Learning to categorize verbs and nouns: studies on Dutch

Erkelens, M.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
4 Category-indicative properties of Dutch input

The results from the production studies in Chapter 3 have shown that children already have initial word class categories from the moment their categorization abilities can be assessed in production. In order to gain more insight into the process of how children learn to categorize, the knowledge of younger children needs to be studied, using data other than from sentence production. Before the perception of categories in younger children is studied in Chapter 5, in this chapter Dutch adult language is studied to identify category-indicative properties that are available to pre-verbal infants. Two properties are under investigation here, namely phonological properties and co-occurrence patterns of words and morphemes. An experimental study on the use of phonological form as an indicator of category by Dutch adults is reported first (§4.2). Then an input study on the presence of so-called ‘frequent frames’ (frequent co-occurrence of two non-adjacent words) is reported (§4.3). The outcomes are discussed and elaborated with a proposal concerning the relevance of frequently co-occurring non-adjacent morphemes in Dutch. This chapter finishes with by presenting and discussing methods to study language perception in infants (§4.4).

4.1. Introduction

As shown in the production studies in Chapter 3, children are able to distinguish verbs and nouns once they produce sentences. Their input language apparently contains category-indicative properties that enable them to start categorizing words as verbs and nouns at a very early stage of development. Since infants become sensitive to the phonemic categories of their first language already between 6 and 9 months of age (Jusczyk, 2002), phonology is the first candidate for category-indicative properties that are available early. If the phonological form of words is indicative of category, children may benefit from these links in an early stage of development. Therefore, the availability of phonology as a category-indicative property in Dutch will be studied here first.

Studies indicate that English-learning infants are able to segment words in an ongoing speech stream from 9 months onwards (Jusczyk, Hohne, & Bauman, 1999; Mattys & Jusczyk, 2001). This ability to segment the utterance into phonological words makes the co-occurrence patterns of those words accessible for these 9-month-olds. The statistical distribution of words in sentences (co-occurrence patterns) is thus the second candidate for a category-indicative property that is
available early. The availability of co-occurrence patterns as an indication of the category of intervening words in Dutch will be studied in § 4.3.

To test whether the available category-indicative properties in Dutch are also used by Dutch infants in early categorization, a proper method is needed to test pre-verbal infants. Young infants’ perception can be measured by means of experimental techniques like the Head-turn Preference Procedure. Such techniques provide important empirical evidence for early language development. However, the methods also have their limitations. In §4.4 the suitability of these methods to test early categorization will be discussed. Attention will be paid to the interpretation of this kind of data in general and more specifically to the interpretation of data on categorization.

### 4.2. Category-indicative phonological properties: an experimental study of adult Dutch

For a long time, phonology and prosody were regarded as unreliable cues for determining categories because of their highly language-specific and arbitrary nature (e.g., Hockett, 1960). However, Morgan & Demuth (1996) introduced the term ‘phonological bootstrapping’, thereby proposing that the speech stream contains phonological information that leads to grammatical information, among which the lexical categories of words. Kelly (1996) showed that the phonological features stress pattern, number of syllables, duration, vowel quality, consonant quality, and number of phonemes correlate highly with grammatical class in English and that human beings are sensitive to these cues. Cassidy & Kelly (1991) showed that English 4-year-olds are sensitive to the number of syllables in words as a cue to their grammatical class. They presented nonsense words of one, two or three syllables and asked the children whether they thought the nonsense word referred to the action or to the object in a videotaped event. The children more often associated the action with the monosyllabic nonsense words than with the tri-syllabic nonsense words.

Subsequent work with artificial learning mechanisms confirms that phonological information can be used for lexical categorization. Durieux and Gillis (2001) showed that learning algorithms can successfully distinguish categories on the basis of phonological information alone. They studied the discrimination of three different lexical categories: verb, noun, and adjective. Their artificial learning system had to

---

predict the lexical category of a word, using only phonological and prosodic information. Several experiments were conducted, varying in target language and cue combination. The overall results showed that, both in English and Dutch, the learning system predicted the lexical categories most successfully if learning was done on the basis of raw segmental data from the target language. Although phonological cues like those presented by Kelly (1996) provide useful bootstrapping information, even purely simple segmental encoding in onset, nucleus, coda and stress pattern provides information about category. Durieux and Gillis (2001) concluded that the link between phonology and lexical category might function as a cue in category acquisition.

Other computational linguists have elaborated on these findings recently. On the basis of a corpus analysis, Monaghan, Chater and Christiansen (2005) showed that phonological cues can be especially useful in categorizing low frequency items. This corpus-based prediction was corroborated in an artificial language learning experiment, suggesting that learners may indeed use this kind of information to categorize items. In a further series of experiments Farmer, Christiansen and Monaghan (2006) showed that in processing English words subjects are sensitive to the word class typicality of a phonological form; a noun that has the typical phonological form of nouns is more easily processed than a noun that does not. Typicality was expressed as the ‘phonological distance’ from other members of the same category. Their results show that English speakers are indeed sensitive to such phonological typicality. Monaghan and Christiansen (2006) concluded, on the basis of their simulations, that although the relationship between individual word forms and meaning is arbitrary, the relationship between groups of word forms and their grammatical category shows a fair amount of systematicity.

A corpus analysis of child-directed speech by Monaghan, Christiansen and Chater (2007) shows that English, Dutch, French and Japanese all contain a considerable number of phonological cues that can lead to the accurate categorization of words as verbs and nouns. Although the specific phonological cues that are informative differ across languages, correlations between certain phonological properties and categories are available in typologically quite different languages such as English and Japanese.

As shown by Durieux & Gillis (2001) and Monaghan et al. (2007), Dutch phonology is also category-indicative to some extent. The next question is of course whether speakers of Dutch are able to use this category-indicative information. Farmer et al. (2006) showed for English that adults are able to use the phonological typicality of words for categorization. The experiment reported in this chapter confirms these results for Dutch, but the method and the type of phonological cues
that are tested differ. The aim was not to test sensitivity to the gradual notion of
typicality, but sensitivity to well-defined properties (impossible forms) of the
phonology of Dutch verbs that were first brought to light by Trommelen (1989).

There are several generalizations in Dutch that relate the phonological form of a
(content) word to its lexical category, as Trommelen showed in 1989 and others
later, as discussed above. An experiment was conducted to test whether in assigning
lexical categories to nonsense words, adult native speakers of Dutch are able to use
the phonological cues that can be deduced from Trommelen’s generalizations.

4.2.1. Research question and predictions

By analyzing all possible clusters in Dutch verbs and nouns, Trommelen (1989)
showed that the complexity of final rhymes in underived words can be an indication
of word class. With respect to verbs, Trommelen observed that, generally speaking,
they conform to a set of phonological restrictions. Many phonological forms that
would make perfect nouns in Dutch do not occur as verbs. Based on Trommelen’s
(1989) observations, the following generalizations can be formulated about the
phonological form of underived verbs in Dutch:

- verbs are monosyllabic;
- verbs do not end in a monophthongal vowel (including schwa), except for ga
- rhymes larger than three elements do not occur in verbs, apart from peins
  [pɛins] ‘to consider’, and veins [vɛins] ‘to pretend’;
- verbs ending in schwa followed by [m] do not occur.

