Complex lexical items

Mos, M.B.J.

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Complex Lexical Items

Published by
LOT
Janskerkhof 13
3512 BL Utrecht
The Netherlands

phone: +31 30 253 6006
fax: +31 30 253 6406
e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: http://www.wordle.net

ISBN: 978-94-6093-026-3
NUR 616

Complex Lexical Items

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Tilburg,
op gezag van de rector magnificus,
prof. dr. Ph. Eijlander,
in het openbaar te verdedigen
ten overstaan van een
door het college voor promoties
aangewezen commissie
in de aula van de Universiteit
op woensdag 12 mei 2010
om 16.15 uur

door

Maria Baukje Johanna Mos

geboren op 11 september 1978 te Amsterdam

Promotor:	prof. dr. A.P.J. van den Bosch
Copromotores:	dr. A.M. Backus dr. A.R. Vermeer
Promotiecommissie:	prof. dr. G. Booij dr. E. Dąbrowska prof. dr. D. Sandra prof. dr. D. Schönefeld

Acknowledgements

As a toddler, I went through a crabby period, so I'm told. I wanted to tell people things, but couldn't. Then one day, after going to the park with my father to feed the ducks, I managed to explain to my mother where we had been. It worked! Reportedly, I was so happy I strutted around the house with a proud swagger for days. That's how important language is.

Growing up around family members who love to read and talk helped me develop my passion for language, as did going to great schools with inspiring teachers. I found out there was such a thing as linguistics and was fortunate enough to study at a great program at the University of Amsterdam. While finishing my Master's thesis, I got a Ph.D. position at the Department of Language and Culture Studies at Tilburg University. It is there, that I wrote the book you now have in front of you.

As diverse as our group of Ph.D. students was, it has been a pleasure to have you as my colleagues, Dorina, Geke, Kasper, Max, Mohammadi, Nadia, Sander, Serpil and Seza. You, and all the other MMS/ICC/T&C/FGW-members (Jeanne, Joost, Véronique and many others) have made my work there thoroughly enjoyable. Outside of the UvT, Anéla helped me keep in touch with fellow Dutch applied linguists. The Anéla board members also happen to be very nice people; I am happy to count myself among you.

I feel I was especially lucky in having Elma working across the hallway from me. Whether is was the soup brigade at lunchtime, or travels to conferences and workshops in Barcelona, Antalya, München and Bergen, we did it together and it is fitting that we're even defending our theses on the same day –thanks for your calm support.

Having three supervisors may sound like a bit much, but in my case they have complemented each other beautifully. Straight-talking Anne was always willing to be convinced to change his mind if my arguments were good enough. Ad's bigger picture and smart thinking tended to make me leave his office with a better question than the one I had when I came in. Antal didn't just step in as promotor, but added his computational expertise and, more importantly, his enthusiasm. All three of you have supported me in the choices I made. Clearly, there would not be a dissertation without you guys. You have truly been my A-team.

The manuscript I submitted was improved significantly because of the comments made by the defense panel. I am grateful to Geert Booij, Ewa Dąbrowska, Dominiek Sandra and Doris Schönefeld for agreeing to be part of this panel. Roel Tanja and Anneke Smits are responsible for making the final product look this good –although I claim full responsibility for all remaining errors.

While I was happy to work in Tilburg, friends and family kept me living in Amsterdam. Micha will always be my *broertje*, but he has also become a great friend, whose food really nurtures the weary linguist. Not many linguists have a mother who knows what a *hapax* is, but I do, and I am very proud of her. Whenever I need her, she is always there for me. Ditske, I wouldn't miss our vacations together for the world, and I hope we'll have many more!

Regardless of whatever other merits it may have, studying Spanish is an excellent way to make friends. I have shared houses, meals and many a hearty laugh with you, Guiris! When I couldn't stop talking about *–baar*, you firmly put my feet back into my Havaianas. Any linguistic frustrations I had could be soothed with a glass of wine or two. I am happy to see our circle grow, with a new generation arriving, and plan to keep sharing meals for a long time to come. Anouk, Jeroen & Martine, Kieran, Linda & Jokim, Mascha, Milan & Annechiene, Myra & Tom, Niels, Paul & Ellen and Rogier & Elisa: bedankt! Linda and Anouk: thank you for agreeing to come all the way to Tilburg to be my *paranimfen*.

I don't know what is going to happen next. All linguistics papers I ever wrote deal with continua in some way, CLI's being no exception. Let's hope that my career in linguistics is far from over, and will continue –wherever it may lead me.

Maria Mos,
Amsterdam, March 2010

Contents

Chapter 1:

Complex lexical items: An introduction

1.0 Introduction

This thesis is about *complex lexical items*: chunks of language which contain more than one element in form and meaning. The main question I will address is what people know about complex lexical items and how they use this knowledge. I argue that language users' knowledge is represented at lexically specific as well as more abstract levels. This first chapter sketches a path describing the development of this knowledge. Employing a usage-based approach to language acquisition, this development is presented as a bottom-up process which starts with concrete form-meaning pairings as the basis for more abstract knowledge. This model of multiple representations is then tested in experimental research (Chapters 2 through 4) and evaluated in the final part of this book.

As a background to the concept of complex lexical items, the first two sections of this chapter give an overview of research on morphologically complex words and multi-word units. Many complex lexical items (hereafter CLIs) are complex words such as compounds and derivations. In linguistics, these morphologically complex words traditionally fall under the heading of morphology. The first section of this chapter places the current research in this tradition. I cannot be exhaustive in the portrayal of morphological research and focus instead on showing that researchers' theoretical approaches and research goals determine to a large extent which morphological processes are studied, and what aspects of these phenomena are highlighted.

Morphology is usually restricted to single-word phenomena, but it has been observed by many linguists that people also use larger combinations as fixed units or lexical chunks. For me, these larger units are CLIs as well. Some of the research on larger combinations of form and meaning for which we have evidence that they are stored is discussed in Section 2.

The third section provides a definition of *complex lexical items* and discusses their complexity and unit status. From this definition it follows that CLIs include more than most morphologists would gather under the notion of morphology. Contrary to some theories, I do not claim that morphology is an autonomous module of language, which functions differently and with other mechanisms than lexical or syntactic phenomena. I will argue instead that knowledge of complex lexical items extends from specific (lexical) representations to general (abstract, syntax-like) patterns.

This is in line with the Construction Grammar approach (e.g. Goldberg, 1995, 2006; Croft, 2001; Boas, 2003) and, more generally, with assumptions shared by cognitive linguists (Langacker, 1987, 1991, 2008; Tomasello, 2003) who reject a strict dichotomy of lexicon and syntax. Section 4 argues how these approaches can accommodate CLIs as part of the linguistic system in general and in speakers' inventory of language structures, or *constructicon*.

Using the insights and conventions of these approaches, it is then possible to sketch the development of knowledge a language user has about a CLI. The path of acquisition, described in Section 5, results in representations which are lexically specific, thus allowing for the storage of information about individual CLIs, alongside more general or abstract representations reflecting the productivity of some patterns. I demonstrate how the findings reported on in Sections 2 and 3 can be accommodated in this model of development and knowledge representation, which will be referred to throughout this book as the MultiRep (for 'Multiple Representations') model.

