar of the language. A vast amount of research has aimed to delineate the processes by which children learn to categorize words into the parts-of-speech of their native language. Researchers favouring *semantic bootstrapping* approaches (Grimshaw, 1981; Pinker, 1984) have proposed that early word categories are formed by grouping together words that refer to the same dimensions of concrete meaning, such as actions or objects. On the other hand, following early proposals by Maratsos & Chalkley (1980), proponents of *distributional bootstrapping* have argued that word categories can be induced by observing that certain groups of words are used in similar linguistic contexts, whether these contexts are defined at the level of words, morphemes, or even phonological or prosodic phenomena.

In recent years, it has become feasible to implement specific distributional bootstrapping proposals as computer algorithms that attempt to categorize words purely by analysing distributional patterns in large corpora of natural utterances (Cartwright & Brent, 1997; Redington, Chater & Finch, 1998). For instance, Redington et al. (1998) found that words in child-directed English speech could be categorized with a high level of success by considering only very local utterance contexts made up of words that occur in close proximity to the target word.

A particularly successful distributional model has been the *Frequent Frames* model of Mintz (2003, 2006a, 2006b). Frequent frames are defined as a disjunct frame occurring around a target word, made up of the word immediately preceding and the word immediately following the target, so that all frequent frames have the form *a _ b*, with *a* and *b* standing for specific words, and the underscore representing a slot that can accept a variety of filler words. For example, in the three-word sequence "a house and", the frame is "a _ and", and the filler is "house". Once all frames of this form have been collected from a corpus, only the most *frequent* ones are retained for the purpose of categorization. This reflects the intuition that, if two words co-occur frequently on either side of another word across several utterances, this is likely to be due to some meaningful linguistic relationship between them. All words occurring in the same frequent frame are assigned to the same category, and frames that have more than 20% overlap in their set of slot fillers have their categories amalgamated into larger, more general categories. This amalgamation step is crucially important: by grouping together frames that accept similar sets of words, the child may be able to hypothesize that a word used in one verb frame may also be legitimately used in another verb frame; without amalgamation, this kind of generalization is not possible.

Frequent Frames provide a very successful categorization of the words that occur in them, with Mintz (2003) reporting values greater than 0.9 for the evaluation measures *accuracy* and *completeness* when the model was implemented on a set of English corpora. However, recently Erkelens (2008) has shown that, in the case of child-directed speech in Dutch, Frequent Frames provide a less accurate basis for part-of-speech categorization than they do for English: whereas the use of Frequent Frames in English yielded an accuracy figure that exceeded the random baseline by 0.52 for tokens and 0.46 for types, a replication with a Dutch corpus could attain an improvement in accuracy over baseline of only 0.33 for tokens and 0.25 for types.

Full-Utterance Frames As Distributional Contexts

An important issue in distributional bootstrapping is to decide on the most appropriate usage contexts to consider for the purpose of categorization. One possible reason for the purported lower utility of Frequent Frames in Dutch

may be that Dutch simply allows a greater amount of flexibility in the range of structures in which particular words are able to occur. This raises the possibility that the contextual window employed by Frequent Frames may simply have been too small, and that it may be necessary to consider a wider amount of lexical context around a word in order to distinguish between different constructions. The maximum amount of context for a word used in an utterance is arguably the entire utterance, and so from a practical point of view it may be useful to explore the use of frames that comprise a full utterance at a time.

Tomasello (2006) has suggested a prominent role in language development for *utterance-level constructions*, expressions that can be used as complete utterances and are associated in a routinized way with certain communicative functions. Pine & Lieven (1993) provide evidence that some children assemble their earliest multi-word utterances by starting with "frozen", unanalysed phrases and proceeding to analyse these into fixed parts with variable slots into which various elements can ultimately be inserted (although some children instead form multi-word utterances by combining familiar single words together).

