Restrictions of frequent frames as cues to categories: the case of Dutch

Marian A. Erkelens
Amsterdam Center for Language and Communication, Universiteit van Amsterdam

The statistical distribution of words in the speech stream provides cues for the lexical category of those words. Syntactic rules of a language determine this distribution of words across sentences; the fact that English determiners have to precede nouns in a noun phrase (syntactic rule) results in the statistical fact that the English determiner *the* is followed more often than by chance by a word that is categorized as a noun in English. As a consequence, the co-occurrence patterns contain information about the syntactic status of words, for example their grammatical category (Mintz, 2003; Monaghan, Chater, & Christiansen, 2005; Redington, Chater, & Finch, 1998). Whereas specific grammatical information probably is inaccessible for young infants in an early stage of language learning, statistical information is accessible from very early on (Gerken, Wilson, & Lewis, 2005; Gomez & Gerken, 1999). Using the statistical distribution of words in the speech stream is therefore an attractive strategy to start categorizing the lexicon.

Mintz (2003) investigated the role of distributional cues in the categorization of verbs and nouns by English infants. Particularly interesting about his work is the fact that he proposes a specific and testable definition of the kind of statistical information children use to categorize. Earlier proposals about the role of distributional information in category learning (e.g., Maratsos & Chalkley, 1980) were rightly criticized for not being precise enough to function as a learning mechanism. Inspired by the results of perception studies showing that very young infants are sensitive to dependencies between frequently co-occurring words (Gomez, 2002; Santelmann & Jusczyk, 1998), Mintz (2003) proposed that frequently co-occurring words with exactly one word intervening (‘frequent frames’) are reliable predictors of the category of the intervening words within such a frame. For example, an English frequent frame is *you X it*, and the words that can occur in *X* (i.e., the frame-based category) are practically only verbs, so this frequent frame should be a reliable predictor of the verb-hood of intervening words. Furthermore, he conducted a perception experiment with 12-month-old infants acquiring English to show that these infants indeed use at least the verbal frequent frames to categorize novel words presented to them (Mintz, 2006). Frequent frames in English are not only available in the input but also result in accurate categories and are used by 12-month-olds to categorize; therefore, Mintz’s proposal that frequent frames play an important role in categorization is very promising in the search for a description of the acquisition of categories.

The aim of the current paper is to investigate whether the availability, accuracy, and use of frequent frames for lexical categorization extends to languages other than English. After all, a learning mechanism should work for every language-learning child.

The present paper reports on a replicative study for Dutch*. It will be shown that Dutch differs from English with respect to the accuracy of the frame-based categories in the input and the use of frequent frames by infants. The paper is organized as follows. The arguments of Mintz’s proposal will be summarized in Section 1. Section 2 then addresses the question whether frequent frames are available in the input to Dutch children and whether they result in accurate categories. Section 3 reports on the experiment that addresses the question whether Dutch children are sensitive to the frequent frames and are able to use them for categorizing novel words. In Section 4 the findings for Dutch will be compared to those for English, resulting in conclusions on the viability of Mintz’s proposal as a general learning mechanism for lexical categorization.

1. Mintz’s frequent frame proposal

I will first consider in more detail Mintz’s (2003) suggestion that the child only needs to focus on the local context of a lexical item to arrive at correct categorization. Since I have replicated both Mintz’s input study and his perception experiment for Dutch, the reader can obtain the details of the methods from Sections 2 and 3; here we will focus on the results for English.

The crucial local context under consideration is that of frequent frames. A frequent frame is defined as a frequently occurring combination of two words with exactly one word position (X) intervening, e.g., *you* X *it* (X = *have, like, show,* etc.). Mintz shows that the set of words X $\{X_1, X_2, \ldots, X_n\}$ within a certain frequent frame forms a category close to the categories in adult English grammar. He took the input from six English child language corpora from CHILDES. All frames (i.e., instances of *a* X *b* where X is variable) in the input speech were counted and tallied for number of occurrences. The items occurring in the 45 most frequent frames of each corpus were assigned a category label (verb, auxiliary, copula, noun, pronoun, adjective, preposition, adverb, determiner, wh-word, “not”, conjunction, interjection). The labels within one such frame-based category were compared to calculate the accuracy. Two different kinds of analyses were performed: a standard and an

* This study forms part of a larger project that will be reported in Erkelens (in progress).

expanded analysis. Under the expanded analysis, all 13 assigned category labels were taken as separate categories whereas under the standard analysis verbs, auxiliaries and copulas were one category as were nouns and pronouns, leaving 10 categories. The exact procedure of accuracy calculation is further exemplified in Section 2 on Dutch. The mean token accuracy across all English corpora was 0.98 under the standard analysis and 0.91 under the expanded analysis. Mean type accuracy was 0.93 in the standard analysis and 0.91 in the expanded analysis. The accuracy of randomly compiled categories was taken as a baseline measure. All resulting accuracy measures of the English input analyzed by Mintz (2003) are represented in Table 1. The accuracy scores of the frame-based categories are significantly higher than baseline in all conditions. Mintz concludes that frequent frames are present in the input to English children and that they form highly reliable cues to categories.

Table 1. Token and type accuracy in standard and expanded analysis including baseline accuracy of random categories for all English corpora (scores taken from Mintz (2003), p. 100).