This chapter concludes with a look ahead at the following chapters and outlines how the different experiments and case studies introduced there can shed more light on the principal question: how is language users' knowledge of CLIs represented and in what context do they call upon which aspects of these representations? More specifically, the experiments described in Chapters 2-4 focus on the developing linguistic repertoire of children in the second half of primary school (ages 9-12). Different experimental techniques are used in order to tap into the use of representations in a variety of task demands. I will argue that these demands influence the type of representation used. The results are contrasted with adult task performance, corpus analyses and frequency-based measures, and are discussed in relation to the MultiRep model.

1.1 Research in morphology

Within linguistics, morphology assumes a position somewhere in the middle of everything. Morphology deals with the structure of words and the concatenation of morphemes. Typically, a morphologist is interested in productive processes. A description of the distribution of morphemes also tends to show that a lot of combinations are fixed (they occur frequently, often with a specialized meaning or phonological contour) or are dispreferred (they occur very rarely). The existence of fixed combinations points at the absence of productivity: storage of these combinations in the inventory of lexical units. In addition to the storage-productivity issue, morphologists will not ignore the fact that morphologically complex words contain morphemes that are used in other contexts. In order to be able to account for such patterns, they will have to incorporate a theory about the structure of the constructicon of speakers, including representations of productive morpheme constructions. Also, the morphologically complex words themselves do not occur *in vacuo*: they are part of phrases, utterances and discourse. For that reason, the study of morphology has to look both at simple μορφ-s (Greek for *form*) and at the bigger context. The storage-productivity question, the linking of morphologically complex words to other, similar, words and to their syntagmatic context entail that lexicon and syntax, knowledge representation and the development of that knowledge in language acquisition are all aspects a morphologist has to contend with.

Following this reasoning, it could be regarded as surprising that not all linguists are fascinated by morphology and make it a central element of their scientific research. In part, this is caused by historical accidents: each subfield of linguistics has its own traditions and topical issues that are studied. Many linguists do, however, study morphology, although they are likely to concentrate on only one of the aspects described above. The remainder of this section describes some types of morphological study, to show how researchers' assumptions determine the phenomena they choose to study and the aspects they focus on as well as the limitations that choice entails. I hope to convince readers that in studying these kinds of linguistic phenomena, they are dealing with some of the most interesting aspects of language: storage and productivity, meaning and form, use and knowledge. Later in this chapter, in Section 5, I introduce a model that can accommodate most if not all of the findings brought forward.

1.1.1 Morphology in theoretical linguistics: Productive processes

In theoretical linguistics within the Chomskyan tradition, morphology has been described as a more or less separate component of the grammar (e.g. Aronoff, 1976, 1994). In the Distributed Morphology approach described in Halle and Marantz (1994), Halle and Marantz speak of "syntax proper" and place morphology in a hierarchical tree below syntax but before vocabulary insertion in an incline that results in PF (Phonological Form). Their main argument for doing so is the observation that morphological operations are constrained by strict local syntactic conditions. In this framework, morphology operates before positions are lexically filled. This means that this theory does not have to account for irregular forms, which are handled by the morphological features in the specific vocabulary items. Throughout their article, the linguistic data Halle and Marantz describe concern inflectional paradigms (case systems, person marking on the verb etc.). The concept of an ideal speaker/hearer in combination with a main interest in highly frequent, basic processes almost naturally guides this kind of research to postulate very abstract, general rules with a large scope. These rules are taken to be part of people's language competence, and deviation from them is due to performance errors and/or the storage of individual items, which, as they are stored, are not linked to this rule. Phenomena such as individual variation between speakers and small-scale patterns with limited productivity do not fall within the scope of this kind of research. It is not a coincidence that this model, which is explicitly linked to a model of syntax, emphasizes inflection to the point where other morphological processes like derivational word formation are left unanalyzed: English inflectional morphological processes are a clear-cut example of very general and regular phenomena (cf. Barðdal, 2008 for a discussion of generality and regularity as aspects of productivity).

Aronoff (1994:166) explicitly states in the conclusion to his book *Morphology by itself* that he *"set out over the last few years to uncover morphological generalizations that are not plausibly analyzed as something else"* and discusses these generalizations in a framework with a large degree of autonomy for morphology. The autonomy of morphology is an oft-returning topic in morphological theory. In my view, the existence of morphological generalizations is not sufficient evidence to stipulate the autonomy of morphology. Even if generalizations could be found which have no relation to systematic semantic differences or to phonological processes present in the language, the fact remains that in the overwhelming majority of cases, a morphological generalization does correspond with a semantic difference. Throughout this thesis I will attempt to show that morphological constructions, their acquisition and processing can be described using templates for constructions that are at work in all of the linguistic system. In that fundamental sense, there is no difference between morphological constructions and, say, argument structure constructions: morphological and syntactic generalizations really are isomorphic, but at different scales. For me, this is sufficient reason to reject the notion of morphology as a completely autonomous component. Anderson (1992) also repudiates such modularity in his book A-morphous morphology but for very different reasons. He does maintain a strict distinction between lexicon and syntax, in his formulation of the lexical integrity hypothesis: *"the syntax neither manipulates nor has access to the internal structure of words"* (Anderson 1992:84) and also claims that inflection, derivation and compounding are essentially different processes. Booij (2005) shows that there is no clear-cut boundary between compounding and derivation, for instance in complex words where the first morpheme can be interpreted as either a word or an affix. He also argues that the claim about syntax not having access to the internal structure of words cannot be upheld (see Booij, 2005, section 2 for a large number of examples).

Regardless of one's view of the relationship between syntax, lexicon and morphology, it seems clear that postulating a schematic template for a morphological construction only makes sense if the template can be shown to be used productively. For some linguists, productivity is an all-or-nothing characteristic of morphological processes. Bauer (2001) distinguishes between a qualitative notion of productivity (availability), which is either present or absent, and a quantitative aspect (profitability), describing the extent to which this process can be used to form new words. Bauer uses this distinction to indicate that a process with extremely limited profitability can still be categorically productive within a specific set of words, i.e. for all words that fit the restrictions on the process. More problematic is his assertion, following Aronoff (1976), that "*the actual production of new words is not necessary to productivity, it is the potential which makes things productive*" (Bauer 2001:21). As Bauer himself acknowledges, taking the *potential* of new words as the defining criterion for productivity makes productivity unidentifiable, because you can only determine if something is a potential word by creating or observing it. At that point, it has become an actual word, and is no longer merely potential. Barðdal (2008, chapter 1) discusses no fewer than nineteen different definitions of productivity -both morphological and syntactic- and reduces these to three main concepts: generality, regularity and extensibility. She argues that generality entails extensibility: if a morphological process is general it must be extensible. The reverse is not true, and regularity often co-occurs with generality and extensibility, but this is not always the case. Processes that are not very general can still be productive, with productivity being determined by type frequency, semantic coherence and the relation between the two. Clearly, this approach views productivity as a gradient phenomenon (see Chapter 3 for a more extensive discussion of productivity).