Given both these pragmatic and theoretical considerations, Leibbrandt and Powers (2008) evaluated a distributional bootstrapping proposal that makes use of schematic representations of complete utterances, with most words in the utterance lexically specified and one or two additional word positions serving as slots, for example "Are you going to X it?" or "That's the X". Under Leibbrandt & Powers's proposal, words that occur in the same full-utterance frame slot are categorized as belonging to the same word category. This approach was highly effective for categorizing word tokens in a natural corpus of child-directed English speech, attaining levels of correctness in part-of-speech classification that were comparable to those achieved by Frequent Frames (Mintz, 2003).

The Function Word - Content Word Dichotomy

Another important factor in distributional bootstrapping proposals is the basic dichotomy that exists in many languages between content word classes (the classes that carry lexical meaning, such as nouns, verbs and adjectives) and function word classes (the classes that are more closely involved with grammar, such as determiners, conjunctions and prepositions).

Attending to the positional relationships between function and content words has been proposed to be of importance to the language-learning child. For instance, Valian and Coulson (1988) found that learning an artificial language is made easier by increasing the frequency of function words that can serve as anchor points for distributional analysis, and suggest that children may seek out the most frequent elements in language in order to learn about the patterns in which parts-of-speech are allowed to occur in a language.

Gerken, Landau & Remez (1990) point out that function words could be crucial in the two tasks of *word segmentation* and *word labeling* (category assignment).

Function words are potentially useful in segmentation because recognizing the relatively small number of function words makes it easier to separate out the far more heterogeneous open-class words that are interspersed between them. Function words could also aid labeling, because they occur in very stereotypical positional relations to open-class words, for instance, "the" is often followed by a noun (or sometimes by an adjective which is followed by a noun), and "-ing" is usually preceded by a verb root.

Because it cannot be assumed that children know *a priori* which words are function words and which are content words, this distinction would have to be learned on the basis of perceptible cues in the language spoken to children. English function words can be identified by a number of phonological cues, including syllable complexity, stress and vowel quality (Morgan, Shi & Allopenna, 1996), and even newborn infants are able to distinguish English function words from content words (Shi, Werker & Morgan, 1999).

Another feature of the distributional approach of Leibbrandt and Powers (2008) that may have contributed to its successful categorization performance is that it attempts to take the function word - content word dichotomy into account by making use of another cue that may plausibly be available to children: most function word types occur more *frequently* in speech than most content word types. Leibbrandt and Powers attempted to approximate the distinction between function words and content words in English by sharply distinguishing between the two sets of frequent and less frequent words, defined as respectively the set of the top *N* most frequently-occurring words in a corpus, and the set of all other words. When creating full-utterance frames for their distributional analysis, they applied the constraint that only frequent words could be used as the lexically-specific words in a frame, and only less-frequent words could be used as the slot fillers that were embedded in the frames.

Adapting The Frequent Frames Model

Erkelens (2008) argues that different cues are useful to differing extents in different languages, and that the occurrence of a word in a frequent frame is not as useful a cue to part-of-speech for the Dutch-learning child as it is for English. While we agree with the former point, we will attempt to show that the utility of Frequent Frames for categorization in Dutch may have been underestimated.

In the remainder of this paper, we report on a series of four experiments intended to investigate whether the Frequent Frames model can be modified to deal successfully with Dutch material. In Experiment 1, we replicate the results of Erkelens (2008) with a larger corpus and Frequent Frame set, and confirm that the unmodified Frequent Frames model is less useful for categorization in Dutch than in English. In Experiment 2, we investigate whether the distributional model of Leibbrandt & Powers (2008) is able to improve over the categorization results of Frequent Frames (we preempt our results here by confirming that it does). As Leibbrandt and Powers's approach differs in two

ways from the Frequent Frames approach, it then becomes important to investigate whether this improved performance is due to both differences, or only one. In Experiment 3, therefore, we modify the Frequent Frames approach to use the complete utterance as the context for categorization, and in Experiment 4, we constrain Frequent Frames to be composed of only frequent (function) framing words, and to take only less-frequent (content) filler words.