	Standard Analysis				**Expanded Analysis**			
	Tokens		Types		Tokens		Types	
Corpus	A*	R**	A	R	A	R	A	R
Peter	0.98	0.49	0.96	0.55	0.97	0.32	0.95	0.49
Eve	0.98	0.51	0.92	0.50	0.91	0.25	0.89	0.40
Nina	0.98	0.48	0.95	0.46	0.98	0.29	0.94	0.36
Naomi	0.97	0.48	0.94	0.49	0.96	0.30	0.93	0.41
Anne	0.98	0.37	0.94	0.41	0.84	0.24	0.90	0.31
Aran	0.97	0.44	0.89	0.42	0.80	0.23	0.87	0.33
Mean	0.98	0.46	0.93	0.47	0.91	0.27	0.91	0.38

* A = Numbers for analysis indicated in upper cell
** R = Numbers for randomly compiled categories

To test whether English infants indeed use the reliable frequent frame information available in the input, Mintz (2006) conducted a perception experiment. In a training phase, 24 English-learning infants of 12 months learned four nonsense words: *deeg, lonk, gorp,* and *bist*. They heard two of the nonsense words in four frequent frames that host verbs (*you_the, to_it, I_you, can_#*). The other two nonsense words were presented in four frequent frames that host nouns (*the_in, your_#, his_on, a_of*). In the test phase infants heard the same nonsense words in either the similar category frames (i.e., a nonsense word trained as a verb inserted into a verb frame) or in a different, inconsistent frame (i.e., a nonsense word trained as a verb inserted into a noun frame). Mintz compared the listening times between the consistent and inconsistent sentences. The 12-month-old infants listened longer to the inconsistent verb frame sentences than to the consistent verb frame sentences. This is interpreted as indicating that they were sensitive to the inconsistent pairing of nonsense nouns with verbal frequent frames. For the nominal sentences he found no differences in listening times. The fact that infants showed different listening times between the consistent and inconsistent verb-frame sentences indicates that they did use the distributional contexts (i.e., frequent frames) in the training phase to categorize the new words. They had never heard the nonsense words being used in the test phase frames before, but still showed different listening times between the verbal and nominal test sentences, at least for the nonsense words modeled as verbs.

Mintz has shown that the frequent frames in the input to six English children result in accurate frame-based categories and that the verbal frequent frames are used by 12-month-olds to categorize nonsense words. To investigate whether these results can be extended to other languages than English, we will now turn to an analysis of Dutch. In Section 2 the input analysis will be repeated for Dutch; Section 3 presents the perception experiment with Dutch infants.

2. Frequent frames in Dutch input

To establish whether the local context of words like frequent frames are available in the input for languages other than English, I have performed an input analysis on Dutch child language corpora. In doing this, I followed the method from Mintz (2003) very closely. I analyzed the input speech received by four Dutch children younger than 2;6 available from CHILDES (Daan and Matthijs from the Groningen corpus (Krikhaar & Wijnen, 1996) and Sarah and Laura from the Van Kampen corpus (Van Kampen, 2004)) for the 45 most frequent frames. General information about these files is presented in Table 2. Although the total number of corpora and utterances examined by Mintz exceeds the numbers in Table 2, the results remain comparable; Mintz (2003) found similar frequent frames and accuracy numbers for each of the six child corpora he examined and the number of utterances in the first four corpora he examined is comparable to the four Dutch corpora examined here.

Table 2. Session ranges for analyzed Dutch corpora, number of utterances, number of tokens and types categorized.

Child	**CHILDES sessions**	**# of utterances**	**Tokens categorized**	**Types categorized**
Matthijs	mat11013-mat20619	16813	2927	319
Sarah	sarah01-sarah19	10710	2186	296
Daan	daa10821-daa20625	13301	2569	324
Laura	laura01-laura22	8811	1948	291
Mean		12409	2407	308

As with the English corpora, all frames in the input speech were counted for number of occurrences, and the 45 most frequent frames were selected for each corpus. The fourth and fifth column of table 2 indicate the number of tokens and types that occur in the 45 most frequent frames of the corpora. For these 45 most frequent frames, all intervening words were listed, resulting in frame-based categories. These frame-based categories were further analyzed to see how they relate to the 'real' categories of the words occurring in them. For Dutch, I used slightly different labels for 'real categories' than Mintz did for English. Instead of a separate category for "not" (the Dutch equivalent *niet* being categorized as an adverb), I included a separate category for proper names since those occurred very frequently within the Dutch frame-based categories. This resulted in the categories verb, auxiliary, copula, noun, pronoun, proper name, adjective, preposition, adverb, determiner, wh-word, conjunction, and interjection. All types in the frame-based categories were assigned one of those category labels. For words in Dutch that can occur in multiple categories, I checked the corpora to see in which category the word was used in the specific frame. To calculate the accuracy of the frame-based categories, the category labels of all possible pairs of items occurring in one frame were compared. Each comparison of two items can result in either a *hit* or a *false alarm*. A hit was counted whenever the two items had the same grammatical category label, and a false alarm whenever the two items had different labels. Accuracy is computed using the following formula:

$$\text{Accuracy} = \frac{\text{Hits}}{(\text{Hits} + \text{False Alarms})}$$

The total number of hits (i.e., instances of two similar category labels) in a frame-based category is divided by the number of hits plus false alarms (i.e., instances of two different category labels). If all types within a frame-based category were from the same 'real' category, the accuracy is 1 since then the number of false alarms is 0. Like Mintz, I performed a standard and an expanded analysis. In the expanded analysis all 13 assigned category labels were taken as separate categories whereas in the standard analysis verbs, auxiliaries and copulas were one category as were nouns, pronouns, and proper names, leaving 9 categories.

Results for the accuracy analyses in the Dutch corpora were quite different from those of the English corpora. The mean token accuracy across all corpora was 0.71 in the standard analysis and 0.56 in the expanded analysis. Mean type accuracy was 0.59 in the standard analysis and 0.40 in the expanded analysis. Similar to Mintz, I compiled random categories in which the tokens of the frame-based categories were randomly distributed among the 45 frequent frames. The accuracy measures of those random categories serve as a baseline for the informativeness of word distribution in Dutch. The Dutch results per corpus are presented in Table 3 and a comparison between the mean accuracy results of all six English and four Dutch corpora is presented in Table 4 below.