There are some exceptions to these phonological restrictions, but in those cases the
words are either denominal, foreign, or extremely rare. The possible phonological
form of an underived verb thus forms a proper subset of the possible phonological
form of Dutch (underived) words. Only nouns can have the full range of possible
phonological forms.

The first of Trommelen’s generalizations requires some explanation. With
respect to their phonological make-up, typical Dutch verbs come in two types:
monosyllabic or bi-syllabic with schwa as the kernel vowel of the final syllable.
Kager & Zonneveld (1985) argue that phonologically speaking, ‘bi-syllabic with
schwa’ in fact means ‘monosyllabic’. If we adopt this analysis, we can confirm the
It is important to note, as Trommelen did, that there are in fact two types of verbs that apparently do not obey the above-mentioned generalizations. First, there are verbs ending in the affix -eer (argument-eer ‘argue’; public-eer ‘publish’). These verbs are clearly derived, whereas the phonological generalizations only hold for un-derived verbs. Second, there are seeming counterexamples like olie [oli] ‘to oil’ and ruzie [ryzi] ‘to quarrel’, but these verbs all have phonologically identical and semantically related nominal or adjectival counterparts. Therefore, these verbs can be safely assumed to be conversions of these ‘underlying’ adjectives, or nouns. For example, the verb olie is evidently derived from the phonologically identical noun olie [oli] ‘oil’, and it can safely be assumed that the same holds for the verbal ruzie and its nominal counterpart. Moreover, independent evidence indicates that the seeming counterexamples can safely be regarded as ‘conversions’ (see Don, 1993).


These connections between the phonological properties of words and their category in Dutch can be translated into predictions for the use of these properties in categorization by adults. The question in (1) follows from this observation; that the possible phonological form of an underived Dutch verb forms a proper subset of the possible phonological form of Dutch underived nouns, and forms the basis for the experiment to be reported in this section.

(1) Do Dutch adults use the differences in the possible phonological forms of verbs and nouns, as observed by Trommelen (1989), to categorize nonsense words?

---

15 So far the following counterexamples were found to the generalization that Dutch underived verbs exist only in a single syllable: bakkelei [bakəlɛi] ‘fight’, plavei [plavɛi] ‘pave’, ravot [ravɔt] ‘romp about’, krioel [kriul] ‘teem’, poleist [pɔlɛist] ‘polish’.
The prediction following from the subset – superset relationship between verbs and nouns is that words with a richer phonological make-up than is allowed for underived verbs will be classified as nouns by adult speakers. That is, words with one or more of the typically nominal characteristics will be classified as nouns. This prediction is formulated in (2).

(2) Dutch adults use phonological characteristics to categorize content words as nouns or non-nouns, because the possibilities for the form of underived verbs constitute a subset of the possibilities for the form of underived nouns.

4.2.2. Method

The experiment used to test the prediction in (2) is a decision task: participants had to decide whether certain nonsense stems were nominal or verbal. Using nonsense stems in isolation is a good way to test phonological awareness. The participants’ semantic, morphological, and syntactic knowledge is ruled out because nonsense stems have no meaning, morphological structure or sentential context. Of course, subjects possibly interpret some nonsense forms as morphologically complex, but the phonological form of the nonsense forms was constructed in such a way that they contained no recognizable Dutch affix. The assumption was that although many nouns in Dutch can be ‘converted’ to verbs without any overt phonological affix, subjects would still tend to first recognize such words as nominal rather than verbal.

The specific properties that define a particular stimulus as ‘nominal’, are the properties listed in (3).

(3) Phonological characteristics typical of underived nouns in Dutch:
   • true polysyllabicity (more than one full vowel)
   • ending in a long vowel
   • ‘more than super-heavy’ rhymes (VVCC or VCCC)
   • final schwa
   • final syllable consisting of schwa, followed by [m]

A set of 20 nonsense stems was created, all phonologically possible stems in Dutch. Of these 20, 11 were designed as nouns, according to the generalizations in (3). The other 9 stems were designed to be ambiguous between nouns and verbs. In Table 4.1, the stimuli are listed together with the design principles.
Table 4.1. Overview of the stimuli used in the classification experiment with their Dutch orthography, phonetics, condition and design principle.

<table>
<thead>
<tr>
<th>Phonetics</th>
<th>Condition</th>
<th>Design principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>donkam</td>
<td>Noun</td>
<td>Bi-syllabic</td>
</tr>
<tr>
<td>meliens</td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td>xiveno</td>
<td>Noun</td>
<td>Final long vowel</td>
</tr>
<tr>
<td>strempa</td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td>smoza</td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td>boogst</td>
<td>Noun</td>
<td>Heavy rhyme</td>
</tr>
<tr>
<td>pirst</td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td>rilo</td>
<td>Noun</td>
<td>Final schwa</td>
</tr>
<tr>
<td>krilø</td>
<td>Noun</td>
<td></td>
</tr>
<tr>
<td>fAl´m</td>
<td>Noun</td>
<td>Final schwa + m</td>
</tr>
<tr>
<td>bok</td>
<td>Ambiguous</td>
<td></td>
</tr>
<tr>
<td>plØÉI</td>
<td>Ambiguous</td>
<td></td>
</tr>
<tr>
<td>dil</td>
<td>Ambiguous</td>
<td></td>
</tr>
<tr>
<td>drüuf</td>
<td>Ambiguous</td>
<td></td>
</tr>
<tr>
<td>blap</td>
<td>Ambiguous</td>
<td></td>
</tr>
<tr>
<td>nort</td>
<td>Ambiguous</td>
<td></td>
</tr>
<tr>
<td>kreī</td>
<td>Ambiguous</td>
<td></td>
</tr>
<tr>
<td>balter</td>
<td>Ambiguous</td>
<td></td>
</tr>
<tr>
<td>klest</td>
<td>Ambiguous</td>
<td></td>
</tr>
</tbody>
</table>

The participants of the experiment were 28 adults whose native language was Dutch. The stimuli were randomized and presented on paper with a brief introduction. The participants were asked to choose for each of the presented words whether they were stems of nouns or stems of verbs. In order to make sure that the subjects understood the notion of a ‘stem’, examples were provided of both a nominal and a verbal stem in Dutch. Moreover, the terms ‘nominal’ and ‘verbal’ were briefly introduced and the task was explained. The precise instructions are included in Appendix 4.1. The task itself was presented as a three-column sheet. In the first column all stimuli were listed. The second and the third column were headed by ‘nominal’ and ‘verbal’ respectively. The participants were asked to tick only one of the columns for each stimulus; the nominal column if they thought the
stimulus was a nominal stem, and the verbal column if they thought it was a verbal stem. All participants confirmed that they understood the task before proceeding.

### 4.2.3. Results

First, a general picture of the outcomes of the decision task will be provided. Then, the results will be analyzed in more detail to determine which phonological generalizations are taken to be the most reliable indicator of category in Dutch.