Experiment 1

Method

The corpus used in these experiments was the Groningen corpus (Wijnen & Bol, 1993), taken from CHILDES (MacWhinney, 2000), and consisting of data from seven Dutch-learning children in the Groningen area, recorded between the ages of 1;05 and 3;07. The corpus was minimally preprocessed for computer-readability, and all sentences uttered by adults were used. Data from all seven children were merged together in order to increase the data set size for the purpose of data clustering.

Frequent Frames were extracted according to the method used by Mintz (2003). Candidate frames were extracted from each utterance in the corpus, by forming a frame from every three consecutive words in each utterance and replacing the middle word with a slot marker. The frames with the highest frequency of occurrence in the corpus were retained as the set of *frequent* frames. Frequency statistics were collected on how often each word occurred in the slot position of each frequent frame throughout the corpus.

The studies by Mintz (2003) and Erkelens (2008) made use of a set size of 45. Because it was desirable in the present work to apply clustering to the data, a slightly larger frame set of 250 frames was used. These top 250 frames were grouped into clusters of frames by means of average-linkage hierarchical clustering (Sokal & Sneath, 1963), with the initial distances between frames given by Spearman's ranked correlation coefficient. Frames were clustered together if they occurred in the corpus with similar sets of slot-filler words.

Clustering makes it possible to make generalizations about the acceptability of words in frames in which they have not been attested in the corpus. If the clustering algorithm produces K clusters of frames, these clusters correspond to K hypothesized categories. Any word token which occurs in any frame belonging to a particular cluster is then assigned to the category corresponding to that cluster.

Evaluation Measures And Significance

All the experiments reported here involve the task of categorizing words into word categories based on the context in which they are used. In each case, there is an empirical allocation of words to unlabelled categories, which needs to be evaluated by a comparison with the "true" distribution of word tokens into their parts-of-speech. This "true" distribution was created by us after manually inspecting each of the particular word tokens in contextual

usage. We made use of the same categories that were used by Erkelens (2008) in her "standard analysis", namely: verbs (including auxiliaries and copula), nominals (nouns, proper names and pronouns), adjectives, prepositions, adverbs, determiners, WH-words, conjunctions and interjections.

Unsupervised categorization models such as Frequent Frames are usually evaluated by means of the mathematical measures *accuracy* and *completeness*, using a *pair counting* approach. A formal definition of these two measures falls outside the scope of this paper, but they can be intuitively understood as expressing the extent to which word tokens assigned to the same category by the model do in fact belong to the same part-of-speech, and the extent to which word tokens which belong to the same part-of-speech were in fact categorized together by the model, respectively.

It is possible to report accuracy and completeness both in terms of the number of word *tokens* correctly categorized and in terms of the number of word *types* correctly categorized; results reported here are based on word type categorization only.

It should be noted that one cannot simply compare accuracy and completeness scores between experiments that make use of different sets of data. Any comparison has at least to take into account the magnitude of the *difference* between accuracy (or completeness) attained by the model, and the baseline accuracy (or completeness) attained by randomly allocating of words to categories.

In order to address this difficulty, we make use of *permutation tests* to assess the significance of differences between evaluation measures, both within and between experiments. Within an experiment, it is possible to assess whether the value of an evaluation measure is significantly higher than the baseline value, by generating a randomized sample of values for that measure and determining how often an equal or higher value occurs in the sample. Between experiments, it is possible to determine whether an obtained value for an evaluation measure in one experiment is significantly better than a value for the same measure in another experiment, by generating a random sample for each experiment separately, taking the differences between pairs of values from the two samples, and comparing this sample of differences to the difference between the originally obtained values.

There is also typically a trade-off between accuracy and completeness, and it is possible to artificially inflate one measure at the expense of the other. For this reason, it is necessary to considered both values together when evaluating the results of an experiment. In addition, we will also report values for the *F measure*, calculated as the harmonic mean of accuracy and completeness, which summarizes both measures and takes on a high value only when they are both high in value.