Statistical analysis of the accuracy measures of both English and Dutch can form the basis for predictions with respect to the informativeness of frequent frames in the two languages. As reported earlier by Mintz (2003), the English accuracy scores were significantly higher than baseline in both the standard analysis and in the expanded analysis. The same is true for the Dutch accuracy scores in both the standard analysis (tokens: $M = 0.38$, $t(3) = 24.6$, $p < .0001$; types: $M = 0.33$, $t(3) = 19.3$, $p < .0001$) and in the expanded analysis (tokens: $M = 0.16$, $t(3) = 30.6$, $p < .0001$; types: $M = 0.16$, $t(3) = 20.6$, $p < .0001$). The differences between tokens and types were significant in the standard analysis both for English ($t(5) = 5.8$, $p < .01$) and for Dutch ($t(3) = 19.5$, $p <. 0001$). In the expanded analysis the token-type difference is only significant for Dutch ($t(3) = 21.7$, $p < .0001$).

This means that for Dutch, the accuracy is higher when the frequency of the items is taken into account. The type-token ratio in Dutch is somewhat higher than in English, which causes a difference between the two languages. In the type analyses the frequency of the items themselves is not taken into account. Based on the fact that the accuracy measures of the frame-based categories in both English and Dutch are higher than baseline, we can make the following prediction for the use of frequent frames by Dutch infants:

(1) Similar to the English infants, Dutch infants will use the informative frequent frames in early categorization.

However, we also see that the accuracy scores for English are higher than those for Dutch. These differences between the two languages are significant in all possible conditions in the standard analysis (tokens analysis: $t(8) = 19.1$, $p < .01$; tokens random: $t(8) = 2.6$, $p < .05$; types analysis: $t(8) = 17.5$, $p < .0001$; types random: $t(8) = 4.6$, $p < .01$) and in the expanded analysis (tokens analysis: $t(8) = 8.4$, $p < .0001$; tokens random: $t(8) = 5.6$, $p < .01$; types analysis: $t(8) = 30.2$, $p < .0001$; types random: $t(8) = 6.7$, $p < .01$). The accuracy measures for English are higher than those for Dutch. However, the baseline measures are also significantly higher than those for Dutch in all conditions (standard tokens: $t(8) = 2.6$, $p < .05$; standard types: $t(8) = 4.6$, $p< .01$; expanded tokens: $t(8) = 5.6$, $p < .01$; expanded types: $t(8) = 6.7$, $p < .001$). The fact that the accuracy of frame-based categories in English is higher than in Dutch leads us to the second prediction:

(2) Dutch infants show a delay in the use of frequent frames for categorization compared to their English age peers.

Whereas the first prediction points to a cross-linguistic similarity, the second prediction entails a cross-linguistic difference. The two predictions do not exclude one another since the cross-linguistic difference is a gradual one; the fact that frequent frames are more informative in English does not mean they are uninformative in Dutch. The predictions will be tested by means of a perception experiment with two groups of Dutch infants: 12-month-olds and 16-month-olds.

Table 3. Token and type accuracy for standard and expanded analysis including baseline accuracy of random categories for all Dutch corpora.

	Standard Analysis				**Expanded Analysis**			
	Tokens		Types		Tokens		Types	
Corpus	A *	R **	A	R	A	R	A	R
Matthijs	0.76	0.43	0.64	0.38	0.58	0.18	0.41	0.18
Sarah	0.69	0.38	0.55	0.31	0.58	0.16	0.40	0.14
Daan	0.69	0.39	0.57	0.34	0.53	0.16	0.38	0.16
Laura	0.69	0.33	0.58	0.29	0.56	0.13	0.41	0.14
Mean	0.71	0.38	0.58	0.33	0.56	0.16	0.40	0.16

* A = Numbers for analysis indicated in upper cell
** R = Numbers for randomly compiled categories

Table 4. Accuracy of frame-based categories in English and Dutch for standard and expanded analysis.

	Standard analysis				**Expanded analysis**			
	Tokens		*Types*		*Tokens*		*Types*	
Language	Anal	Rand.	Anal.	Rand.	Anal.	Rand.	Anal.	Rand.
English	0.98	0.46	0.93	0.47	0.91	0.27	0.91	0.38
Dutch	0.71	0.38	0.58	0.33	0.56	0.16	0.40	0.16

3. Testing the use of frequent frames by Dutch infants

To test the predictions following from the input analysis, I conducted a perception experiment with Dutch infants. Based on the statistical analyses of the frame-based category accuracy rates for English and Dutch input, I extracted two predictions for a perception experiment with Dutch infants. Prediction (1) entails that, similar to the English infants from Mintz's (2006) experiment, Dutch infants will use frequent frames to categorize content words. The frame-based categories resulting from frequent frames are significantly more accurate than the random baseline categories as was the case for English. Since the English-learning 12-month-old infants from Mintz's experiment used frequent frames for categorization, we expect, based on the significant differences between frame-based and random baseline categories, that Dutch 12-month-old infants will also use frequent frames for categorization; therefore, I first tested 12-month-old Dutch infants.

However, the accuracy scores for Dutch are significantly lower than the accuracy scores for English across the board. This may cause a delay in the use of frequent frames by Dutch infants because the information Dutch infants get from the input is less consistent. Therefore, they probably need more input to reach the same categorization level as the English infants. Prediction (2) entails that Dutch infants show a delay in the use of frequent framesfor categorization compared to their English age peers. We expect that older Dutch infants have progressed in the use of frequent frames for categorization compared to the 12-month-olds; therefore, I also tested a group of 16-month-old Dutch infants to see whether they differ from the English and Dutch 12-month-olds.

3.1 Stimuli

The stimuli consisted of sentences in which novel words are embedded in eight different frequent frames. The selection of the four novel words used in the experiment was based on their phonological make-up and their perceivability for adult native speakers of Dutch. Since Dutch nouns have more possibilities in terms of their phonological structure than verbs (Trommelen, 1989), I had to make sure the nonsense items were phonologically ambiguous between nouns and verbs. Otherwise the phonology could give extra cues as to the category of the novel items. Therefore, I took novel items with a segmental make-up similar to those that were classified equally often as nouns and as verbs by Dutch adult speakers (Don & Erkelens, fc). Two novel items have the segments CCVC (*plif* and *klot*) whereas the other two have CVVC (*daap* and *sook*). Each type of frame contains two novel items, one from each segmental make-up. The division of novel words over frame types was counterbalanced across test groups A and B. As a result, *sook* and *plif* are verbs for the A-group whereas they are nouns for the B-group and vice versa: *daap* and *klot* are nouns for the A-group, whereas they are verbs for the B-group.