Figure 4.1. Percentages of noun and verb responses per stimulus designed either as a noun (in CAPITALS) or as ambiguous in the phonology experiment with Dutch adults.

A first look at the data shows some general patterns across the participants. Figure 4.1 depicts the percent of participants that judged each stimulus as either nominal or verbal, with the stimuli ordered from left to right in terms of decreasing likelihood of being categorized as a noun. For example, a large majority of the participants categorized the stimuli ‘riele’, ‘donkam’, ‘giveno’, and ‘strempa’ (to the left) as nominal, but ‘pluig’, ‘krei’, and ‘drauf’ (to the right) verbal. This indicates that participants judged the stimuli as different with respect to word class, and this difference is significant ($\chi^2 (19) = 117.1; p < .01$).

Where does this significant difference come from? In other words: did the participants indeed respond ‘nominal’ more often to the stimuli that were designed...
as nominal? And did they respond at chance level to the stimuli designed as ambiguous? Table 4.2 shows the relationship between the design of the stimuli and the responses. For the stimuli designed as ambiguous, the participants slightly preferred a verbal response. The pattern for the stimuli designed as nominal clearly shows the expected asymmetric pattern: the subjects categorize almost 80% of the nominally designed stimuli as nominal stems \( (\chi^2 (1) = 102.0 ; p < .001) \). For the stimuli designed as ambiguous the difference was also significant, although less convincingly \( (\chi^2 (1) = 6.4 ; p = .025) \). Since 11 items were designed as ‘nominal’, subjects could answer ‘noun’ 11 times for the items of that design. On average, subjects provided the answer ‘noun’ 8.64 times (range 5-11, SD = 1.97) and the answer ‘verb’ 2.32 times (range 0-7, SD = 2.02). There was variation across subjects: 3 subjects gave the answer ‘noun’ only 5 times, 3 subjects 6 times, and 1 subject 7 times; the other 21 subjects answered ‘noun’ 8 times or more. The results show a clear preference of the participants for a nominal categorization of nonsense words that were designed as phonological nouns and a weaker preference for verbal categorization of nonsense words that were designed as ambiguous.

Table 4.2. Division of responses over stimulus designs in raw numbers (and percentages) in the phonology experiment with Dutch adults.

<table>
<thead>
<tr>
<th></th>
<th>Noun</th>
<th>Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>242 (79%)</td>
<td>65 (21%)</td>
</tr>
<tr>
<td>Ambiguous</td>
<td>105 (42%)</td>
<td>145 (58%)</td>
</tr>
</tbody>
</table>

The principles that were used to design stimuli as nominal did seem to influence participants to categorize stimuli as nominal stems. There were five different types of nominal stimuli, each corresponding to one of the generalizations found by Trommelen: bi-syllabic, stimuli ending in a long vowel, stimuli with super heavy rhymes, stimuli ending in a schwa, and stimuli ending in a schwa followed by the labial nasal [m] (see Table 4.1). In Table 4.3 the scores for the stimuli grouped in the five types are listed.

Four design principles triggered the participants to judge the corresponding stimuli as nominal in more than 80% of the cases. The exception was the principle of ‘more than super heavy’ rhymes, which resulted in categorization as nominal in only 55% of the cases. As shown above, a statistically significant difference was found in response patterns for all stimuli designed as phonological nouns. A one-way ANOVA of response x design principle shows a significant effect of design principle on the responses \( (F (5,19) = 13.3, p < .001) \). Post hoc comparisons show...
that the response pattern for the stimuli based on four of the five principles differs significantly from the response pattern on stimuli designed as ambiguous (Tukey’s HSD procedure). Only the principle of heavy rhyme does not cause a significant difference and hence does not contribute to the overall significant effect.

Table 4.3. Division of responses over design principles in raw numbers (and percentages) in the phonology experiment with Dutch adults.

<table>
<thead>
<tr>
<th></th>
<th>Noun</th>
<th>Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>bi-syllabic</td>
<td>48 (85 %)</td>
<td>8 (15 %)</td>
</tr>
<tr>
<td>final long vowel</td>
<td>68 (83 %)</td>
<td>14 (17 %)</td>
</tr>
<tr>
<td>more than super heavy rhyme</td>
<td>31 (55 %)</td>
<td>25 (45 %)</td>
</tr>
<tr>
<td>final schwa</td>
<td>49 (88 %)</td>
<td>7 (12 %)</td>
</tr>
<tr>
<td>final schwa + nasal</td>
<td>46 (82 %)</td>
<td>10 (18 %)</td>
</tr>
</tbody>
</table>

4.2.4. Conclusion

Of the five phonological principles presented by Trommelen, four principles do indeed provide enough cues for adult Dutch speakers to categorize stems as nominal. The results of this experimental study show that adult speakers of Dutch are able to make use of phonological cues in the discrimination of nouns and verbs. This result supports the idea that children could use such phonological properties in the process of categorization. Of course, whether this is indeed the case can only be established on the basis of experimental results from children.

4.3. Category-indicative co-occurrence patterns: a study of Dutch input

As shown in the previous section phonology is a possible categorization cue for children. The distribution of words and morphemes in a sentence is possibly also a good indicator of category membership from very early on. Because they are determined by phonological, semantic, and syntactic restrictions, the co-occurrence patterns of words and morphemes in a language are far from random. Maratsos and Chalkley (1980) presented a theory according to which children construct syntactic categories on the basis of the statistical distribution patterns of words. They proposed that children keep track of the co-occurrence patterns of all words and affixes in the ambient language, and form groups of words that have similar or identical distributional properties in a sentence. These groups of words then correspond to syntactic categories. After critical comments from Pinker (1984),
Maratsos narrowed down his proposal to verbs. Maratsos (1990) suggested that, whereas an object meaning might be the best ‘central binder’ for the category of nouns, co-occurrence patterns are still the best ‘central binder’ for the category of verbs. One of the problems Pinker (1984) indicated was: how does the child know which information it has to pay attention to in performing the co-occurrence analyses? Pinker illustrated the problem as follows: the child could assume on the basis of the utterances *John eats meat, John eats slowly, and the meat is good* that *the slowly is good* is a grammatical English sentence, which it is not. Maratsos (1990) agreed with this criticism, conceding that the ‘unbiased inductive processor’ proposed in his earlier article makes it impossible to pick the right level of analysis for nouns. However, he maintained that the grammatical words preceding and following verbs should function as cues for categorization.

Brown (1957) was the first to show that children are actually able to use the morpho-syntactic environment in which a nonsense word is presented to them to infer its possible category. He showed 3- to 5-year-old children a picture of hands kneading a confetti-like material in a bowl, accompanied by a nonsense word description of the scene as either *sibbing* or *a sib*. The children who had heard the verbal form *sibbing* understood the nonsense word *sib* as referring to the action of kneading (i.e., they assigned it a verbal meaning), whereas the children who had heard the nominal form *a sib* thought the word referred to the bowl (i.e., they assigned it a nominal meaning). Children could infer the possible categories of the nonsense words on the basis of their distributional environment. The presence of the –ing ending led them to categorize *sib* as a word referring to an action, whereas the presence of the determiner *a* led them to categorize it as a word referring to an object. This study can be seen as evidence that morpho-syntactic co-occurrence information is accessible to 3-5 year-old children and that they can use this information to make inferences about properties of the nonsense words.

