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.
1 Introduction

In 1947 Reichenbach set a problem to his students of interpreting the following sentence: “John where Jack had had had had had had had had had had had a better effect on the teacher” (Reichenbach, 1947: Exercise 3-4, p.405; solution p.417). The eleven occurrences of the word had can only be construed as an acceptable sentence if the uses are distinguished in terms of word category, as shown in the punctuated version: John, where Jack had had "had," had had "had had"; "had had" had had a better effect on the teacher. In order to properly understand sentences, the linguistic categorization of their contents is indispensable. Categorization therefore is a cognitive skill that is crucial not only to understanding, but also to producing language. For instance, adult speakers are aware of the different uses of a word such as destroy in comparison to the uses of a word such as apple. Speakers of English can say destroy an apple, but they cannot say apple a destroy. The different ways these two kinds of words are used reflect the fact that they belong to different categories. Traditionally, destroy-like words are called verbs, and apple-like words are called nouns. Since categorizing verbs and nouns is crucial for understanding and producing language, acquiring these categories is integral to language acquisition in general. The focus of this thesis is therefore on when and how children acquire these lexical categories. The studies will be limited to verbs and nouns, since it is assumed here that the results can be generalized to other categories as well.\footnote{Other lexical categories are adjectives and, according to some researchers, manner adverbs and prepositions.}

This chapter first introduces the notions relevant to this topic: verbs, nouns, and the development of categorization in children, followed by the research question and the outline of this thesis.

1.1. Linguistic properties of verbs and nouns

At first sight, the categorization of content words into verbal and nominal categories seems quite straightforward. There is a ‘common sense’ consensus on the categorization of words such as apple as nouns and words such as destroy as verbs, roughly based on their meaning, namely apple refers to an object, destroy to an action. Furthermore, we use knowledge about the category of these object and action words as a basis for producing and interpreting sentences. If we learn and categorize novel words, we are able to apply all the properties of the category to the novel word.
at once. For example, if an unfamiliar object is named dack, we are able to produce and interpret a sentence such as here are two dacks. We unconsciously interpret a novel word for an object as a noun and assume that all morphological and syntactic properties that are related to the category noun are applicable to the novel word (e.g., taking the plural marker -s, modification by two, and functioning as the head of the noun phrase two dacks). A similar story holds for novel action words. If an unfamiliar action is named gorping, we assume that to gorp is a verb and we are able to produce and interpret sentences such as he gorps (where the novel word is the head of the verb phrase with he as its subject and is marked accordingly with a third person singular -s). We use our knowledge of the categories verb and noun to apply all kinds of linguistic properties to new members of these categories. Knowledge of the lexical categories verb and noun seems basic and can be used to determine the combinatorial possibilities of a given new word. Therefore, knowledge of these categories early in the course of acquisition would be helpful for children in categorizing new words in grammatical sentences.

Even though verbal and nominal categories are intuitively straightforward, the precise linguistic details are not so clear. A diverse collection of linguistic properties correlates with these categories but there is no agreement as to which of these properties actually define verbs and nouns and which are only indicative of verbs and nouns. For a mathematical category such as ‘square’, the defining properties are fairly straightforward: we can categorize a given shape as a square if it has the following three defining properties “A 4-sided regular polygon with all sides equal and all internal angles 90°”. These properties are necessary and sufficient for ‘squarehood’: if a figure has these properties, it is a square; if it does not, it is not a square. Unfortunately the lexical categories verb and noun do not have such necessary and sufficient defining properties that linguists agree upon. This is not to say that adult speakers do not have clear intuitions about verbs and nouns. They can in fact use a number of properties as an indication of which category the word belongs to (e.g., the word’s meaning, the morphemes attached to a word, the position of a word in a sentence). However, this does not mean that these properties define the category. All squares, besides the defining properties, also have other properties that are indicative of ‘squarehood’. For example, they have a measurable surface and they have parallel sides. These properties are applicable to ‘squarehood’, but do not define it, since, for example, there are far more types of objects with measurable surfaces than squares. Such properties that are common to a category, but do not necessarily define them, will be called indicative properties from now on. For verbs and nouns there is no agreement on which of the properties that are
indicative of their status are necessary to define them. In Chapter 2 different attempts to define verbs and nouns will be discussed.