The following selection criteria formulated by Mintz (2006) were also the basis of the selection of frequent frames for the Dutch experiment: (1) frames have to be among the frequent frames in as many different corpora as possible, and (2) no overlap between framing words is allowed. Just as Mintz did, I added four training sentences (one for each novel word) that did not contain frames, but bigrams (i.e., two frequently co-occurring words where the first word serves as the context for the other), to make sure infants segment the novel words. By presenting these at the end of the sentence, the chance that infants do indeed segment the novel words is high. This resulted in the selection of three verbal frames and three nominal frames that are frequent in as many different corpora as possible, and one verbal and one nominal bigram.

For verbal frequent frames in Dutch, the first selection criterion is easily met; a considerable number of verbal frequent frames occur in all four Dutch input corpora. However, the second selection criterion cannot be met for Dutch verbal frames; there are no three different verbal frames without any overlap in framing words. This is a consequence of the fact that verbs are more richly inflected for person and number in Dutch than in English. If we want to present only frequent frame information to the infants and not additional morphological information, the novel words cannot be inflected. Therefore, only frames in which the intervening word is a bare stem can be selected. In Dutch, only first person singular (*ik* 'I') and second person singular inverted (*je* 'you') have a bare stem. This means that one of the pronominals *je* and *ik* has to be used twice in the experimental frames. The four verbal frames that were among the frequent ones in all of the four input corpora are *wat X je* 'what X you', *dan X je* 'then X you', *ik X het* 'I X it' and *ik X niet* 'I X not'. Out of these four frames, three have to be selected as experimental frames. Because the pronoun *het* 'it' is ambiguous in Dutch in the sense that it can also be an indefinite determiner ('the'), I chose not to select *ik X het*, leaving the three verbal frames *wat X je* 'what X you', *dan X je* 'then X you', and *ik X niet* 'I X not'. Since *ik* is the only pronoun preceding a verb that selects a stem, the only sentence-final verbal bigram possible in Dutch is *ik X* 'I X'.

The nominal frames should also be frames that are among the most frequent in as many corpora as possible. However, there are not enough nominal frames among the frequent frames from the corpus analysis since they occurred in the Dutch input speech much less frequently than the verbal frames. I had, therefore, to add frames that do not occur in the 45 most frequent frames from the corpora. Since Dutch has gender marking on the determiner, I furthermore had to make sure that no gender violations could occur within the nonsense items. All stimuli are selected in a way that all nonsense items have common gender: they take as definite determiner *de*, as indefinite determiner *een*, and as demonstratives *die* and *deze*. The three nominal frames selected for the experiment are *de X van* 'the X of', *een X op* 'a X on', and *die X die* 'that X that'. The only remaining word with common gender to use for the bigram is *deze*, so the nominal bigram for Dutch will be *deze X* 'this X'.

In Table 5, the resulting set of verbal and nominal frames to be used in the experiment with their accuracy and frequency scores are presented.

Table 5. Frequency and accuracy of frames used in the stimuli across the four Dutch CHILDES corpora (Matthijs, Sarah, Daan, Laura)

Frame	Mean accuracy (standard, tokens)	Mean frequency (% of total # of utterances)
dan X je	0.96	0.51
wat X je	0.99	1.73
ik X niet	0.60	0.34
Total Verbs	*0.85*	*0.86*
de X van	0.95	0.15
een X op	0.73	0.23
die X die	0.43	0.15
Total Nouns	*0.71*	*0.18*

The overall frequency of the frames used in the stimuli (see table 6) is comparable to the frequency of the English frames used in the experiment by Mintz (2006).[1] Although the Dutch verbal frequent frames seem slightly more frequent and the Dutch nominal frequent frames slightly less frequent than the English ones, these differences were not significant ($\chi^2 = 2.00$, $p = .16$).

Table 6. Frames used in experimental stimuli: frequency of occurrence in input corpora as percentage of the total number of utterances in those corpora.

	English (% of total # of utterances)	Dutch (% of total # of utterances)
Verbal frames	0.75	0.86
Nominal frames	0.32	0.18
All frames	0.54	0.52

The Dutch infants learned the four nonsense words *sook, plif, daap,* and *klot*. Two of these nonsense words were trained in four frequent frames that host verbs (*dan_je, wat_je, ik_niet, ik_#*). The other two nonsense words were presented in the four frames that host nouns (*de_van, een_op, die_die, deze_#*). Each nonsense word is thus presented in four different frames, but these frames are all either nominal or verbal. A word trained in nominal frames is a nonsense noun, and a word trained in verbal frames is a nonsense verb. If the infants use the distributional cues of the frequent frames, they have learned in the training phase that two of the nonsense words can occur in verbal frequent frames whereas the other two nonsense words can occur in nominal frequent frames. In the test phase infants heard the same nonsense words in either a similar category frame (i.e., a nonsense verb in a verb frame) or an inconsistent frame (i.e., a nonsense verb in a noun frame). For example, if the nonsense verbs were *sook* and *plif* and the nonsense nouns *daap* and *klot* (group A), a child heard in the training phase sentences like *dan plif je de kamer* 'then, you plif the room', *ze zien dat ik sook* 'they see that I sook', *ik zie de daap van pappa* 'I see the daap of daddy', and *die klot die daar staat* 'that klot that stands overthere'. For this child, consistent sentences in the test phase would be *wat sook je daar* (*sook* in a verb frame) and *die daap die daar staat* (*daap* in a noun frame) and inconsistent sentences would be *ik klot niet hoor* (*klot* in a verb frame) and *er ligt een plif op de kast* (*plif* in a noun frame). The complete set of stimuli, with the Dutch frames and novel items following the experimental set-up of Mintz (2006), is presented in Appendix I.