Results

Firstly, an analysis of the 250 most frequent frames in the pooled corpus confirmed the conclusion drawn by Erkelens

(2008) that frequent frames are not as reliable a basis for the categorization of Dutch words as they are for English words. The categorization accuracy for the top 250 frames from the pooled corpus (displayed in Table 1) was 0.60, against a random baseline of 0.31, i.e. categorizing on the basis of frequent frames rather than by randomly assigning categories improved accuracy by only 0.29. This result is comparable with Erkelens's (2008) 0.25 increase in accuracy, suggesting that the differences between that study and the current experiment (a different corpus and a larger set of frequent frames) did not materially affect the results. As might be expected, completeness was almost equal to the random baseline (and near zero), as frame clusters have not yet been created. These results confirm that the individual frames have some utility in predicting the part-of-speech of their slot-filler words, but that their accuracy is far from perfect.

Table 1: Accuracy and completeness for the top 250 frequent frames before clustering, against results from Mintz (2003) and Erkelens (2008). Random baseline figures in italics.

Study	Language	Measure	Value
Mintz (2003)	English	Accuracy	0.93 *(0.47)*
Erkelens (2008)	Dutch	Accuracy	0.58 *(0.33)*
Experiment 1	Dutch	Accuracy	0.60 *(0.31)*
Experiment 1	Dutch	Completeness	0.01 *(0.01)*

Hierarchical clustering was applied to the frames based on the distributional patterns of their filler words[1]. The results are shown in Table 2. Accuracy decreased sharply, as should be expected, as assigning every frame to its own unique category corresponds to the maximum attainable accuracy value. While completeness increased by a large amount in absolute value, it did not exhibit a large advantage over the random baseline completeness value.

Table 2: Evaluation of word token categorization after hierarchical clustering of the top 250 frequent frames into 12 clusters. Random baseline figures in italics.

Accuracy	0.429 *(0.327)*
Completeness	0.405 *(0.308)*
F	0.417 *(0.317)*

Discussion

These results replicate the findings of Erkelens (2008) that frequent frames have some utility as a basis for the prediction of the part-of-speech of Dutch words, but that they are not nearly as reliable as they are in English.

---

[1] The number of clusters produced by hierarchical clustering affects the obtained results. Procedures exist for choosing an optimal number of clusters, but for the sake of consistent comparison across the four experiments described here, the number of clusters produced was fixed at 12 in each experiment.

Experiment 2

Method

As in Experiment 1, we made use of the Groningen corpus. Here, however, we attempted to apply the lexically-specific frame approach proposed by Leibbrandt and Powers (2008). A list was compiled of the most frequently occurring word types in the Groningen corpus. This requires a choice of an arbitrary frequency cutoff point, and in this experiment the top 300 most frequent words were selected as the frame-building words. This set included the most common function words in Dutch, including pronouns (*ik*, *hij*, *ze*), determiners (*een*, *de*, *het*, *deze*, *dat*), and forms of the copula (*ben*, *zijn*) as well as a number of common content words.

All utterances were rewritten as lexically-specific frame candidates, by replacing every word that was not on the frequent-word list by a placeholder symbol X. From this set of candidate *dichotomous full-utterance frames*, the 250 frames with the highest frequency of occurrence were retained for analysis.

As in Experiment 1, co-occurrence data was collected about the frequency with which different words occurred in each of the frames, and the set of frames was clustered based on similarity in their sets of filler words. Note that, because of the way in which the frames were constructed, all slot fillers were taken from the set of less-frequent words.

Results

A number of intuitively sensible Dutch full-utterance frames were produced by this process, for example "Daar is de X" ("There's the X"), "Gaat ie X?" ("Is he going to X?") and "Heel X" ("Very X"), frames which could reasonably be expected to take noun, verb and adjective fillers respectively. Note that none of these example utterance structures could have been covered by the Frequent Frames approach, as the slot word occurs at the end of the frame in each case.

Table 3: Evaluation of word token categorization after hierarchical clustering of the top 250 dichotomous full-utterance frames. Random baseline figures in italics.

Accuracy	0.752 *(0.431)*
Completeness	0.407 *(0.233)*
F	0.528 *(0.302)*

The results of categorization evaluation after clustering are shown in Table 3. Accuracy, completeness and F were all significantly higher than baseline, as assessed by a permutation test ($p < 0.01$). Furthermore, categorization performance was significantly better than in the Frequent Frames approach of Experiment 1, as assessed by a permutation test of F value differences ($p < 0.01$).