Various researchers have attempted to quantify productivity by looking at type-token ratios, neologisms and hapaxes in corpora (cf. Baayen & Lieber, 1991; Plag, 2003). These measures are rooted in the observation that new instantiations of a morphological process must be formed productively. The instantiations that are new are found in synchronic corpora as infrequent types or even as a hapax legomena. These measures are not completely objective: how many types and tokens are found depends on the size of the corpus someone decides to use, but also on what is included as an instantiation of a morphological process. The word *openbaar* 'public' serves to illustrate this point: Dutch has a derivational affix *–baar*, which usually takes verbal stems and is semantically more or less equivalent to English *–able*. *Openbaar* looks like an instantiation of this pattern: *open* is a verbal stem (as well as an adjective), and *openbaar* functions as an adjective. On the other hand, it is semantically opaque and it differs in its stress pattern from all other adjectival *–baar* forms (with stress on the last syllable, *baar*, for *openbaar* while other words have their main stress on a syllable that is part of the verbal stem). Whether this word is an instantiation of the *–baar* pattern and should be included it in a corpus search for the *–baar* affix is therefore debatable.

With these quantifying approaches, it is possible to rank morphological processes in terms of productivity within a given corpus. Assuming that a corpus is representative of a language as a given speech community uses it, this ranking reflects the productivity of these processes for these speakers. It is important to note that in these quantifying approaches, productivity is treated like a characteristic of a morphological process that is stable: the measures do not tap in to any differences between speakers (i.e. something may be productive for some speakers, but not all), let alone differences within a single speaker, between discourse contexts (i.e. depending on the topic, genre, addressee etc. a speaker may or may not be able to productively use a morphological pattern). In Chapter 3, I define productivity as dependent on speakers: a morphological process can only be productive by virtue of speakers being able to form new words with it. Productivity as described in the current section is discussed in terms of general patterns, often making use of abstract rules, in principle unrelated to individual speakers' knowledge.

1.1.2 Morphology in psycholinguistics: Processing

Psycholinguistic research aims to discover how speakers' knowledge is applied when they process language. With regard to morphological processing, the central question is whether language users process a complex word as a whole, or as a concatenation of morphemes, i.e. do they decompose a complex word in reception and compose it from its morphological parts in production? Over the past thirty-plus years, a whole range of possibilities has been proposed, ranging from the "affix-stripping" model (Taft & Forster, 1975), which assumes obligatory morphological decomposition, to so-called supralexical models (e.g. Giraudo & Grainger, 2001) in which morphemes are only accessed *after* the whole word has been recognized.

An intermediate position in this debate has been taken up by *dual route accounts*, where some complex words are (de)composed while others are not. This model is mainly used to account for the co-occurrence of regular and irregular forms, e.g. the English past tense with regular pairs like *walk-walked* and irregular ones like *sing-sang*. In such accounts, two essentially different mechanisms are suggested to work in parallel: a general rule through which all regular forms are generated, and a list of irregular forms. Thus, for the English past tense, the rule takes care of the forms in *–ed* and a list provides all other forms. Pinker and Prince (1994) argue in favour of such a dichotomous distinction between regular and irregular forms.

Conceiving of morphological processes as general, abstract rules is not limited to theoretical linguistics. The dual route proposal suggests that there are differences in processing regular and irregular forms. This has been empirically investigated by a large number of researchers, mainly using lexical decision experiments (for an overview of morphological effects in lexical decision tasks see Feldman, 2000, and contributions in Baayen and Schreuder, 2003). These experiments are based on the observation that reaction times to existing words depend on a number of factors, among which frequency is the most influential. Frequency is a facilitatory factor in word recognition. If regular forms, such as most plural nouns in English, are formed productively, the recognition of regular singular or plural forms should be influenced equally by the summed frequency of both forms. Baayen, Dijkstra and Schreuder (1997) tested this hypothesis and found evidence for the storage of frequent regular forms. Rather than discard the dual route account, they let go of the condition that only irregular forms are stored in the list and argue that this part of the dual route is larger than previously assumed. In more recent work, Baayen has abandoned the dual route account; he now favours a memory-based model of morphology, with a much stronger emphasis on stored exemplars (Baayen, 2008). Alegre and Gordon (1999) also found frequency effects for inflected forms, but only if those forms were above a frequency threshold of about 6 per million. This is only compatible with an account that allows for the storage of at least some frequent regular forms. Not all psycholinguistic research seems to point in the same direction, however: Clahsen and Felser (2006) observe that children (aged 5-12) process morphologically complex forms in fundamentally the same way as adults do and argue that this processing can best be described with a dual route model. They suggest that differences they observed between adults and children are caused by children's "*slower and less accurate lexical access and retrieval than in adults*" (Clahsen & Felser 2006:8).

Dual route models do not incorporate these mechanisms of morphological processing in a larger account of the language system: how does the morphological rule that is suggested relate to other rules in the language system? What does the structure of the lexical inventory look like, aside from a list of irregular forms, and how are sentences formed? These questions may seem irrelevant if one is only interested in the processing of morphologically complex words.[1] Evaluating the suggested processing

[1] For other critiques of the dual route account, see Eddington (2000) who suggests alternative single-route models, Keuleers and Sandra (n.d.) who find similarity effects for regular forms which cannot be accounted for in a dual route model, and Kapatsinski (2005) who argues that

mechanism, however, is problematic if it is unclear how it relates to the storage of specific items in the lexicon and to other productive syntactic processes. Even if some results of lexical decision tasks fit in with a dual route model, in normal language processing speakers always combine lexical items -whether morphologically simple or complex- to form meaningful utterances. Are the two proposed mechanisms for morphological processing (rule and list) restricted to such lexical items, or should they be viewed as much more general mechanisms, which also operate at the level of sentential structures?

In a number of experiments, Kuperman and colleagues investigate production and reading of complex words in Dutch and Finnish, using acoustic duration in production, reaction times in lexical decision and gaze measures in eye-tracking experiments. By combining these online measures with detailed data about their test items (e.g. whole word frequency, stem frequency, affix productivity etc.) they are able to show that processing complex words involves all of these lexical measures. For example, in reading Dutch compounds like *werkgever* 'work-giv-er, employer', they observed *"simultaneous effects of compound frequency, left constituent frequency and family size early (i.e., before the whole compound has been scanned), and also effects or right constituent frequency and family size that emerged after the compound frequency effect"* [2] (Kuperman 2008:45). The complex interaction of these effects and their temporal development cannot be accounted for in a dual route model or in a supra- or sublexical model. For this reason, Kuperman puts forward the Probabilistic Model of Information Sources (PROMISE). This model assumes that a language user will use all information as soon as it becomes available. In this, Kuperman follows Libben (2006) who argues that language users *"maximize their opportunities for comprehension by the simultaneous use of all processing cues available to them, and all processing mechanisms that they have at their disposal, including retrieval*

there is no single default stem extension for Russian verbs but two equally productive classes, which, again, a dual-mechanism model cannot incorporate.