3.2 Procedure

The experiment is based on the Head Turn Preference Paradigm with a training phase (Kemler et al., 1995). It consists of three phases: a training phase, a contingency training phase, and a test phase. In the training phase, the child hears the training sentences, which depend on the group to which she belongs, as outlined in Table 7: group A hears the first set of training sentences and group B the second set, differing only in the frames in which the novel words are embedded. The training sentences are played continuously divided over six

[1] For this comparison I used only the first four English corpora (Peter, Eve, Nina, Naomi) since the number of utterances in these four corpora is comparable to the number of utterances in the Dutch corpora analyzed here.

randomized blocks. Each block contains all training sentences in a random order. I made sure that in the randomized orders no two sentences with the same frames followed one another, no three sentences with the same novel word occurred in succession, and no block began with the *dan X je …* 'then you X…' as a first sentence. The reason for this last criterion was that the conjunct *dan* 'then' in Dutch is always preceded by some content that locates the time or place of the following utterance. Therefore, it would have been very unnatural to start a block of sentences with this kind of word. During training the lights as described below are initially influenced by the child's head-turns towards the lights, but once the training trial starts, the sentences keep playing and the light keeps burning independent of the child's gaze direction, to make sure every infant hears all training sentences. Each training trial lasts 19 seconds, adding up to a total of 114 seconds for the entire training phase.

Between the training and test phases, a brief contingency training phase is inserted. The aim is to accustom the infants to the influence their behavior has on the lights and the auditory stimulus, i.e., that there is a contingency between their head turn, the lights, and the sounds. Since this contingency is used in the test phase, infants had to complete the contingency training successfully. In this phase, the center light is activated until the child turns her head towards the center. Then the light goes out and a light on one of the sides is activated. As soon as the child looks at the sidelight, a tone of 500 Hz is presented from the loudspeaker behind that light. The tone takes 1 second and is repeated as long as the child looks at it, with pauses between each tone of 100 milliseconds. If the child looks away, or if the 15 repetitions have ended, the center light flashes again and a second, similar trial starts.

After two contingency training trials, the test phase starts. The method is similar to that of the contingency training phase. The tones are replaced by the test sentences in Table 1. Each sentence is repeated 15 times in a trial, with pauses of 500 milliseconds. There are 8 test trials: one for each test sentence. These trials are presented randomly. Both groups hear the same test sentences since the consistent sentences for group A are automatically inconsistent for group B and vice versa. I measured the time their head is directed towards the sidelights for each sentence and compare the total amount of time for consistent and inconsistent sentences.

The presentation of the stimulus to either side of the child was randomized by the computer but constrained so that no more than three consecutive trials are presented from the same side.

3.3 Principled coding system for exclusion of infants

Since the participants in this kind of perception experiment are very young, their behavior is also very variable, which results in high numbers of excluded participants. The reported dropout rates for Headturn Preference experiments vary between 15 and 40% (Kemler et al., 1995). Participants are excluded from analysis for various reasons, mostly due to 'inattentive behavior'. However, this general criterion is operationalized with diverse measures such as (among many others) 'failure to look for an average of at least 3 seconds during test', 'crying' (Gomez & Gerken, 1999), 'excessive fussiness' (Gomez, 2002), and 'unresponsiveness or falling asleep' (Santelmann & Jusczyk, 1998). Although every study reports on exclusion criteria, there is no principled coding system for exclusion of infants. Therefore, a principled coding system was developed (Erkelens & Polišenská, in preparation).

Secondly, the coding system was used to rate each infant based on her behavior during the experiment only. The only valid reason to exclude infants from the analysis of an experiment on the basis of behavior is serious doubt about whether the infant pays attention to the experiment at all. The coding system distinguishes four different groups of behavioral signals that differ in seriousness with respect to the attentiveness of the infant and are marked by the colors red, yellow, blue, and green (in descending order of seriousness - see Appendix II). All infants were coded for their behavior based on these signals. A selection of the infants was double-checked by a second researcher to make sure that the criteria are objectively interpretable. There was 100% agreement between two researchers on the code.

Red-coded infants are automatically excluded from further analysis whereas all other infants are included. The encodings enable checking of interactions between behavior and results. In case the behavior during the experiment has significant effects on the results, one could make a motivated decision to later exclude yellow and / or blue infants (cf. Erkelens & Polišenská, in prep.).

3.4 Participants

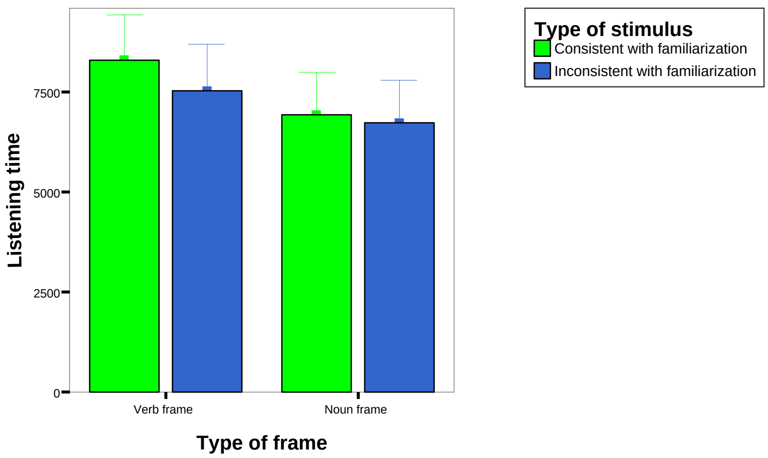
For the younger age group (12 months), I tested 66 infants who were all healthy, growing up in a monolingual Dutch environment, without risk for dyslexia or developmental delays as a consequence of prematureness. Three infants were excluded before encoding because of experimenter or computer errors. Of the remaining 63 infants the behavior was encoded based on the coding system devised by Erkelens & Polišenská (in prep.). The behavior of the infants in this age group was quite attentive: 44 infants were coded green, 17 were coded yellow, and only 4 were coded red. No infant showed differential behavior during parts of the experiment, so the blue code was not applied. The average age of the 59 infants left after exclusion of the red-coded infants was 1:00;07 (range 0:11;22-1:01;00). Group A consisted of 30 subjects and group B consisted of 29 subjects. There were more girls (34) than boys (25) in the sample.