With the development of more sophisticated methods for testing younger children, it has become clear that children are able to segment familiar words from the speech stream by 8 months (see Jusczyk & Aslin, 1995 for English; Höhle & Weissenborn, 2003 for German). If children can segment words, the co-occurrence patterns of words can become available to them.

Transitional probabilities between syllables have also been explored in recent studies as possible indicators of word boundaries. If infants can identify word boundaries, they may be able to determine the important distributional environments indicative of categories. An early study of this kind of statistical learning by Saffran, Aslin & Newport (1996) shows that 8-month-old English-learning children are able to segment words from the speech stream solely on the basis of transitional
probabilities. The statistical properties of phonological forms help infants to detect word boundaries and so to detect the co-occurrence patterns of words. Subsequent studies show that by the age of 18 months English infants can detect not only words but also frequently-occurring relationships between two words or morphemes that are separated by other words (e.g., the relationship between auxiliary *is* and progressive ending –*ing* in English) (Santelmann & Jusczyk, 1998). German children are able to do the same around 19 months of age (Höhle, Schmitz, Santelmann, & Weissenborn, 2006). These long-distance relationships between two words in a sentence are especially well perceived by children if the intervening material is highly variable (Gómez, 2002).

These findings have important implications for category acquisition, according to Mintz (2003). He proposed that the long-distance relationships between frequently and invariably co-occurring items (‘frequent frames’) provide the crucial contexts children need for early categorization. Earlier proposals based on the relevance of co-occurring words mainly focused on ‘bigrams’, i.e., a target word X with either the preceding or the following word as a cue (e.g., Cartwright & Brent, 1997; Redington, Chater, & Finch, 1998; Mintz, Newport, & Bever, 2002). The advantage of focusing on the relationship of the to-be-categorized word with the combination of both the preceding and the following word, according to Mintz, is that it results in more accurate categories than the bigram analyses provided. He showed for English that such a mechanism would result in very reliable links between distribution and category, and proposed that this mechanism can be used in language acquisition in general.

I have investigated this proposal for Dutch. If local co-occurrence contexts such as frequent frames are indeed a general category learning mechanism, the links between the frequent frames and category have to be reliable in all languages. Mintz (2003) looked at the input to English-learning children. Mintz’s input study of English will be reported first to provide more detail about his frequent frame proposal. This study was also replicated for Dutch, which will be reported in §4.3.2. The results will show which co-occurrence information might be indicative of category in Dutch.

### 4.3.1. Mintz’s frequent frame proposal

Mintz (2003) suggests that the child only needs to focus on the local context of a lexical item to categorize it correctly. 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 with an English input study that the set
of words \( X \} \{X_1, X_2, ..., X_n \} \) within a certain frequent frame forms an accurate
category compared to the adult syntactic category.

Mintz took all the adult input utterances from six English child language corpora
in the CHILDES system in which the children were younger than 2;6. Table 4.4
presents general information about these corpora. Analyses were performed on both
tokens and types, because a token-based analysis reflects the frequency of words and
a type-analysis reflects the diversity of words. If these two measures lead to
significantly different categorization results, either the frequency of a word’s
occurrence or the number of different types of words is important for the frequent
frame mechanism to work. In §4.3.2 I will show that the difference between tokens
and types is more relevant for Dutch than it is for English.

Table 4.4. Session ranges for analyzed English input corpora (alphabetically);
number of utterances, number of tokens and types categorized by 45 most frequent
frames (Mintz, 2003: 96).

<table>
<thead>
<tr>
<th>Child</th>
<th>CHILDES sessions</th>
<th># of utterances</th>
<th>Tokens categorized</th>
<th>Types categorized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>anne01a-anne23b</td>
<td>26199</td>
<td>4389</td>
<td>405</td>
</tr>
<tr>
<td>Aran</td>
<td>aran01a-aran20b</td>
<td>20857</td>
<td>5628</td>
<td>620</td>
</tr>
<tr>
<td>Eve</td>
<td>eve01-eve20</td>
<td>14922</td>
<td>3513</td>
<td>400</td>
</tr>
<tr>
<td>Naomi</td>
<td>no1-n58</td>
<td>6950</td>
<td>1617</td>
<td>297</td>
</tr>
<tr>
<td>Nina</td>
<td>nina01-nina23</td>
<td>14417</td>
<td>6265</td>
<td>469</td>
</tr>
<tr>
<td>Peter</td>
<td>peter01-peter12</td>
<td>19846</td>
<td>5690</td>
<td>446</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>17198</td>
<td>4517</td>
<td>439.5</td>
</tr>
</tbody>
</table>

From the 17,198 input utterances Mintz extracted all trigrams, i.e., three-word
sequences. For example, six trigrams can be extracted from the utterance *do you
want to give it to me:* (1) *do you want*, (2) *you want to*, (3) *want to give*, (4) *to give it*,
(5) *give it to*, and (6) *it to me*. The middle word of each trigram was then replaced by
a variable \( X \), resulting in the frames of co-occurring words with exactly one word
position intervening. The second trigram of the example above would result in the
frame *you \( X \) to*. All frames (i.e., instances of a \( a X b \) where \( X \) is variable) in the input
speech were counted. The frame *you \( X \) to* thus occurs not only in an utterance with
*you want to*, but also in utterances containing the trigrams *you try to* or *you also to*. 
A subset of 45 most frequently occurring frames was selected for each corpus. For each of these 45 most frequent frames, all intervening words were listed, resulting in frame-based categories. The frame-based category resulting from the frequent frame you X to consists of words such as want, try, and also. These frame-based categories were further analyzed to see how they are related to the adult lexical categories of the words occurring in them.

Each type in a frame-based category was labeled for its actual syntactic category in English (verb, auxiliary, copula, noun, pronoun, adjective, preposition, adverb, determiner, wh-word, “not”, conjunction, interjection). For example, in the frame you X to, want was labeled as a verb, also as an adverb, etcetera. To calculate the accuracy of the frame-based categories, the category labels of all possible pairs of intervening items were compared to see if they counted as a ‘hit’ or a ‘false alarm’. A hit was counted for two items with the same grammatical category label, and a false alarm for two items with different labels. Accuracy is computed by the formula in (4).

\[
\text{Accuracy} = \frac{\text{Hits}}{\text{Hits} + \text{False Alarms}}
\]

The total number of hits (i.e., instance of two identical category labels) in a frame-based category is divided by the number of hits plus false alarms (i.e., instance of two different category labels). If all types within a frame-based category were from the same syntactic category, the type accuracy would be 1, since then the number of false alarms is 0; the higher the accuracy of a frame, the more consistent the frame-based category. Take, for example, the previously-mentioned frame you X to in which 2 verb types (want and try), and 1 adverb type (also) occur. The category labels of the types are now compared to each other and each comparison renders either a ‘hit’ or a ‘false alarm’. If the types want and try are compared with each other, they result in a hit, because both are labeled as verb. If want and also are compared with each other, they result in a false alarm, because they have different category labels: verb and adverb. This example results in 1 hit (comparison of want and try), and 2 false alarms (comparisons of want and also, and of try and also). The resulting accuracy of the category based on this frame would be the total number of hits divided by the total number of hits + false alarms: 1 divided by 3 = 0.33, a relatively low accuracy.