[2] Family size is discussed in Section 1.4. This is a measure of the number of different words with the same constituent, e.g. *werkuren* 'working hours' is a family member of *werkgever*.

from memory and compositional computation" (as summarized in Kuperman 2008:71). Note that this is not a dual route model: all lexical items are processed in the same way, namely by employing any cues that are available. This suggestion is compatible with the model of multiple representations that I introduce in Section 5 of this chapter.

1.1.3 Less general morphological patterns

In the studies described above, the morphological processes investigated occur very frequently, and a majority of them concern inflection. This type of research succeeds in capturing broad generalizations, but also has limitations. Studies written from a theoretical linguistic point of view tend to turn a blind eye to individual variation and small-scale systematicities, e.g. they look at –*ed* as the past tense affix for verbs in English and view all other verbs as irregular, whereas there are subregularities for quite a few strong verbs too. There are some studies, however, that do look at such subregularities. Albright (2002) developed a computational model to describe verb classes in Italian. He posited rules for which the reliability was tested in the model on different phonological environments. The resulting stochastic rules combined both general patterns and generalizations for subregularities, which were more specific and more reliable than general rules. An example of such a subregularity is the fact that a large majority of verbs ending in –*end* belong to the –*ere* verb class (the environment reliability of it belonging to the –*ere* class was 0.84, with 1.00 indicating absolute reliability based on the model's exposure to phonologically similar words). Albright then constructed novel verbs matching with these specific phonological environments, which he called *islands of reliability*. He asked native speakers to rate the acceptability of these novel verbs in inflected form. These ratings were found to correlate strongly with the reliability of the specific *island* the form belonged to: speakers rated a form of the novel verb *aggi**end**-* much higher with an inflectional ending congruent with the –*ere* class (mean rating 4.12 on a 7-point scale) than with –*ére* (mean rating 1.90, environment reliability 0.05). Albright concludes that this finding cannot be accounted for in a dual route model. Albright and Hayes (2003) followed a similar procedure, this time looking at the English past tense. Again the results pointed at the existence of islands of reliability (e.g. "suffix [-t] to stems ending in voiceless fricatives") as mental representations, from which the authors conclude that *"inductive learning of detailed generalizations plays a major role in language"* (Albright & Hayes 2003:154).

Dąbrowska (2008) also studied a pattern that displayed subregularities. She asked native speakers of Polish to supply dative endings (elicited in sentence contexts) to nonce nouns. These nouns were phonologically constructed to be of different gender -masculine, feminine or neuter- and belonged to a densely or sparsely populated phonological neighbourhood, i.e. containing either a lot of nouns with the same ending or few. She found effects of gender, neighbourhood density and of the educational level of the participants. Scores were higher on high-density nonce words than on low-density ones, an effect that a categorical rule cannot account for. Participants also did better on feminine and masculine nonce nouns than on neuter forms. Dąbrowska attributes this finding to the fact that corpus counts show these latter forms to be very rare. She tested separately if participants were able to deduce the nonce nouns' gender from their phonological form, which they were. This means that errors in applying the correct dative ending are not caused by participants' inability to assign gender to novel words, indicating that they have difficulties with the dative endings specifically. The large individual differences in performance correlated strongly with participants' educational level. Dąbrowska discards test-wiseness of the highly educated participants as a cause for this difference, because all participants did well on the high-density feminine forms. She suggests that *"[t]he actual determinant of an individual's productivity with dative inflections is (…) the number of nouns experienced in the dative case (which) may depend more on the type of texts an individual has been exposed to"*. According to Dąbrowska, exemplars of inanimate, neuter nouns in the dative case are mainly restricted to formal written texts, because they tend to be high-register or even archaic (Dąbrowska 2008:947). The observed differences can therefore all be retraced to frequency in the input: subgeneralizations reflecting regularities based on gender and phonological neighbourhood.

1.1.4 Frequency and morphology

The fact that frequency plays an important role when it comes to the storage of lexical units is a recurrent observation for seemingly very different phenomena. I will discuss four kinds of frequency effects in morphology here: the role of type and token frequency in the productivity of morphologically complex words and in semantic drift, the effects of relative frequency (of stems and complex forms) and the morphological family size effect. Whereas the independence of frequent types and semantic drift are related to the frequency of whole instantiations, relative frequency and the family size effect have to do with the frequency of the stem of a complex word.

In an influential article on regular morphology and the lexicon, Bybee (1995) suggests that *"high token frequency forms will (...) not contribute to the productivity of a pattern"* (Bybee 1995:434). She explains this as follows: because frequent forms are often rote-learned, they will be stored relatively independently. For a frequent complex form, this means that they are "*less likely to participate in schemas*" (ibidem). A second and related observation is that of *semantic drift*. Frequent instantiations of a construction over time often undergo semantic specialization or change until they no longer are semantically transparent. An oft-cited example is *dirt-y*, where the morphological composition is still visible, but the meaning no longer links closely with the stem *dirt*. Semantic drift is a consequence of a form being stored independently from the pattern it instantiates: a word like *dirty* can only obtain a meaning that cannot be derived from the general pattern if it is stored as a lexical item in the first place. Although semantic drift is a reliable indicator that a form is stored, it is not a prerequisite: we may store forms that are completely transparent instantiations of a more general pattern.

The frequency of the parts of a complex word is central to the notion of *relative frequency*, which has been put forward in work by Hay and Baayen, among others (Hay, 2001; Hay & Baayen, 2002, 2005). Their corpus study showed that a complex low-frequency form is more likely to be semantically non-transparent if it is composed of even-lower-frequency parts. On the other hand, a more frequent complex form may still be highly decomposable if the stem it contains is more frequent than the complex form. In an experimental setting, Hay asked speakers to rate morphologically complex words in terms of their complexity. The participants saw pairs of words with the same affix. For each pair, one word was more frequent than its stem and the other word had a stem with higher frequency than the complex word (e.g. *respiration* versus *adoration*).[3] For each pair, all participants indicated which word was more complex, with complex words defined as "*a word which can be broken down into smaller, meaningful, units*" (Hay 2001:124). Mean ratings reflected not only the frequency of the complex forms they were shown, but also the frequency of the stems. If complex forms were frequent *relative to their stem*, they were rated as less complex. Both the corpus findings and the experimental results suggest that the storage of a complex form depends on its own frequency and that of its parts.

Finally, psycholinguistic experiments, and more specifically lexical decision tasks, have shown that the type frequency of the stem is also relevant for the recognition of words, yielding what is known as the *morphological family size effect* (cf. De Jong, Schreuder & Baayen, 2000; Bertram, Schreuder & Baayen, 2000; Feldman, Soltano, Pastizzo & Francis,

[3] Frequencies in the CELEX corpus are 39 and 49 for respiration and adoration respectively. Stem frequencies are 4 and 218.