For the older age group (16 months) I tested 64 infants that were all healthy, monolingual Dutch infants without risk for dyslexia or developmental delays as a consequence of prematureness. Four infants were excluded before encoding because of experimenter or computer errors. Of the remaining 60 infants the behavior was again encoded based on the Erkelens & Polišenská (in prep.) coding system, resulting in the following numbers: 35 green, 1 blue, 15 yellow, and 9 red encodings. The average age of the 51 infants left after exclusion of the red-coded infants was 1:04;03 (range 1:03;08-1:04;28). Group A consisted of 25 subjects and group B consisted of 26 subjects. There were more boys (27) than girls (24) in the sample.

3.5 Results

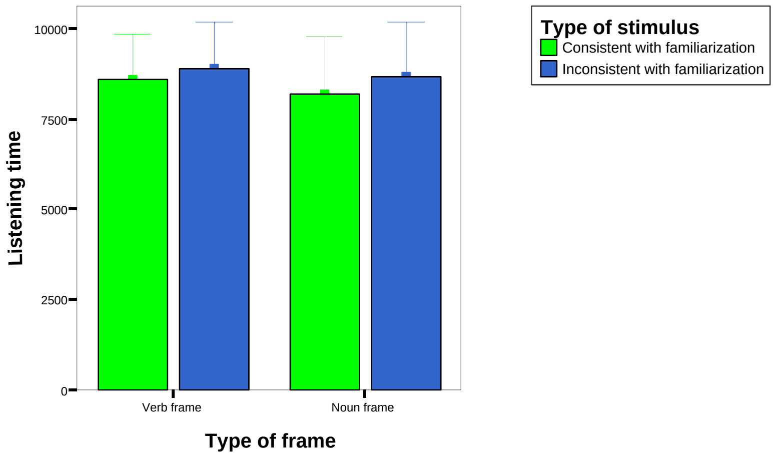
The expected result following from the first prediction was that 12-month-old Dutch infants also differed in listening times between items that were consistent and inconsistent with training as the English 12-month-old infants did. Figure 1 shows the listening times for the 12-month-olds. We see that the difference between consistent and inconsistent stimuli is greater for the verb frame items (8.3 and 7.5 seconds respectively) than for the noun frame items (6.9 and 6.8 seconds respectively). I conducted a mixed ANOVA (between subjects: test group (2), gender (2), behavior (2) x within subjects: stimulus type (2), frame type (2)) on the loglinear values of the listening times because the listening times were significantly different from a normal distribution (Kolmogorov-Smirnov test for normal distribution: $Z = 3.84$, $p <. 001$). The only significant results of the ANOVA were a significant main effect of frame type ($F(1,51) = 8.72$; $p < .01$) and a significant two-way interaction of frame type x behavior ($F(1,51) = 4.46$, $p < .05$). There was no effect of stimulus type at all, which means that these Dutch results are different from the English results by Mintz (2006). Whereas Mintz found a significant difference in listening times between inconsistent and consistent verb frame stimuli, the results from my experiment show no difference at all between inconsistent and consistent items. Furthermore, the between-subjects effects test group, gender, and behavior each had no significant effect on average listening times. Finally, the 12-month-olds listened longer to verb frame items than to noun frame items. The first prediction is not borne out by the data.

Figure 1. Average listening times per stimulus type, split up per frame type (in msec) for the 12-month-olds



The second prediction of the experiment was that Dutch infants show a delay in the use of frequent frames due to the accuracy differences between English and Dutch frame-based categories. Although we have seen that the Dutch 12-month-olds did not use the frequent frames for categorization, it still might be the case that the 16-month-olds do. Figure 2 shows the listening times per frame type for this age group. We see minor differences in the listening times. The difference between consistent and inconsistent stimuli is somewhat greater for the noun frame items than for the verb frame items. Again I conducted a mixed ANOVA (between subjects: test group (2), gender (2), behavior (2) x within subjects: stimulus type (2), frame type (2)) on the log values, which revealed that none of the factors had a significant main effect. The main conclusion of this experiment with older infants compared to the younger infants is that the frame type effect has disappeared, but none of the other factors have a significant influence on the listening times. So far, the second prediction does not get support from the 16-month-olds' data.

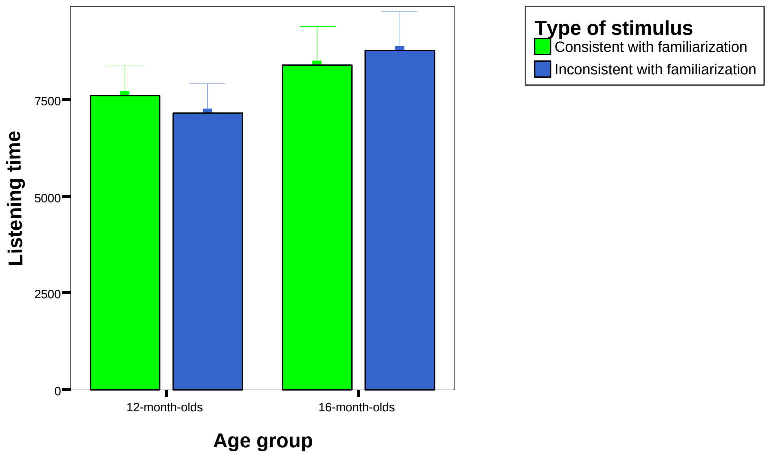
Figure 2. Average listening times per stimulus type, split up per frame type (in msec) for the 16-month-olds



We have seen that neither age group differed significantly in their listening times for consistent and inconsistent test stimuli. However, they did show different tendencies for preferences. The 12-month-olds tended to listen longer to consistent verb frame sentences than to all other sentences whereas the 16-month-olds tended to listen longer to inconsistent verb and noun frame sentences than to consistent verb and noun frame sentences, which still might point to a development towards the English pattern. We will see whether these differences between the age groups are significant. If they are, it shows that there is at least some development between 12 and 16 months with respect to the categorization of novel words and the second prediction would be somewhat supported.