Mintz (2003) performed two different kinds of accuracy analyses. In his expanded analysis all assigned category labels were taken as separate categories, whereas in his standard analysis verbs, auxiliaries and copulas were counted as a
single category, as were nouns and pronouns. Since the number of possible false alarms with fewer categories is lower, the standard analysis always results in the same or higher accuracy measures than the expanded analysis. The mean token accuracy across all corpora was 0.98 in the standard analysis and 0.91 in the expanded. Mean type accuracy was 0.93 in the standard analysis and 0.91 in the expanded analysis. The average accuracy of the categories in the six corpora taken separately ranged from 0.80-0.98.

The most likely explanation for these high accuracy measures is that the specific distributional contexts of frequent frames yield reliable categories. To test this, Mintz calculated the accuracy of random categories for each corpus. The random categories consisted of the words from the original frame-based categories but now randomly divided across the categories. So all tokens that were categorized by the frequent frames (see Table 4.4) were randomly redistributed over 45 new, similarly sized categories. The scores of the resulting categories provide an indication of how high the accuracy is if all tokens within the frame-based categories are randomized and new categories are compiled of equal size and number. The only difference between these random categories and the frame-based categories is the frequent frame information. The average random scores for English varied between 0.23 and 0.55. Since these random measures differed significantly from the measures of the earlier analyses, Mintz concluded that frequent frames contribute important information for categorization.

All the scores of the six corpora of input to the English-speaking children Mintz analyzed are shown in Table 4.5. Each row in the table represents one of the corpora from CHILDES, listed in alphabetical order. For each corpus, the average accuracy scores in all analyses are provided. The second and third columns, for example, provide the accuracy of the categories of tokens in the standard analysis in both the frame-based condition (A for ‘analyzed’) and the baseline condition (R for ‘random’). The accuracy scores of the frame-based categories are significantly higher than baseline in all conditions. The difference between tokens and types was only significant in the standard analysis. The frequency with which a word type appears thus does not play an important role in the frame-based categories of English.

Mintz (2003) concluded that the specific distributional context of frequent frames forms a highly reliable bootstrap for categories in English. Moreover, he proposed that using distributional information for categorization could be a generally valid mechanism in child language acquisition. He acknowledged that some typological differences might affect the success of such distributional mechanisms. However, he suggested that languages with freer word order may still
show enough consistency in distributional patterns, for example in terms of frequently co-occurring morphemes rather than words. In the following section, the success of this proposal will be investigated for Dutch, a language typologically related to English. The outcomes of the study of frequently co-occurring words in the Dutch input will be reported.

Table 4.5. Token and type accuracy for standard and expanded analysis (A), including baseline accuracy of random categories (R) for all English corpora (Mintz, 2003: 100).

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Standard Analysis</th>
<th>Expanded Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tokens</td>
<td>Types</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>R</td>
</tr>
<tr>
<td>Anne</td>
<td>0.98</td>
<td>0.37</td>
</tr>
<tr>
<td>Aran</td>
<td>0.97</td>
<td>0.44</td>
</tr>
<tr>
<td>Eve</td>
<td>0.98</td>
<td>0.51</td>
</tr>
<tr>
<td>Naomi</td>
<td>0.97</td>
<td>0.48</td>
</tr>
<tr>
<td>Nina</td>
<td>0.98</td>
<td>0.48</td>
</tr>
<tr>
<td>Peter</td>
<td>0.98</td>
<td>0.49</td>
</tr>
<tr>
<td>Mean</td>
<td>0.98</td>
<td>0.46</td>
</tr>
</tbody>
</table>

4.3.2. Frequent frames in Dutch input

To establish whether frequent frames are available in the input from other languages than English, the input from Dutch child language corpora was studied, in which the method developed by Mintz (2003) was followed very closely. I analyzed the input speech to the same four Dutch children, which was available in CHILDES, as studied in the longitudinal production study reported in §3.3. They were younger than 2;6. General information about these files is presented in Table 4.6. 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. Although the total number of corpora and utterances examined by Mintz was greater than in the data available for this analysis, this is not of major consequence.

As with the English corpora, all frames in the input speech were counted and the 45 most frequent frames were selected for each corpus. The fourth and fifth column

---

16 This study was also reported in Erkelens (2008).
of Table 4.6 indicate the number of tokens and types that are contained within 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 are related to the syntactic categories of the words occurring in them. For Dutch, I used slightly different labels for these categories than Mintz did for English. Since the Dutch negative particle *niet* ‘not’, unlike its English counterpart, has the same distributional properties as all other Dutch adverbs, it was categorized as an adverb, and not as a separate category as Mintz created for “not”. A separate category for proper names was included for two reasons. They occurred very frequently within the Dutch frame-based categories and in the production data (§3.3) proper names were used differently by children than by the adults. The categories used in this Dutch input study are thus verb, auxiliary, copula, noun, pronoun, proper name, adjective, preposition, adverb, determiner, wh-word, conjunction, and interjection. All word types that occurred in the frame-based categories were assigned to one of these categories. For words in Dutch that can occur in multiple categories, the corpora were checked to see in which category the word was used in the specific frame. Just as in Mintz (2003), a standard and an expanded analysis were performed. In the expanded analysis all 13 assigned category labels were taken as separate categories, whereas in the standard analysis verbs, auxiliaries and copulas formed one category, as did nouns, pronouns, and proper names, leaving nine categories. Accuracy scores were computed in the same way as Mintz (see §4.3.1).

Table 4.6. Session ranges for analyzed Dutch input corpora (alphabetically); number of utterances, number of tokens and types categorized by 45 most frequent frames.

<table>
<thead>
<tr>
<th>Child</th>
<th>CHILDES sessions</th>
<th># of utterances</th>
<th>Tokens categorized</th>
<th>Types categorized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daan</td>
<td>daa10821-daa20625</td>
<td>13301</td>
<td>2569</td>
<td>324</td>
</tr>
<tr>
<td>Laura</td>
<td>laura01-laura22</td>
<td>8811</td>
<td>1948</td>
<td>291</td>
</tr>
<tr>
<td>Matthijs</td>
<td>mat11013-mat20619</td>
<td>16813</td>
<td>2927</td>
<td>319</td>
</tr>
<tr>
<td>Sarah</td>
<td>sarah01-sarah19</td>
<td>10710</td>
<td>2186</td>
<td>296</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>12409</td>
<td>2407</td>
<td>308</td>
</tr>
</tbody>
</table>
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 (English: 0.98) and 0.56 in the expanded analysis (English: 0.91). Mean type accuracy was 0.59 in the standard analysis (English: 0.93) and 0.40 in the expanded analysis (English: 0.91). The range of average accuracy scores for the separate corpora was 0.38-0.76 (English: 0.80-0.98). Although Mintz (2003) does not report the range of accuracy measures over the frames, such a range provides insight into the contribution of each frame to the average accuracy scores. If the range over frames is limited, each frame is a reliable categorizer, whereas large ranges indicate a high variability of informational value over frames. For Dutch, the accuracy scores of the type-based analyses ranged from 0.00-1.00 and those of the token-based analyses from 0.18-1.00. The Dutch frequent frames vary enormously in their categorization power.