2004). The morphological family size of a word is the number of complex words (derivations, compounds) in which that word occurs. In Dutch, the word with the largest family is *werk* 'work', whose family members include *bewerken* 'cultivate', *werkuren* 'working hours' and *werkster* 'cleaning lady'. The family size effect is facilitatory and a type effect. This means that a word with a large family size is recognized faster than one with a small family, all other things being equal. Chapter 2 rapports on a lexical decision task investigating the family size effect in children's reaction times.

The family size effect in lexical decision tasks provides evidence that the morphological structure of complex words is somehow stored, or at least that stems are somehow units within the representations for the complex words. Duñabeitia, Perea and Carreiras (2008) focused on the role of affixes. They too developed a series of lexical decision tasks, but this time with masked priming: prior to each stimulus, participants saw a prime. This prime was flashed on the screen so briefly that they were not aware of having seen it. Still, the nature of the prime influences reaction times. Duñabeitia and colleagues found that reaction times to a morphologically complex word were faster after a prime with the same affix in a nonsense symbol string, i.e. flashing ####er aids in the recognition of WALKER as an existing word. Orthographic overlap between prime and stimulus always facilitates recognition, but Duñabeitia observed a clear dissociation between 'simple' orthographic priming and morphological priming. [4]

1.1.5 Applied linguistics: Assessing morphological knowledge

In applied linguistics, morphology is often approached from the perspective of proficiency, usually that of children or second language learners. Especially in work that is pedagogically motivated, research centres on testing people's ability to produce and understand morphological patterns correctly. Essentially, this type of research is aimed at investigating a language user's ability to use morphological patterns, which really is the

[4] Rastle, Davis and New (2004) observed priming effects in lexical decision tasks for stems (i.e. CLEANER primes *clean*), but also for pseudo stems (e.g. BROTHEL primes *broth*). According to the authors, this suggests a level of representation *"at which morphological decomposition is defined on a purely orthographic basis, where words are segmented simply because they have a morphological structure"* (Rastle et al. 2004:1095). This analysis may be restricted to reading, but clearly there is still a lot to find out about priming effects in general and their implications for models of representation of morphological knowledge in particular.

reflection of a morphological pattern's productivity. Examples of such work include Carlisle (1988, 2000), Freyd and Baron (1982), Nagy, Diakidoy and Anderson (1993) and Nunes and Bryant (2006). In these studies, knowledge of morphology is often explicitly linked to other aspects of language, such as reading and writing ability (e.g. Nunes & Bryant, 2006, who looked at the relation between first graders' morpheme awareness and their spelling ability). Nagy et al. (1993) and Carlisle (1988), using a number of experimental tasks, observed a clear developmental pattern in children's proficiency. Nagy and his colleagues conclude that children's knowledge of suffixes still grows after fourth grade, even after the effects of other, possibly related factors such as test-taking skills have been taken into account. They suggest that this late development is caused by the relative abstractness of the information that derivational suffixes convey and by the fact that these forms mainly occur in written language and formal discourse (Nagy et al. 1993:156-7). The children in Carlisle's 1988 study (English native speakers in the fourth, sixth and eighth grade) were better at analyzing the morphemic structure of derived forms than at producing them and they were more successful with oral performance than with writing. In a later study (Carlisle, 2000), she found that for both third and fifth graders performance on defining complex words was significantly correlated with awareness of morphological structure, which was tested separately. The different tasks Carlisle constructed to test children's proficiency are all offline tasks: in contrast to the lexical decision tasks described in the previous section, children had time to reflect on their responses. Most of the tasks Carlisle discusses are typical of the classroom: they involve explicit definitions, morphological decomposition etc.

The results of these tasks are informative with regard to what children can do, but one must be careful to draw direct conclusions from them about language use. Anglin (1993) touches upon this issue when he suggests that we must distinguish between words that have been learned as a whole and those that are interpreted with the help of knowledge of morphological structure. He too observed a gradual increase in children's ability for what he calls *morphological problem solving*: figuring out the meaning of a complex word through its structure. For him, this is a reason to *"emphasize the need to distinguish between words for which there are distinct representations in the mental lexicon and words that are not so represented but are rather potentially knowable through morphological analysis and composition"* (Anglin 1993:130, emphasis in the original). Anglin does not discuss how we can distinguish between those two. In fact, the identification of a lexical item as either represented separately or known through morphological analysis is a very real problem. In both cases a language user may be able to use and define this item correctly. This problem, the impossibility to identify instantiations of a complex pattern as either productively formed or used as a chunk, is an issue I will return to at various points throughout this book, when discussing children's performance on linguistic tasks (Chapter 2) and in the chapter on productivity (Chapter 3).

Some of the tasks that are used in this kind of research contain pseudo words. Berko (1958) first employed pseudo words to test children's ability to use a rule. When a child is asked to, say, form the plural of an existing word like *tree*, a correct response *(trees*) may be the result of rule application or retrieval from memory of this particular form. With pseudo words, the second option ceases to exist: participants in the experiment cannot have the plural in their mental inventory. Freyd and Baron (1982) utilized this advantage in their experiment in which children had to learn pseudo words and derivations from that word (e.g. skaf – steal, skaffist – thief). They found that children did draw on these morphological relations when they learned the words, although there were large individual differences. There are some important limitations to these kinds of experiments with pseudo words. The children in Berko's experiment performed much better on existing words than on the pseudo words. This seems to undermine Berko's conclusion that these children *"operate with clearly delimited morphological rules*" (Berko 1958:269). If they do, how can this difference in performance be explained? One possibility is that children can make use of these rules, but do not do this with (most) words they already know. That means that the performance on items or tasks with pseudo words, although they reflect an ability of children, may not be an indication of their everyday language processing.

Research into the acquisition of morphology it not only limited to experimental tasks. Especially for younger children, investigators often rely on spontaneous speech data. As with elicited data, a spontaneous speech corpus has its drawbacks. An important issue is the density a corpus needs to have: what percentage of children output must be recorded in order to be reasonably certain about, for instance, developmental sequences? Tomasello and Stahl (2004) show that this depends on the frequency of the phenomenon that is studied and on the density of the recordings. According to them, the most common sampling density, one hour a week, is insufficient to state much about relatively rare phenomena. If one is interested in, say diminutives in Dutch, and a child uses *koekje* 'cookie' in a recording, this need not mean that this child is able to use the diminutive *–(t)je* with other nouns as well. It may simply have the lexical item *koekje* stored as an unanalyzed whole. Sometimes researchers come across clear examples of children employing morphological analysis (1) or productively forming morphologically complex words like compounds (2) and derivations (3), but these instances are too rare to draw any conclusions from them other than that children sometimes perform these analyses (examples are taken from Elbers and Van Loon-Vervoorn 2000).