Both age groups made minor differences in listening times between consistent and inconsistent test sentences, but the preference was opposite: the 12-month-olds listened longer to the consistent items whereas the 16-month-olds listened longer to the inconsistent items. The opposite patterns are clearly observable from the graph in Figure 3.

Figure 3. Comparison of age groups for listening time per stimulus type.



For the inconsistent stimuli, the difference between the age groups is considerable: 7.1 seconds for the younger group and 8.8 seconds for the older group. This difference is significant ($t(108) = -2.08$, $p < .05$), indicating that the 16-month-olds listened longer to the inconsistent items than the 12-month-olds. This seems to be an effect of stimulus type on age group that needs further exploration by means of an ANOVA. In the comparison of the age groups we have an additional between-subjects factor 'age', which results in a mixed ANOVA (between subjects: age (2), test group (2), gender (2), behavior (2) x within subjects: stimulus type (2), frame type (2)) on the loglinear values. The main effect of age is marginally significant ($F(1) = 3.55$, $p = .06$), which indicates that the two age groups differ from each other in listening times to the stimuli. This marginally significant main effect results neither in a significant interaction of age and consistency ($F(1,93) = 0.15$, $p = .70$) nor in a significant interaction of age and frames (F(1,93) = 0.70, p = .40). Summarizing the comparison of the two groups, the apparent significant difference in listening times to inconsistent items revealed by the t-test is not confirmed by the ANOVA. Although there is a marginally significant main effect of age on the values of the listening times, there are no significant interactions.

4. Conclusion and discussion

Based on the comparison between Dutch and English with respect to the informativeness of frequent frames, I predicted that Dutch infants would use the frequent frame information for early categorization, albeit somewhat delayed due to the differences in accuracy of the frame-based categories. The results of the perception experiment conducted with Dutch infants of 12 and 16 months of age revealed that Dutch infants use the frequent frame information in a different way than the English 12-month-old infants from Mintz's (2006) perception experiment. Both prediction (1) and prediction (2) are not borne out by the data from the perception experiment reported here although the delay predicted by prediction (2) may be even longer than 4 months. I will come back to this point later in this section. After training of novel words in verbal and nominal frequent frames, the English infants showed longer listening times to sentences that contained inconsistent pairing of verbal frequent frames and novel words whereas both groups of Dutch infants did not show any significant differences between consistent and inconsistent pairings. Since frequent frames are as frequent in the English input as they are in the Dutch input, the difference cannot be explained in terms of frequency. If frequency is the important factor, Dutch infants should be as good as the English infants in detecting the frequent frames. Furthermore, since the accuracy of the Dutch frame-based categories is significantly higher than the accuracy of Dutch random baseline categories, the frequent frames do give cues to the category of the intervening words in Dutch. However, the frame-based categories for English are much more accurate than the Dutch frame-based categories. This fact makes the frequent frame cues less reliable for Dutch children to use in initial categorization.

The experiment reported here shows that the frequent frame information as defined by Mintz (2003) is not universally used as an early cue for lexical categories. Whereas the English frequent frames provide reliable information with respect to the category of the intervening elements, the Dutch frequent frames provide less reliable information. The consequence of this difference is attested in this experiment. The Dutch 12-month-olds, as well as the Dutch 16-month-olds, did not categorize novel words presented in verbal frequent frames whereas the English 12-month-olds categorize novel words presented in verbal frequent frames. If frequent frame information were a universal learning mechanism for categories, this would imply that Dutch infants start to categorize at least five months later than English infants. Since this is not what we find in the production data of English and Dutch children (Bloom, 1970; Erkelens, 2006), there must be some other mechanism that is used by the Dutch infants for early categorization. Although Dutch and English are both Germanic languages, there are considerable differences in word order and morphology. For example, Dutch has richer inflectional morphology than English, and Dutch word order differs between main and subordinate clauses. As a consequence of these differences, distribution in terms of morphemes instead of words is potentially more informative for categories in Dutch. Furthermore, as Freudenthal, Pine & Gobet (in press) point out, the inclusion of sentence ends as framing elements possibly improves the frequent frame model. In their model of a syntax-acquiring child, such an improvement provides data that are closer to real children's production data than the original model as presented by Mintz in 2003. In further experiments, the possibilities of inflectional morphemes or sentence ends as framing elements should be investigated to see whether Dutch infants may rely more on these cues than purely on the lexical words as framing elements.

All in all, this experiment shows that there are language-specific restrictions on infants' use of distributional cues like frequent frames to bootstrap lexical categories. One way to understand these results is that a complete set of cues is in principle available in all languages but not to the same extent (cf. Monaghan, Christiansen, & Chater, 2007). The relative strength of the cue in a certain language probably determines the moment and importance of the use by language learning infants. If a certain cue very strongly points to the adult categorization of the language – as frequent frames seem to do for English – infants use the cue early and strongly. However, if a cue has weaker links to the adult categorization of the language – as frequent frames in Dutch – infants maybe never use the cue since there are other, better cues for their native language.

APPENDIX I - Training and test sentences for the perception experiment with Dutch infants.