The Dutch frequent frame *ik X het* ‘I X it’ provides a good example of the difference in accuracy scores between Dutch and English. The frame-based category of its English counterpart *I X it* consists solely of verbs in the input to English Peter (Mintz, 2003). However, the frame-based category of the Dutch frame *ik X het* in the input to Dutch Matthijs contains not only verbs, but also the preposition *aan* (which is often used in the construction *aan het X-en* with a meaning similar to the English progressive) and the adverbial particle *dan* ‘then’. Quite a number of these adverbial particles seem to disturb the co-occurrence patterns across frequent frames in Dutch. It is possible that this language-specific property of Dutch makes it difficult for children to trace the dependency between the framing words (see the discussion later on in §6.3). However, it is not clear whether there is an overall major typological difference causing the lower accuracy measures.

Just as in Mintz’s study, I compiled random categories in this study in which the tokens of the frame-based categories were randomly distributed across the 45 frequent frames. The accuracy measures of these random categories serve as a baseline for the informativeness of word distribution in Dutch. The Dutch results for each child are presented in Table 4.7 and a comparison of the mean accuracy results of all six English and four Dutch corpora is presented in Table 4.8. Both the numbers for the analyses (‘A’) and the randomly compiled categories (‘R’) are provided.

Statistical analysis of the accuracy measures of both English and Dutch allows us to compare 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 the expanded analysis. The same is true for the Dutch accuracy scores in both the standard analysis (tokens: t(3) = 24.6, p <
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 the difference between the two languages here. On the basis of the fact that the accuracy measures of the frame-based categories in both English and Dutch are significantly higher than baseline, frequent frames in Dutch are potentially informative with respect to the category of intervening items.

Table 4.7. Token and type accuracy for standard and expanded analysis (A), including baseline accuracy of random categories (R) for analyzed Dutch corpora.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Standard Analysis</th>
<th>Expanded Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tokens</td>
<td>Types</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>R</td>
</tr>
<tr>
<td>Daan</td>
<td>0.69</td>
<td>0.39</td>
</tr>
<tr>
<td>Laura</td>
<td>0.69</td>
<td>0.33</td>
</tr>
<tr>
<td>Matthijs</td>
<td>0.76</td>
<td>0.43</td>
</tr>
<tr>
<td>Sarah</td>
<td>0.69</td>
<td>0.38</td>
</tr>
<tr>
<td>Mean</td>
<td>0.71</td>
<td>0.38</td>
</tr>
</tbody>
</table>

However, the accuracy scores are higher for English than 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: $t(8) =$
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 for English 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: $t(8) = 6.7, p < .001$). So, although frequent frames are informative for both languages, the higher accuracy measures for English indicate that frequent frames are more informative about category membership in English than in Dutch.

4.3.3. Conclusion

Co-occurrence patterns are indicative of the category of words both in English and in Dutch. However, the specific local context of frequent frames as operationalized by Mintz (2003) provides more accurate categories in English than in Dutch. As Mintz suggested, the granularity of the frequent frame may differ across languages. The word level may be too coarse a unit in Dutch. Dutch has more morphology on verbs and nouns than English. Therefore, it seems worthwhile to consider the level of granularity that is most informative for Dutch categories. The attested differences between Dutch and English with respect to frequent word frames may be a consequence of the available material in the two languages. Suppose that the learning mechanism used by children is to search for the most frequent non-adjacent elements (i.e., frames). If this search starts from the smallest frame possible (the most local context), then Dutch children may already have a large number of morpheme frames at their disposal that provide them with morpheme frame-based categories before they ever arrive at the word frames investigated in this input study.

In order to test this suggestion, a method is needed to assess young children’s use of frames for categorization. Unlike the use of phonology for categorization by young infants, the use of frames has been investigated earlier. In the next section these studies will be evaluated to determine whether the method used there, perception experiments, would be suitable for testing whether Dutch children use frequent frames defined in terms of either words or morphemes.

4.4. Perception experiments as a method to test early categorization

Perception data allow us to study linguistic development before children actually start to talk. Over the last decades the methods of conducting perception experiments
have improved enormously; there are now suitable methods for the observation of specific aspects of linguistic comprehension in very young children. In a recent overview of research on infant speech perception, Gerken & Aslin (2005) show that techniques such as High Amplitude Sucking (HAS, e.g., Jusczyk, Rosner, Cutting, Foard, & Smith, 1977) and the Head-turn Preference Procedure (HPP, Hirsh Pasek et al., 1987; Kemler Nelson et al., 1995) have provided many insights into the phonological and lexical development of infants between 6 and 12 months old. These same techniques are currently used increasingly often to investigate the morphological and syntactic development of infants. Before reviewing perception methods as a means of studying syntactic development and, more specifically, verbs and nouns, the interpretation of data from perception experiments will be considered in general.

4.4.1. General considerations for interpreting perception data

The perception of language can be tested only indirectly, by observing children’s behavior. Speech stimuli are presented to infants while their behavioral response (e.g., sucking, looking) is measured. A dependent measure that is often used in perception experiments with young children is the duration of the behavioral response. It has to be assumed that this response (i.e., the duration of a look) is a measure of the child’s listening time to the stimulus. The growing body of evidence that head-turn experiments provide in young children confirms the likelihood of this assumption. If children differ significantly in the time they spend listening to stimuli in two different conditions, this indicates sensitivity to the variable on which the two conditions differ. Research questions that can be answered with such experiments need to focus on children’s sensitivity to a certain property of language. Perception experiments cannot be used to answer questions about the grammar that underlies children’s ability to distinguish these stimuli. For example, the question whether the child’s category of verb is defined by the presence of a specifier position, as Baker (2003) proposes, cannot be tested with a perception experiment because it is impossible to design two sets of stimuli that differ only on this particular property. However, a very carefully designed experiment can be used as a possible window on children’s sensitivity to linguistic patterns.

By controlling for factors that are irrelevant to the research question and manipulating the factors that are relevant, different details of language can be investigated separately. Perception experiments thus can provide evidence for theories about the properties of the input that play a role in the identification of categories by infants. Since the design of the stimuli determines the interpretation of perception data to a considerable extent, phonological, morphological, and syntactic
factors that might disturb the picture must be controlled for. Using nonsense words instead of real words can also control for lexical factors. In this section some of the perception studies on syntactic development and category acquisition will be reviewed to show that perception experiments provide a good method for testing which co-occurrence patterns Dutch infants use for early syntactic categorization.