(1) Klokhuis, dat is een huisje voor de pitten (Tim, 4;2)
Clock-house, that is a house_DIM for the pips
'An apple core, that is a little house for the pips'

(2) Oorlogmannen (Thomas, 4;5)
War_men
'Soldiers'

(3) Afpeuken (Thomas, 5;8)
Off_butt_ing
'Tipping the ashes off a burning cigarette'

This is not to say that spontaneous speech corpora cannot be used to study the acquisition of morphology. Blom (2003) is a good example of an in-depth study which looks at longitudinal data from a developmental perspective. She used data from Dutch monolingual children and tried to describe their acquisition of finiteness. In her data analysis, Blom distinguishes four developmental stages, but Nap-Kolhoff (2010, chapter 7) argues that this distinction is an artefact of the research design and that the data are better accounted for as evidence for a gradually increasing language repertoire, with abstractions coming about slowly on the basis of the input children receive. For a more extensive discussion of spontaneous data in finiteness and inflection acquisition research, see Nap-Kolhoff (2010).

The performance on tasks by children mentioned in the preceding paragraphs shows a few clear patterns: there are large differences between children who do the same task, as well as between tasks. Overall, children's ability to consciously use the morphological structure of complex words to perform well on them is shown to grow with age, although it is not altogether clear whether this can be interpreted as evidence for the storage of morphological structure, since older children outperform younger ones on a variety of possibly related measures (vocabulary size, test-taking skills, cognitive development etc.). Children with a larger vocabulary, for instance, are more likely to have stored an individual complex word, and may therefore perform better on a task, without this being an indication that they master productive use of the morphological pattern.

Differences in performance between tasks may well be caused by the nature of these tasks. Generally, a distinction is made between online and offline tasks. The former measures language processing, whereas the latter concerns those tasks in which participants can, so to say, 'step away' from the process and think about the response they give to a test item. Many of the experiments discussed here are prototypical offline tasks: children are asked, for instance, to learn the meaning of new words by heart (Freyd & Baron, 1982) or to define complex words (Carlisle, 2000). Examples of online experiments include the lexical decision tasks discussed in Section 2.4 of this chapter. It may well be the case that children make use of different aspects of their knowledge of complex words and structures, depending on the task requirements. This is recognized in work by, among others, Hulstijn and N. Ellis (Hulstijn, 2005; N. Ellis, 2005), who distinguish between explicit and implicit knowledge. The main difference between these two is whether someone has explicit awareness of -in this case- the morphological structure and regularities "*underlying the information one has knowledge of, and to what extent one can or cannot verbalize these regularities*" (Hulstijn 2005:130). Clearly this distinction is related to online and offline tasks, although it is not exactly the same. In Chapter 2 a number of experiments I developed are reported on: a word definition task, a word formation task and a lexical decision task. Here too, the differences in performance between tasks are significant. The notions of online and offline tasks and implicit and explicit knowledge will prove helpful in accounting for these differences (see Sections 2.1.3 and 2.1.2).

In research on the developing lexicon of children, another aspect that has generated a considerable amount of research is the notion of 'deep word knowledge'. This reflects the idea that knowledge of a word incorporates frequent associations with other words and information about relations with other words (such as oppositions, hyponyms etc.). Various researchers have investigated ways to measure such knowledge (e.g. Read, 1993, Strating-Keurentjes, 2000, Schoonen & Verhallen, 2008) and have emphasized the importance of these aspects of knowledge in language teaching (Nation 1990). Often, the relations that are investigated are limited to a small number of relations (super- and subordinates, opposites, (near-) synonyms). Moreover, these relations are usually investigated for a restricted part of the lexicon, especially concrete nouns such as types of clothing, colours, furniture etc. Assuming that languages learners develop their linguistic repertoire on the basis of the input they receive, relations between lexical items will be established through distributional properties. These can be 'horizontal' or 'vertical'. Horizontal relations are relations between items that co-occur: collocations of all types. For a noun like *vakantie* 'vacation', this might include frequently occurring modifiers (*grote vakantie* 'summer holiday'), verbal frames (*op vakantie gaan* 'go on vacation'), determiners (*de vakantie* 'the vacation') etc. These are syntagmatic relations; this knowledge informs you about likely linguistic context given the item. 'Vertical' relations are paradigmatic: they are relations between items that may occupy the same slot given the linguistic context. These may be opposites (*the store is open/closed*), near-synonyms (*how was your vacation? Great/super!*), super- and subordinates (*I just bought a new pair of shoes/ pumps*) etc. This type of distributional information also supplies a language learner with the clues needed to form word classes (cf. Tomasello, 2003).

It is not surprising that applied linguists emphasize proficiency and knowledge, since this is at the heart of their research questions. By not linking it to a model of the linguistic system, however, it is not always clear how the knowledge investigated in this type of research is connected to other aspects of the networks of representations that together form our language system, which we use to produce and understand meaningful discourse. This field of research seems to view morphological knowledge either as pertaining to the domain of the lexicon or neglect to link it to other aspects of a user's linguistic proficiency.

1.1.6 Summary

In sum, in Section 1.1 I have looked at morphological research of very productive phenomena from a theoretical perspective, at processing of (mainly) inflectional morphology, at modelling more complex systems with subregularizations, and at the development of morphological knowledge in language learners. This is just a subset of research in the field of morphology, but this overview was intended to raise three issues. First, researchers who study morphology are often not explicit about the ways in which the phenomena they study relate to other aspects of linguistic knowledge. If they are, they will either say that morphology is a separate module (e.g. in Distributed Morphology, Halle & Marantz, 1994) or that the same mechanisms apply in all of language learning (e.g. N. Ellis, 2005). Second, linguists who are interested in the description of general and productive morphological phenomena tend not to pay much attention to differences, neither at the level of subregularities nor in variation between speakers. When, on the other hand, proficiency is investigated, researchers are not always very clear about what they see as the system that needs to be acquired. In both cases, this is understandable, and there are attempts to bridge this gap (e.g. Dąbrowska, 2008). In this book, I attempt to do the same: give a theoretical description of the phenomenon of interest and look at what speakers know about the phenomenon, both explicitly and implicitly, through their language use. The phenomenon I am concerned with in this book is *complex lexical items*, a definition of which is given in Section 3. More specifically, the focus will be on the productive use of patterns of which complex lexical items are instantiations and on the storage of individual complex lexical items that instantiate them. First, I present a short overview of research regarding larger combinations of words that may be used as lexical items.

1.2 Larger lexical units

Morphology is usually viewed as the study of structure within word boundaries, but we know that people also store bigger chunks of language. Like morphologically complex words, these chunks contain more than one meaning-carrying element. In morphology, the elements are stems and affixes, the latter of which do not occur outside the complex words. For larger lexical units, this is usually not true: they mostly consist of words that can be found in other contexts as well. Because this is the case, we must assume that these words have their own lexical entry, i.e. they are stored independently from the larger chunk. Yet, there are a number of reasons to assume that many larger combinations are stored as well. These are discussed in Sections 1.2.2 and 1.2.3. Subsection 1.2.1 provides a brief summary of research on larger lexical units, and starts with remarks on the definition and identification of these chunks.