For now a number of properties indicative of verbs and nouns will be considered briefly to obtain an idea of the difficulty of understanding their nature. For example, the word forms *apple and *destroy display different combinatorial properties at three levels: the sentence, the phrase, and the word. The English form *apple can be the head of the argument phrase (sentence level, 1a), can be preceded by a determiner (phrase level, 1b), and can be inflected for number (word level, 1c). The form *destroy cannot be used in any of these combinations, see (2).

(1) a. I [see [an apple]NP ]VP  
b. the / an apple  
c. two apple-s

(2) a. *I [see [a destroy]NP ]VP  
b. *the / *a destroy  
c. *two destroy-s

In turn, *destroy can be the predicate (3a), can be combined with a subject that determines its form (3b), and can be inflected for tense (3c). These are all combinatorial properties *apple cannot have, see (4).

(3) a. I [destroy [furniture]NP ]VP  
b. he destroy-s  
c. destroy-ed

(4) a. *I [apple [furniture]NP ]VP  
b. *he apple-s  
c. *apple-d

The morphemes, words, and phrases with which the target words *apple and *destroy can be combined are all category-indicative properties. However, they are not necessarily defining properties of the categories verb and noun. Some forms, such as English *walk, have both the verbal and the nominal combinatorial possibilities mentioned above, see (5).

(5) a. I had [a nice walk]NP  
b. the walk (in the country)
Depending on the combination in which *walk* is used, it can be interpreted as either a noun or a verb. That the context of a sentence is needed to determine the category of a form is not unusual in English nor in Dutch. But it is extremely common in languages such as Samoan, Tagalog and Mundari (Hengeveld, Rijkhoff, & Siewierska, 2004). For example, in Samoan the form *lā* ‘sun’ (see 6) can be interpreted both as noun (6a) and as verb (6b) depending on the combination of words.

(6) a. ‘Ua mālosi le lā.
   PERF strong ART sun
   ‘The sun is strong.’

b. ‘Ua lā le aso.
   PERF sun ART day
   ‘The day is sunny.’

(Mosel & Hovdhaugen, 1992, ex. (4.44): 80)

Both (6a) and (6b) have the same morpho-syntactic structure: a predicate phrase marked with the perfective marker ‘ua, followed by an argument phrase marked with the article *le*. The same word form *lā* can be used both as predicate (6b) and as argument (6a), without any changes in the morpho-syntactic structure of the sentence. The different interpretations of the word form are determined by its position in the sentence structure. This raises the question whether the verbal and nominal properties are connected to the word forms at all. Possibly it is the morpho-syntactic environment in which the form is used that determines the category of the word form.

There are two different views of these uses of word forms in multiple categories. A common morphological analysis of the multi-categoriality of *walk* is that the nominal form *walk* in (5a-c) is derived from the verbal form *walk* in (5d-f) by means of a derivational operation. In the same way, the nominal form *growth* is derived from the verbal form *grow* (7a). The difference between these two derivational operations is that the operation for *grow-growth* is overtly marked with the morpheme –*th*, whereas the *walk-walk* operation is not overtly marked, but has a zero-morpheme –*ø* in (7b).
This latter form of derivation by means of a zero-morpheme is called zero-derivation or conversion (Kiparsky, 1982; Don, 2004). The basic assumption underlying such a derivational analysis is that words are specified for category in the lexicon, for example apple is specified as a noun (see Lieber, 1981; Di Sciullo & Williams, 1987; Lieber, 1992). This view of categorization is called lexicalist, because it assumes that category is a lexical property. Since apple is a noun, the constituent of which it is the head is a noun phrase.

Several linguists have challenged this lexicalist view of categorization. They propose that categorization is a property of the wider linguistic environment in which a word is used (so apple is the head of a nominal phrase because it is inserted in a nominal environment) rather than a property of the word itself (e.g., Hengeveld, 1992b; Goldberg, 1995; Marantz, 1997 - §2.2, §2.4, and §2.5). This second kind of analysis means that Samoan lā in (6b) is not the result of a derivational process in the lexicon. It is the syntactic position in the predicate phrase preceded by 'ua that marks the item as verbal. On its own, the item is category-less. It is the skeleton of the sentence that is marked for categories, not the word forms that are used in it. Note that the English data in (1)-(4) then need a different explanation; in these theories the ungrammaticality of (4a) cannot simply be attributed to the fact that apple is a noun. In fact, Borer (2003) would claim that this sentence is grammatical, but is simply not used in conventional English. A number of linguists with quite different theoretical backgrounds are in agreement on this point, i.e., that category is a property of the morpho-syntactic structure. But the debate has not been resolved.