4.4.2. Perception of syntactic structure

Although perception techniques were first used to study phonological and lexical development, more recent experiments also address syntactic development. The crucial difference between sounds and words on the one hand and syntactic structure on the other hand is that syntactic structure entails relationships between two or more units in a sentence. For example, languages differ in the order in which heads and complements are expressed in a sentence. This difference is traditionally called the OV/VO parameter (e.g., Neeleman & Weerman, 1997). Acquiring the directional relationship between heads and complements in a sentence is in essence a syntactic rather than a lexical development. In a perception study using the HAS procedure Christophe, Nespor, Guasti, Dupoux, and Van Ooyen (1997) showed that infants as young as 3 months distinguish the prosodic patterns of OV and VO languages. OV languages systematically stress the right part of a phonological phrase, whereas VO languages systematically stress the left part. The infants Christophe et al. (1997) tested were sensitive to these differences. They propose that children use the prosodic structure to bootstrap the place of the head in a phrase. As the authors acknowledged, this finding does not imply that children have ‘set’ the parameter for their language. However, it does tell us that very young infants are sensitive to input characteristics related to syntactic phenomena. Even before they segment words (the earliest age for which word segmentation is reported is 7.5 months: Jusczyk & Aslin, 1995), they are sensitive to prosodic characteristics indicative of syntactic structure. This means that precursors of syntactic development may be present at birth, or even before.

There are, of course, syntactic dependencies that are not adjacent, such as agreement relationships. For example, there is always at least a verbal stem – and often additional linguistic material – between the English auxiliary *is* and the progressive ending *–ing*. Using the HPP technique, Santelmann & Jusczyk (1998) show that at 18 months infants are sensitive to the syntactic dependency between the auxiliary *is* and the progressive ending, whereas they are not sensitive to an

---

17 Some researchers state that syntactic relationships like head-complement are also learned lexically first. This does not change the assumption that the relationship between heads and complements is a syntactic one.
unnatural dependency between the auxiliary *can* and the progressive ending. They are still sensitive to the dependency between *is* and *–ing* if as many as three syllables are inserted between them, but not four or more. If we compare these results to what we know about children’s production data, it is striking that children can already trace these syntactic dependencies long before they are able to produce them. The dependencies tested in these studies are investigated on the basis of concrete linguistic input (prosody and real words). The children’s sensitivity to these syntactic phenomena could stem from learned morpho-syntactic properties of the adult language such as basic word order and verbal inflection. An equally plausible explanation, however, is that they are sensitive to frequency in the linguistic input. In English, *is* and *–ing* are likely to occur more frequently together in a sentence than *can* and *–ing*. The results would then only indicate that 18-month-old children are able to detect frequently occurring invariable non-adjacent morphemes in their input.

Results from artificial grammar studies in infants support the former explanation, i.e., that 18-month-olds have indeed acquired the basic word order and verbal inflection of English. Gómez & Gerken (1999) trained 12-month-old infants with stimuli generated by an artificial grammar. Since these children had never heard sentences from this grammar before, they could not already know the frequency of certain co-occurring words. The only possible way for the children to know things about the grammar was to learn them from the few stimuli presented to them in training. The HPP experiments by Gómez and Gerken (1999) show that these one-year-olds were able to distinguish grammatical from ungrammatical word orders, even if other words were used than those presented in training. Children use the co-occurrence information from concrete linguistic stimuli to make abstractions and to learn word order rules. These outcomes make a stronger case for HPP experiments as a means to test children’s knowledge of grammar. However, in experiments with a real grammar (and stimuli), the interpretation that previous knowledge of the grammar is decisive for performance cannot be excluded.

On the basis of these studies of syntactic development it can be concluded that perception experiments are a promising method to answer questions about the mechanisms that are at work in early syntactic development.

### 4.4.3. Perception of verbs versus nouns

The relevant question for purposes of the current research is of course whether perception experiments are suitable to test the acquisition of verbs and nouns. As indicated in §1.1, different kinds of properties (morphological, syntactic, semantic) relate in different ways to these categories. Some of these properties are phonetically
perceptible in a linguistic utterance (e.g., the distribution of morphemes and words in a sentence), whereas other properties are not (e.g., the meanings of the words). In perception techniques such as the HPP, phonetically perceptible properties play a crucial role in testing, since the stimuli are presented auditorily. As a consequence, all perception studies on categorization focus on the relevant phonetically perceptible linguistic context that is needed to arrive at categorization of lexical items (e.g., their distribution) and not on the properties that are not directly perceptible (e.g., their meaning and syntactic features).

As reported in §4.3, scholars have proposed that function words that surround lexical items with especially high frequency play an important role in learning to categorize these lexical items. Theories of the relevance of co-occurrence patterns for the identification of categories have been tested in various languages. In an HPP experiment with German children, Höhle and her colleagues (Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004) tested the role of function words such as determiners and pronouns in the categorization of nonsense words. They trained children on sentences with nonsense words (glamm and pronk) modeled either as nouns (preceded by a determiner) or as verbs (preceded or followed by a pronoun). Examples of the training sentences are: 

- *zwischen den Bäumen pronk* ‘he did not pronk so often between the trees’
- *dieses Glamm war ganz bunt und wunderschön* ‘This glamm was entirely colorful and beautiful’.

In the test phase, children were presented with the same nonsense words preceded by either the indefinite determiner *ein* ‘a’, or the third person pronoun *sie* ‘she’ (*ein glamm, *sie glamm, *ein pronk, sie pronk*). The children’s listening times to the combinations consistent with the training phase (*ein glamm* if *glamm* was a noun in training, *sie pronk* if *pronk* was a verb in training) were compared to the children’s listening times to the ungrammatical, inconsistent combinations (*sie glamm* if *glamm* was a noun in training, *ein pronk* if *pronk* was a verb in training). The 12-month-old infants did not show a difference in listening times, but the 16-month-olds did: they listened longer to *sie glamm* than to *ein glamm* if *glamm* was a noun in training.

The conclusion that the researchers draw from this experiment is that 16-month-old German children are able to categorize the nonsense words as nouns on the basis of the determiners. It is questionable, however, whether this experiment actually shows that children have a grammatical representation of nouns. Children were familiarized with sentences in which the phonetically perceivable distributional cues

---

18 Note that the 3rd person singular simple past in German regularly receives an inflectional ending *-te*. Irregular verbs do not receive an ending, but only undergo vowel change and this is what the researchers tried to model here. The word form *pronk* has a phonological make-up similar to that of the past tense of some existing irregular verbs in German (e.g., *trink-/trank-* ‘drink / drank’).
pointed to the category of the nonsense words. If children use the distributional cues to categorize the nonsense words, it can be said that they have categorized the nonsense words together based on the distributional property ‘[determiner] _’, which is indicative of the nominal category in German. The fact that they are able to group new words with other words that follow a determiner in German does not necessarily imply that they have a category of nouns with any properties that go beyond that of following a determiner. It is possible that these children have a category of determiner-following words that also includes adjectives (e.g., in a sentence such as *der rote Glamm* ‘the red glamm’, the adjective *rote* obeys the criterion of the ‘[determiner] _’ category). What this experiment clearly does show is that 16-month-old German infants use determiners to identify this construction.