1.2.1 Defining and identifying larger lexical chunks

A first question that needs to be addressed is why one would want to assume that combinations of words are stored. This brings up the problem of identification: how can we be certain that a sequence of words, spoken by a particular speaker in a specific context, is retrieved from memory as one unit? Both of these questions implicitly define larger lexical chunks as *being stored as a unit*. It is in this respect that they differ from other sequences of words, which are produced by combining stored units. Larger units themselves can be combined with other constructions and units to form utterances and conversational turns. The defining criterion is thus that a larger lexical chunk consists of more than one word but is used as a single 'building block' in the production and reception of language. Wray (2008) seems to go a step further in her definition of a *morpheme equivalent unit* (MEU, her label for instances of formulaic language) to label instances of formulaic language and says that a MEU: *"…is processed like a morpheme, that is, without recourse to any form-meaning matching of any sub-parts it may have."* (Wray 2008:12). This seems to impose a restriction on MEUs that goes beyond storage as a unit: psychologically, the unit cannot be internally complex, i.e. there cannot be any analysis of its internal structure. For many sequences of words that have been identified as examples of larger lexical chunks, this poses a big problem. As they combine with other units in sentence structures, idioms, for instance, undergo adaptation to the larger context: a verb will appear in an inflected form, the ordering of the constituents may depend on the information structure determined by the discourse context etc. If the idiom were *processed like a morpheme*, it seems impossible to explain how this happens. For me, storage like a unit does not necessarily entail that a lexical unit is completely unanalyzed or morpheme-like (see Section 1.6).

Although there seems to be a consensus that storage as a unit is a necessary condition for a sequence of words to be identified as a lexical unit, opinions differ with regard to what sequences should be included and accepted. A veritable plethora of names and labels has been applied to larger lexical chunks, ranging from *multiword sequences* and *formulaic sequences* to *unanalyzed chunks* and *idioms* and *collocations* (cf. Wray, 2002 for a discussion of these labels and Fillmore, Kay & O'Connor, 1988 for a clear typology of idioms within a cognitive framework). Some of the labels are more specific than others.

Idioms are probably the clearest and most uncontroversial case of larger lexical chunks. The meaning of an idiom like *kick the bucket* appears completely unrelated to the words it is composed of. In many other cases, conceptual metaphors can be identified that underlie the idiom (e.g. *to spill the beans* has been argued to have THE MIND IS A CONTAINER and IDEAS ARE ENTITIES associated to them, Gibbs & O'Brien, 1990). In either case, it is impossible to reconstruct the meaning of the whole from its parts, let alone compositionally produce the idiom –why do we say *spill the beans* and not *spill the rice*?

A second undisputed category of larger lexical chunks are those fixed expressions that contain words that are restricted to one or a few combinations: they contain words that co-occur almost exclusively with each other (e.g. *holier than thou, mutatis mutandis*). For these two types of larger lexical chunks, combining existing lexical entries for the individual words cannot lead to a correct interpretation, where 'correct' is the interpretation given by native speakers.[5]

Collocations form a more contentious category of larger lexical units. Many word combinations seem completely regular, but at closer inspection turn out to be conventional collocations (e.g. *white wine* for a yellowish drink, *long* but not *large* stories, *strong* but not *powerful* coffee etc.). Some of

[5] This is in fact rather simplistic. If, as is done here, we take the meaning of words and larger lexical chunks to be what people have stored, what is the yardstick to judge an interpretation to be 'correct' and how can we find out what the meaning of a lexical unit is? Corpus studies can provide valuable information, because a (good) corpus provides a sample of instances where a lexical unit is used. Since use reflects what is stored, this gives us an -indirect- picture of meaning.

these combinations are discussed in Section 1.3. There is no agreement on whether collocations should be classified the same as idioms. From the point of view of storage, the answer has to be affirmative: we can only account for the occurrence of collocations if they are psycholinguistically real, that is if these collocations are stored as units in speakers' constructicons. Collocations not only include adjective-noun combinations, but also many other conventional expressions like *at school* (compare [?]*at kindergarten*). Many of these conventions are usually not thought of as collocations. Prepositional verbs, for instance, are not fully transparent: the exact choice of preposition cannot (always) be deduced from its semantics, something which becomes clear when we look at differences between typologically related languages. In Dutch someone dreams *of* or *about* something (*dromen van/over*), but in Spanish you dream *with* (*soñar con*). The same occurs with adjective-preposition collocations like *angry at* (see Chapter 4) and with fixed verb-object pairs (e.g. *make a decision, take a photograph*). All of these examples are semantically transparent, which means that someone who does not have the combination stored as a unit may still be able to understand them, but they are conventions: if you want to produce them, you do need to know exactly which preposition goes with *angry* etc.

Depending on the word orders of a language, these conventional combinations do not always occur consecutively. The first part of Wray's definition explicitly allows for this: *"[a MEU is] a word or word string, whether incomplete or including gaps for inserted variable items …"* (Wray 2008:12, emphasis added). If these discontinuous items are also included as larger lexical units, this is another reason why *processing like a morpheme* is untenable as a defining criterion. The simple fact that the lexical item consists of two parts separated by other units means that there must be some amount of internal structure. Langacker, who uses the term *complex expression* for larger lexical units already recognizes this: *"A complex expression involves an intricate array of semantic and symbolic relationships, including the recognition of the contribution made by component structures to the composite whole (…) even when its component words occur discontinuously to satisfy the dictates of higher-order grammatical constructions that incorporate it."* (Langacker 1987:475).

Wray's definition also includes incomplete word strings, which I take to mean lexical chunks that are partially underspecified (these are *formal* or *lexically open* idioms in the typology proposed by Fillmore, Kay and O'Connor 1988). These patterns, labelled *partially schematic units* in Doğruöz and Backus (2009), include some lexically specific element(s) and an open slot. A nice example is the Dutch collective construction *met zijn NUMBER-en* (literally: with his NUMBER-en) described by Booij (n.d.). This construction is used in sentences like (4), in the context of entering a restaurant and letting the person who welcomes you know that you would like a table for four:

(4) Wij zijn met zijn vieren.
We are with his four-AFFIX
"There's four of us"

At first glance, it may not be clear why this should be classified as a larger unit. Booij points out, however, that the marker on the number, glossed in (4) as AFFIX, looks like a normal plural, but is not. In all other contexts, the plural of *zeven* (seven) is *zevens* as in (5). Within the collective construction context, the form *zevenen* is used (6).

(5) Ik heb drie zevens op mijn rapport
I have three seven-PL on my report_card

(6) Ze komen met zijn zevenen
They come with his seven-AFFIX
"There will be seven of them"

The lexically fixed elements *met zijn –en* and the slot, where any number can be included, therefore have to be stored as a unit: there is no way to explain the *–en* affix as a regular pattern and this construction as an instance of the general plural schema.

In sum, it has been suggested that idioms, expressions with words that only occur inside these, continuous and discontinuous collocations of many different types and partially schematic units are all lexical units. The remainder of this section is devoted to examples of research showing that these larger units are not only theoretical proposals but that there is evidence that these units are real; they are stored and used by individual speakers.

1.2.2 Evidence for the existence of larger lexical chunks

In the previous subsections I have argued that the defining criterion for a lexical chunk is that they are stored as units. There is evidence for the storage of different kinds of larger lexical chunks, some of which will be discussed here. In trying to investigate larger lexical units, researchers have to devise experiments that show differences in processing between potential larger lexical chunks or similar-length sequences of words that are not collocations, idioms etc.