Discussion

This experiment has shown that the frame selection approach used by Leibbrandt and Powers (2008) produces frames that are far more reliable indicators of the part-of-speech of a word in Dutch than the standard Frequent Frames proposed by Mintz (2003). As stated, this approach differs from Frequent Frames in two ways, making use of full-utterance contexts and accepting only less-frequent word (content word) fillers. In the next two experiments, we attempt to determine whether the improved performance shown here is due to one, or both, of these properties.

Experiment 3

Method

From each utterance in the corpus, candidate frames were extracted that contained all the words in the utterance except for one target word, which was turned into a variable slot (so that each utterance yielded as many candidate frames as there were words in the utterance). For example, the utterance "Dat is het vliegtuig" yielded the frames "X is het vliegtuig", "Dat X het vliegtuig", "Dat is X vliegtuig" and "Dat is het X". The most frequently occurring of these candidate *frequent full-utterance frames* were selected for evaluation. As before, word occurrence frequencies were calculated for each frame, and frames were clustered together based on the patterns of words that occurred in their slots.

Results

Evaluation results are shown in Table 4. While all evaluation measures were significantly greater than baseline ($p < 0.01$), the full-utterance frames did not provide a better basis for categorization than Frequent Frames; on the contrary, the categorization using Frequent Frames in Experiment 1 performed significantly *better* than the categorization with full-utterance frames, as assessed on a permutation test of differences in F measures ($p < 0.01$).

Table 4: Evaluation of word token categorization after hierarchical clustering of the top 250 frequent full-utterance frames. Random baseline figures in italics.

Accuracy	0.428 *(0.267)*
Completeness	0.249 *(0.156)*
F	0.315 *(0.197)*

Discussion

Clearly, the relatively low utility of Frequent Frames in categorizing Dutch words shown in Experiment 1 was not merely due to an insufficient amount of contextual information. In this experiment, increasing context to comprise the whole utterance did not improve categorization, as one might have expected from the superior results with dichotomous full-utterance frames in

Experiment 2, and it seems that the success of those frames had nothing to do with their being based on full utterances.

Experiment 4

Method

Candidate frames were extracted from the corpus in the same way as for Experiment 1, i.e. the candidate frames were all Frequent Frames. However, following the approach taken in Experiment 2, we retained frames for the final evaluation set only if both the frame-building words (i.e. the first and third words) occurred in the list of the most frequent words in the corpus. Equally importantly, only words that were not in the frequent-word list were accepted as slot fillers for the frames. The most frequent such *frequent dichotomous frames* were selected and clustered as in the previous experiments.

Results

Evaluation results are shown in Table 5. It is immediately noticeable that the values of all measures are higher than for any of the other 3 experiments. For instance, comparison with Table 1 reveals that accuracy in this experiment is similar to the level of accuracy attained by Mintz (2003) with English frames. All measures are significantly above their baseline ($p < 0.01$), and categorization performance is significantly greater than for Experiments 1 and 3 ($p < 0.01$), but not significantly different from Experiment 2.

Table 5: Evaluation of word token categorization after hierarchical clustering of the top 250 frequent dichotomous frames. Random baseline figures in italics.

Accuracy	0.921 *(0.611)*
Completeness	0.513 *(0.340)*
F	0.659 *(0.437)*

Discussion

In this experiment, we have seen evidence that, contrary to the results of Erkelens (2008), Frequent Frames may yield high accuracy and completeness in categorizing Dutch words, *provided* that the frames are composed of the set of frequent words in Dutch (usually function words) while the categorization targets are taken from the relatively less frequent words, i.e. essentially content words.

Every one of the 250 frames in Experiment 1 was already composed of two frequent words. However, only 105 of those frames were retained in the evaluation set for Experiment 4. Therefore, the large improvement in categorization performance was due to the requirement that fillers should be less-frequent words, thereby effectively dropping function word fillers from the categorization. This seemed to result in a great number of frames being added to the evaluation set that were strongly associated with only one content word class (verbs, nouns, or adjectives).