In sum: verbs and nouns have a number of recognizably different properties that enables language users to categorize them. However, defining verbs and nouns is not straightforward. The study of how children learn to categorize verbs and nouns can help to illuminate the relation between all properties indicative of category and their linguistic definition. In the next section, different facets of how children learn to categorize will be considered.

1.2. Categorization in development

Categorizing linguistic elements is part of the task children face in learning to categorize the world around them. In the literature on general cognitive development, much attention has been paid to the question of how infants learn to
categorize. Mandler (2000) describes the earliest stages of categorization by infants as a joint force of perceptual and conceptual categorization. Children categorize things together if they are perceptually similar. For example, children already categorize different dogs together in one category and different cats in another category when they are 3 months old (Quinn, Eimas, & Rosenkrantz, 1993). The basis of this categorization is the similarity of facial features between all dogs, being distinctively different to the facial features of cats (Quinn & Eimas, 1996). Perceptual categorization allows infants to identify different categories in the world. The identification of category does not automatically provide a meaningful interpretation of that category. Having a perceptual category of all animals with dog-like facial features does not automatically provide the child with information about the nature of dogs. To learn about the nature of dogs, the child needs a conceptual category, a mental representation of the category DOG. Mandler & McDonough (1993) show that infants as young as 9 months categorize animals different from vehicles, even if the exemplars of different categories were perceptually similar. Researchers differ in their opinion about whether perceptual categorization precedes conceptual categorization or vice versa. Mandler (2000) proposes that both processes are at work at the same time. In essence, learning to categorize consists of two different tasks: learning to identify the relevant categories (constructing perceptual categories) and learning to understand the nature of these categories (constructing conceptual categories).

If we assume that these general categorization processes also hold for linguistic categories, children not only have to learn to identify verbs and nouns as different categories (perceive verbs and nouns differently), they also have to learn what verbs and nouns are (use verbs and nouns in an adult-like manner). All linguistic properties that are indicative of category probably trigger categorization in children. Eventually, however, children have to learn which of these indicative properties constitute the defining properties of verbs and nouns to acquire a mental representation of verbs and nouns.

The interplay between cognitive and linguistic categories in the acquisition of word meaning has been the subject of much earlier work on the acquisition of verbs and nouns. Studies on lexical development in a number of different languages such as Hebrew, English, German, Turkish, Japanese, Kaluli, and Mandarin Chinese have shown that the majority of words in early child vocabularies are object words (‘nouns’) rather than action words (‘verbs’) (Gentner, 1982; Dromi, 1987; Clark, 1993). This observation has led to a so-called ‘noun-verb debate’ (Imai, Haryu, Okada, Lianjing, & Shigematsu, 2006), which asks where meaning comes from: whether the conceptual categories underlying word meaning are determined by
INTRODUCTION

general cognition or shaped by linguistic categories (see Bowerman, 2000 for an overview of the discussion on this issue). The outcome of the debate on noun and verb categorization seems to be that the acquisition of verbs (and grammatical words) is predicted to be more prone to language-specific influences than the acquisition of nouns. A number of studies have shown that certain relational and event concepts are indeed influenced by language in the sense that the conceptual categories of speakers differ depending on their first language (e.g., conceptual categories of ‘space’: Bowerman & Choi, 2001; conceptual categories of ‘cutting and breaking’: Majid, Staden, Boster, & Bowerman, 2004). The extent to which object concepts are different from these relational and event concepts is a matter of current research (e.g., Malt, Sloman, & Gennari, 2003).

In the noun-verb debate, the category labels verb and noun are roughly used for ‘words for actions’ and ‘words for objects’. However, as already briefly illustrated in §1.1, there are more properties associated with verbs and nouns than just meaning. Furthermore, ‘action’ or ‘object’ meaning is not a reliable property for defining the nature of the categories. Not all words that behave as nouns in sentences refer to objects (pain, crisis in (8a)) and not all words that behave as verbs refer to actions (know, love in (8b)).