GROUP A	
Verb Frame training sentences	*Noun Frame training sentences*
Dan *plif* ***je*** de kamer	Ik zie ***de*** *daap* ***van*** pappa
Dan *sook* ***je*** de kamer	Ik zie ***de*** *klot* ***van*** pappa
Ze zien dat ***ik*** *plif.*	Hij wil ***deze*** *daap*.
Ze zien dat ***ik*** *sook.*	Hij wil ***deze*** *klot*.
Wat *plif* ***je*** daar?	Er ligt ***een*** *daap* ***op*** de kast.
Ik *sook* ***niet*** hoor.	***Die*** *klot* ***die*** daar staat.
GROUP B	
Verb Frame training sentences	*Noun Frame training sentences*
Dan *klot* ***je*** de kamer.	Ik zie ***de*** *sook* ***van*** pappa.
Dan *daap* ***je*** de kamer.	Ik zie ***de*** *plif* ***van*** pappa.
Ze zien dat ***ik*** *klot.*	Hij wil ***deze*** *sook*.
Ze zien dat ***ik*** *daap.*	Hij wil ***deze*** *plif.*
Wat *klot* ***je*** daar?	Er ligt ***een*** *sook* ***op*** de kast.
Ik *daap* ***niet*** hoor.	***Die*** *plif* ***die*** daar staat.
TEST ITEMS	
Consistent-A, Inconsistent-B	*Inconsistent-A, Consistent-B*
Wat *sook* ***je*** daar?	***Wat*** *daap* ***je*** daar?
Ik *plif* ***niet*** hoor.	***Ik*** *klot* ***niet*** hoor.
Er ligt ***een*** *klot* ***op*** de kast.	Er ligt ***een*** *plif* ***op*** de kast.
Die *daap* ***die*** daar staat.	***Die*** *sook* ***die*** daar staat.

APPENDIX II - Criteria of the system for encoding infants' behavior in an HPP-experiment

Code	Criteria
Red	• Refusing to sit still and managing to get off the parent's lap • Crying; giving the impression that the infant is distressed by the situation (face and body expression) • Not reacting to the lights • Not completing the experiment • Looking at the lights for less than one sentence for one or more trials. • Looking at the lights for more than the stimulus lasts in at least half of the experiment.
Yellow	• Staying on parent's lap (sitting or standing) and reacting to the stimuli • Looking back-and-forth multiple times during trials (re-orientation within 2 seconds). • Being physically active in between the trials (turning the whole body to the parent; moving the whole body from the waist up while looking around; waving arms; shaking head).
Green	• Sitting still on the caregiver's lap • Reacting spontaneously to the lights • Giving the impression that she is alert • Giving the impression that she feels relaxed
Blue	• Behaving according to the 'green' code for at least half of the experiment. • Behaving according to the 'yellow' code for the rest of the experiment.

Acknowledgments

I am grateful to all participants, their parents, and to Elma Blom, Daniela Polišenská, Theo Veenker, and Kamila Polišenská for making the experiments possible. Furthermore, I would like to thank Elma Blom, Daniela Polišenská, Robert Cloutier, Jan Don, and Anne Baker for their useful comments both on the research in general as specifically on this paper.

References

Bloom, L. (1970). *Language development form and function in emerging grammars*. Cambridge, Mass.: the M.I.T.press.

Don, J., & Erkelens, M. A. (fc). Possible phonological cues in categorical acquisition: evidence from adult categorization. *Studies in Language (Special issue on Parts-of-Speech)*.

Erkelens, M.A. (in progress). *The acquisition of verbs and nouns by Dutch children*. PhD thesis, ACLC, Universiteit van Amsterdam.

Erkelens, M. A. (2006). *Children know how to use their verbs, nouns, and adjectives.* Paper presented at the Conference on Universality and Particularity in Parts-of-Speech Systems, Amsterdam.

Erkelens, M. A., & Polisenska, D. (in preparation). *A principled coding system for excluding infants in Headturn Preference experiments.* Unpublished manuscript, ACLC, Universiteit van Amsterdam.

Freudenthal, D., Pine, J. M., & Gobet, F. (in press). *Simulating the Noun-Verb Asymmetry in the Productivity of Children's Speech.* Paper presented at the Eighth International Conference on Cognitive Modelling.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of child language, 32*(2), 249-268.

Gomez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science, 13*(5), 431.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-136.

Kemler, N. D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The Head-Turn Preference Procedure for Testing Auditory Perception. *Infant Behavior and Development, 18*(1), 111-116.

Krikhaar, E., & Wijnen, F. (1996). Children's Categorization of Novel Verbs: Syntactic Cues and Semantic Bias. In E. V. Clark (Ed.), *The Proceedings of the Twenty Seventh Annual Child Language Research Forum* (pp. 197-206). Stanford, CA: CSLI.

Maratsos, M. P., & Chalkley, M. A. (1980). The Internal Language of Children's Syntax: The Ontogenesis and Representation of Syntactic Categories. In K. E. Nelson (Ed.), *Children's Language* (Vol. 2, pp. 127-214). New York: Gardner.

Mintz, T. H. (2003). Frequent Frames as a Cue for Grammatical Categories in Child Directed Speech. *Cognition, 90*(1), 91-117.

Mintz, T. H. (2006). 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.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The Differential Role of Phonological and Distributional Cues in Grammatical Categorization. *Cognition: International Journal of Cognitive Science (Cognition), 96*(2), 143-182.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis Cross-linguistic evidence in language acquisition. *Cognitive psychology, 55*(4), 259.

Redington, M., Chater, N., & Finch, S. (1998). Distributional Information: A Powerful Cue for Acquiring Syntactic Categories. *Cognitive science, 22*(4), 425-470.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: evidence for limitations in processing space. *Cognition, 69*(2), 105-134.

Trommelen, M. (1989). Lettergreepstruktuur en woordkategorie. *De Nieuwe Taalgids, 82*, 64-77.

Van Kampen, J. (2004). The learnability of syntactic categories. In J. v. Kampen & S. Baauw (Eds.), *Proceedings of GALA 2003* (pp. 245-256). Utrecht: LOT.