Mintz (2006) conducted a similar experiment with English-speaking children to test his proposal about the role of frequent frames in learning to categorize (§4.3.1). This experiment showed that it is possible to test the use of this cue by infants, which is especially relevant for the present study on Dutch. He taught English-speaking children four nonsense words (*deeg, lonk, gorp*, and *bist*), embedded in the specific distributional contexts of frequent frames. In a familiarization phase, 12-month-old children heard two of the nonsense words in four frames (embedded in longer utterances) that host verbs (*you_the, to_it, I_you, can_#*). The other two nonsense words were presented in four frames that host nouns (*the_in, your_#, his_on, a_of*). Every nonsense word was thus presented in four different frames, all either nominal or verbal. A word familiarized in nominal frames is a nonsense noun, and a word familiarized in verbal frames is a nonsense verb. In the test phase children heard the same nonsense words in either a similar category frame (i.e., a nonsense verb in another verb frame) or in the other, inconsistent frame (i.e., a nonsense verb in a new noun frame). For example, if the nonsense verbs were *deeg* and *lonk* and the nonsense nouns *gorp* and *bist*, a child heard in the familiarization phase ‘*can you deeg the room*’, ‘*she wants to lonk it*’, ‘*I see the gorp in the room*’, and ‘*here’s a bist of a dog*’. For this child, consistent sentences in the test phase would be ‘*can you lonk the room*’ and ‘*I see the bist in the room*’ and inconsistent sentences would be ‘*she wants to gorp it*’ and ‘*here’s a deeg of a dog*’. Mintz compared the listening times to the consistent and inconsistent sentences. For sentences with a verb frame, the 12-month-old children listened longer to the inconsistent sentences than to the consistent ones. This means that they were sensitive to the inconsistent pairing of nonsense nouns with verbal frequent frames. For the sentences with noun frames, he found no differences in listening times. To illustrate with our examples children thus listened significantly longer to ‘*she wants to gorp it*’ (*gorp* was familiarized in a noun frame) than to ‘*she wants to deeg it*’
(deeg was familiarized in a verb frame) but they did not listen longer to 'here's a deeg of a dog' (deeg was familiarized in a verb frame) than to 'here's a bist of a dog' (bist was familiarized in a noun frame). The fact that the children discriminated 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 used in the test phase frames before, but they still differentiated between the verbal and nominal frames, at least for the nonsense words modeled as verbs. Mintz suggests that differences of frequency in the input explain why the infants did not distinguish between test conditions for the nonsense words modeled as nouns. Whether these 12-month-old English-learning children also have a grammatical representation of the category verb cannot be assessed with this kind of experiment. However, this experiment does establish that these English-learning infants used frequent frames for categorization.

4.4.4. Conclusion

Perception experiments can be used to study the categorization abilities of pre-verbal infants. The results of previous experiments primarily provide information about the co-occurrence properties of the input children use in the process of categorization. Such co-occurrence patterns can be modeled in detail and irrelevant factors can be controlled for. If designed properly, perception studies thus provide informative data on the mechanisms that are at work in the process of category learning. Especially Mintz’s (2006) experiment suggests that an HPP experiment could be a good method for testing whether Dutch infants use frequent frames for categorization. The next chapter reports two HPP experiments that investigate the use of frequent frames by Dutch infants.

It is not clear that the findings from perception studies on German and English can be automatically applied to other languages. There are cross-linguistic differences in the age of production of determiners (see Rozendaal & Baker, 2008 and references therein), which might mean that there are also cross-linguistic differences in the age of perception. French children are faster at producing determiners than English children, and English children are faster than Dutch children. At least in part, this is a consequence of the use of determiners in the input of these three languages. Rozendaal and Baker’s study of the input to French, English, and Dutch children reveals not only that Dutch has fewer obligatory contexts for the production of determiners, but also that Dutch adults omit

---

19 Two of the input corpora analyzed by Baker & Rozendaal (2008) were also used for the input study reported in §4.3.2 here, namely those of Matthijs and Sarah.
determiners in obligatory contexts in their speech to children more often than English and French adults do. As a consequence, Dutch children probably have a harder time detecting those elements in the input and using them in early categorization. Although Höhle and Weissenborn (2003) have shown that the perception of function words in German seems to be parallel to that in English, it cannot be assumed that co-occurrence cues are accessible to Dutch children as early as they are to English children. It is a question for further research to determine whether there are differences across languages in the usefulness and accessibility of co-occurrence properties as indicators of category membership.

A clearly defined co-occurrence pattern such as a ‘frequent frame’ can be incorporated into stimuli presented to children in a perception experiment. However, the stimuli have to be modeled carefully in such a way that the co-occurrence information is the only information that differs between variables. As we have seen in this chapter, the phonological form of words and the syntactic structure of the sentence also provide information about the category of the word. The influence of the phonological form can be controlled for by taking word forms that are equally likely to be phonological nouns, or phonological verbs in the language (see §4.2). The influence of the syntactic structure can also be controlled for by keeping the properties of the wider syntactic environment similar in both conditions of the experiment.

4.5. Conclusion

The experimental study with Dutch adults reported in §4.2 showed that the phonological form of Dutch words is indicative of their category: Dutch adults were able to categorize nonsense stems as nouns based on their phonological make-up. To test whether children also use this cue in early categorization is more difficult. It is not possible to use the design of the reported experiment with young children. The first problem is the presentation of nonsense stems in isolation. If children hear a nonsense word without any morpho-syntactic context, the task is too complex and unnatural for them to provide reliable results. Second, children of course cannot be asked to tick a mark for verb or noun, so the method of collecting responses has to be adjusted. It is very difficult to design the categories verb and noun as answer options. One way to do this is by using a picture selection task. The method used by Cassidy & Kelly (1991) can be seen as a version of such a selection task. They used a video in which the action stood for the verb answer and the object for the noun answer, the child had to select one of the options. Such a method, however, can only be used with children of at least 2 years of age (Gerken & Shady, 1996). Since the
production studies in Chapter 3 showed that 2-year-old children are already adult-like in their categorization, a different method such as a perception experiment is needed to test whether infants use phonological properties as indicators of category at earlier stages of development.

The input study of frequent frames reported in §4.3 revealed that frequent word frames are available as indicators of category in Dutch. The resulting frame-based categories are, however, less accurate compared to the adult lexical categories than they are in English. It was therefore hypothesized that frequent morpheme frames may prove to be better indicators of category in Dutch than frequent word frames.

The method used in the perception experiments discussed in §4.4 proved to be suitable for testing categorization in pre-verbal infants. The HPP has already been employed to test the use of frequent word frames by English 12-month-olds. In principle, the method is also suitable to test the use of phonological form as category-indicative property. However, it is more logical to start out by replicating the experiments on co-occurrence patterns than to design an experiment for testing the phonology. Therefore, the use of frequent frames will be tested first. The experiments on Dutch infants’ use of these frames are described in Chapter 5. The predictions for the experiments are drawn directly from the comparison of the input studies on frequent frames in English and Dutch discussed in this chapter. The studies described in this chapter thus have shown that it is possible to study categorization in younger infants by investigating their use of category-indicative input properties such as phonology and co-occurrence patterns. In the next chapter the use of the category-indicative property ‘frequent frames’ by Dutch infants will actually be tested.