(8) a. I [see [the pain (in your eyes)] NP ]VP
    I [see [the crisis (in the world)] NP ]VP

b. I [know [that song] NP ]VP
    I [love [that song] NP ]VP

That ‘having an object meaning’ is an indication of nounhood and ‘having an action meaning’ an indication of verbhood does not mean that object words and action words are nouns and verbs. Meaning can be used to identify different categories: a category of action words and a category of object words. However, the ultimate representation of verbs and nouns involves more than just meaning.

The present study will not tackle the subject of linguistic categorization from the perspective of the learning of word meaning. It focuses in the first place on the nature of the structural properties of verbs and nouns to inquire how children learn what these categories are. Furthermore, the role of category-indicative properties other than semantics will be studied. The next section specifies the approach taken in this thesis by presenting the research question and the outline of the study.
1.3. Research question and outline of the study

The two kinds of categorization described in the previous section, perceptual and conceptual categorization, serve as the basis for the research strategy used in this study. In order to categorize verbs and nouns, children need to acquire a mental representation of these categories and they need a perceptual trigger to identify the different categories in speech. Both processes take place in the child during acquisition, but not necessarily at the same time and not in the same way. This study focuses on how Dutch children learn to categorize verbs and nouns and thereby aims to provide insight into both the mental representations of verbs and nouns and the process of learning to identify verbs and nouns in the speech stream. The central research question is stated in (9).

(9) When and how do children learn to categorize verbs and nouns?

As will be shown in Chapter 2, different hypotheses exist among linguists as to what the defining properties of verbs and nouns in the mental grammar can be. To decide which theory is most compatible with child language production data, three facts about the theories have to be established. The first is self-evident: for a meaningful comparison of theories, the theories have to have different ideas about verbs and nouns. In Chapter 2, six theories of the representation of verbs and nouns in the mental grammar will be presented and evaluated with respect to their predictions regarding learnability. These theories were selected because they focus on syntactic properties. They come from diverse traditions, namely the generative (Marantz, 1997; Baker, 2003; Borer, 2003), the functional (Hengeveld, 1992b), and the constructionist (Croft, 2000; Goldberg, 2006) tradition. Secondly, it is crucial whether these proposals can be translated into predictions so that their compatibility with child language data can be tested. The gap between a theory of grammatical structure and predictions for the acquisition of this grammar might be difficult to bridge. Therefore, special attention is paid to how the theories can be translated into predictions for child language data. Both the theories and their predictions will be discussed in Chapter 2.

Chapter 3 handles on the third fact to be established, namely whether it is possible to test the predictions on child language data. Since child language utterances do not come with flags on them telling what the structure of the utterance is, we need other methods leading to an interpretation of the data. The decision as to which theory is most compatible with child language data requires appropriate methods for obtaining and interpreting child language data. In Chapter 3 the predictions from the theories will be tested on Dutch children’s production data.
Special attention will be paid to the interpretation of longitudinal corpus data. The results of these production studies reveal how Dutch children’s representations of verbs and nouns develop over time and what these representations consist of. As will become clear in Chapter 3, the production data show that children apparently have already learned to successfully categorize most of their vocabulary before they produce sentences (sentence production occurs around two years of age). So there seems to be some mechanism available for categorizing at a very young age.

Chapter 4 investigates the category-indicative properties of the input available at an early stage that potentially play a role in the early categorization mechanism operating in Dutch children. The focus will be on the perceptual properties that are phonetically perceivable from the speech stream, namely, phonological properties and distributional co-occurrence patterns. Children are confronted with all the category-indicative properties available in their input. In order to find out whether children are able to detect these indicative properties and use them for categorization, the predictions from one of the input studies in Chapter 4 are tested.

Chapter 5 presents the results of two perception experiments testing the predictions from the input study. The specific property focused on is the frequent co-occurrence of two elements in language, so-called ‘frames’, that can be used as indicators of the category of the form that falls between these two elements. The use of such frames by Dutch children will be investigated and compared to their use by children from earlier studies of English. The results of these perception studies provide insight into the mechanisms that play a role in identifying categories from the speech stream.

The results of the production studies from Chapter 3, the input studies from Chapter 4, and the perception studies from Chapter 5 together provide a picture of when and how Dutch children learn to categorize verbs and nouns. This picture will be summarized in Chapter 6. The outcomes of this study show that children start learning to categorize verbs and nouns at an early stage of language development and that they use language-specific properties from the input to arrive at these early categories. The discussion in the final chapter will focus on the learning mechanisms used in early categorization and the representation of verbs and nouns in the adult grammar.