General Discussion

The experimental results in this paper show that a simple distributional approach is effective in categorizing words in Dutch child-directed speech. Both the dichotomous full-utterance frame approach of Leibbrandt & Powers (2008) used in Experiment 2 and the Frequent Dichotomous Frames approach of Experiment 4 yielded significantly better categorizations than either Frequent Frames or Frequent Full-Utterance Frames, with no significant difference between the two models relative to their random baselines (although Frequent Dichotomous Frames achieved a higher F value in absolute terms, and so may arguably be preferred). By contrast, extending the context used for distributional analysis to a full utterance as in Experiment 3, paradoxically decreased performance. This suggests that the extent of context used in distributional bootstrapping may be less crucial than the kinds of words used for frames and fillers respectively

A simple modification to the Frequent Frames model is therefore able to overcome the shortcomings with Dutch material identified by Erkelens (2008). We suggest that the reason why the Frequent Dichotomous Frames model yields such a successful categorization is that the less-frequent words being categorized are mostly content words. In other words, it may be the case that the only useful targets for distributional analysis are the content word classes such as noun, verb, adjective and adverb. While pronouns, auxiliary verbs, etc. can also be identified by their distribution, these words may simply be learned on a one-by-one basis.

The weaker results of the original Frequent Frames model may have been due to a conflation of legitimate content word contexts with other cases where a function word in the frame slot indicates a different linguistic construction. For instance, the Frequent Frame "die _ wel" accepts a variety of verbs, and the pronoun "die" serves as the subject of the verb. When the slot is filled by the adverb "ook" to form "die ook wel", however, this is a different construction where "die" is the object of a verb outside the frame, or else the subject of an unstated verb. Eliminating function-word fillers avoids the conflation of contexts.

While the results from this corpus analysis speak less directly to how children *actually* learn parts-of-speech than the results of an experimental study would, they demonstrate the feasibility of exploiting a particular form of information in child-directed language. In concurrence with the proposal by Valian & Coulson (1988) that function words serve as anchor points indicating the structure of an utterance and facilitating distributional analysis, we suggest that it would be useful for children to make a distinction between function and content words, based on various cues such as phonology, greater occurrence frequency, etc. When children encounter a function word occurring in the slot position of what would normally be a distributional frame, they would then be able to avoid carrying out the normal process of categorizing the word on the basis of the frame, and to treat the function word as part of the structural information in the utterance only.

References

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition, 63*, 121-170.

Erkelens, M. (2008). Restrictions of frequent frames as cues to categories: the case of Dutch. *Supplement to the Proceedings of the 32nd Boston University Conference on Language Development (BUCLD 32).*

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology, 26*(2), 204-216.

Leibbrandt, R. E., & Powers, D. M. W. (2008). Grammatical category induction using lexically-based templates. *Supplement to the Proceedings of the 32nd Boston University Conference on Language Development (BUCLD 32).*

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* (3 ed. Vol. 2: The database). Mahwah, NJ: Lawrence Erlbaum.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91-117.

Mintz, T. H. (2006a). Finding the verbs: Distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs*. Oxford: Oxford University Press.

Mintz, T. H. (2006b). Frequent frames: Simple co-occurrence constructions and their links to linguistic structure. In E. V. Clark & B. F. Kelly (Eds.), *Constructions in acquisition*. Stanford: CSLI Publications.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 263-283). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pine, J. M., & Lieven, E. (1993). Reanalysing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language, 20*, 551-571.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*(4), 425-469.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*, B11-B21.

Tomasello, M. (2006). Acquiring linguistic constructions. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology*. New York: Wiley.

Valian, V. & Coulson, S. (1988). Anchor points in language learning: the role of marker frequency. *Journal of Memory and Language*, *27*, 71-86.

Wijnen, F. & Bol, G. (1993). The escape from the optional infinitive stage. In A. de Boer, J. de Jong & R. Landeweerd (Eds.) *Language and Cognition 3*, University of Groningen, Dept. of Linguistics.