Ehrismann (2009) looked at the processing of sequences of words that were idiomatic phrases and those that were not, but with the same constituent structure (examples 7 and 8).

(7) Het neusje van de zalm
The nose of the salmon
"the cat's whiskers"

(8) De sleutel van de voordeur
The key of the front door

Ehrismann provided participants in his experiment with sentences on a computer screen, which then had to be copied on another screen. Each time a participant switched between the original sentence and the copy they were making, this was logged. Ehrismann hypothesized that such switches would occur at points between different lexical units.[6] By contrasting the switch behaviour between idiomatic sequences such as (7) and non-idiomatic phrases such as (8), this hypothesis was tested. Constituents like the one in (7) required fewer switches to copy than those in (8). This shows that idiomatic structures were processed more as a unit.

Schilperoord and Cozijn (2010) investigated the unit status of the same type of idiomatic expressions, using eye-tracking equipment in a reading task. They observed that participants had more difficulty with anaphoric expressions when these related to part of an idiomatic expression. They needed more time to resolve the anaphora that referred to the first NP in (7), *het neusje* than in (8), *de sleutel*. Schilperoord and Cozijn interpret this finding to indicate that within idiomatic expressions the parts are less available for resolution, meaning that the expressions are processed as a unit. Conklin and Schmitt (2008) also looked at how people read idiomatic expressions, and found that their participants read idiomatic expressions more quickly than non-formulaic sequences that were matched for length. As with the problems in anaphoric resolution, this difference in reading times can be accounted for by assuming that the idiomatic expressions are stored as units: because non-formulaic sequences consists of combined lexical units, they take longer to process than idioms which are stored as one unit. In a lexical decision experiment for children (fifth graders) Qualls, Treaster, Blood and Hammer (2003) found that idioms were recognized faster than matched non-idiomatic sequences. They also observed a frequency effect for the idioms, supporting their conclusion that these idioms are recognized as lexical items.

[6] In Chapter 4 I report on an experiment using the same task to investigate the unit status of fixed adjective-preposition combinations like *trots op* 'proud of'.

Other types of evidence for the existence of larger units (cf. also Wray, 2008 for an overview) come from the field of second language acquisition. Boers et al. (2006) looked at perceived fluency in English as a second language. They found that blind judges rated L2 learners of English as more fluent speakers when they used more collocations and idiomatic expressions. This proves that use of such items is a part of proficiency in a language. Code-switching data, in which bilingual speakers use two languages at the same time, also points at fluent code-switchers' frequent use of larger lexical chunks. In his discussion of Turkish-Dutch code-switching data, Backus (2003) identified a number of different types of units used by code-switchers. These include idioms and various kinds of collocations. Backus suggests that it is necessary to recognize these larger combinations of words as lexical units to further our understanding of code-switching.

Finally, neurolinguistic data constitute yet another source of evidence for the existence of larger lexical units. Tremblay (2009) investigated *lexical bundles*: sequences of words that occur very frequently, and do not necessarily form one structural item (i.e. a constituent) or a clear semantic unit (cf. Biber 1999:990-1024). Examples include *in the middle of* and *I don't know whether*. Tremblay employed a variety of experimental tasks, with measures ranging from chunk recall to speed of production and semantic ratings, all of which point at facilitatory effects for lexical bundles. Utterances with lexical bundles are more likely to be remembered, will be produced faster (they have an earlier speech onset) and are rated semantically 'better' than utterances without such bundles. In an innovative design, these behavioural measures are coupled with neurolinguistic data. Tremblay's ERP results show that "*sequence-internal trigrams and single words modulate recall in addition to whole-string probability of occurrence [which] suggests that four-word sequences are both stored as wholes and as parts*" (Tremblay 2009:67).

In sum, research showing that some sequences of words are stored and processed as units comes from very different sources. The data described above are from monolingual speakers, learners of a second language and speakers who use two languages at the same time. They include offline judgments of fluency and semantic acceptability as well as online reading and copying data and neurolinguistic evidence. All these data provide converging evidence that larger (i.e. multi-word) lexical units are psycho-linguistically real: people store them and use them. It is now time to combine what has been said about morphologically complex words and larger lexical units and recognize that they are really very similar: they are complex lexical items.

1.3 Complex lexical items: A definition

Complex lexical items are strings of language in which more than one meaning-carrying element can be recognized yet which are very likely candidates to be stored as units in people's linguistic repertoires. Complexity here should not be understood as something which is difficult, but only as an containing internal structure. A templatic representation is given in (9)

(9) E_1 (...) E_2 (...) E_n

The capital letter E indicates a meaning-carrying element: morphemes or words. As stated above, a complex lexical item must consist of at least two of these. There might be more than two, as is indicated by the last part of the representation (E_n). The different elements form one word or two (or more) words, which may or may not always immediately follow each other. The possibility of discontinuous complex lexical items is indicated by (…) between the elements. Particle verbs such as *to call (…) up* are an example of a discontinuous complex lexical item. For these items, the exact position of the elements is determined by the syntactic context in which they occur. This does not take away from the fact that these fixed combinations still form one complex lexical item.

Both the *complexity* and *unity* elements of this definition will be discussed below. In what follows here, it will become clear that a CLI's *complexity* is something that is relatively easy to determine with linguistic analysis. At this point, I want to stress that complexity need not be visible to a language user, although it may be. In discussing the *unity* of CLIs there will be a focus on the larger units that language users store. A model incorporating both complex and unitary representations for CLIs will be introduced in Section 1.5. At that point, I will also describe generalized knowledge representations in the form of (partially) abstract constructions, which are needed to explain the production and acceptability of novel forms. Throughout this text, CLI stands for Complex Lexical Item.

1.3.1 The complexity of CLIs

One of the defining elements of a CLI is that it contains more than one meaning-carrying element. This means that morphologically simple words like *school* are not CLIs. The definition of a CLI speaks of 'meaning-carrying elements' because it includes both morphemes and words. A CLI contains at least two morphemes or words; this is what makes a CLI *complex*.

Basic examples of one-word CLIs are morphological derivations such as *eetbaar* (eet-baar, eat-able, 'edible') and compounds such as *fietssleutel* 'bike-key, key to open the lock on a bike'. Each of these consists of two parts. *Eetbaar* is a combination of the verbal stem *eet* 'eat' and the adjectival suffix –*baar* which is more or less equivalent to English –*able* (see Chapter 2 for a detailed analysis of this Dutch suffix). The bound morpheme –*baar* is found with a variety of verbal stems. Its semantics when combined with a verbal stem can be paraphrased as 'property describing the possibility that something can be V-ed', which in this case leads to 'can be eaten' or 'edible'. The compound *fietssleutel* is a concatenation of two simple nouns, *fiets* 'bicycle' and *sleutel* 'key'. Used together they designate a kind of key, i.e. used for locks on bikes. This type of N-N compound, where the left part specifies a type of the right part (an XY is a Y with some relation to X) is very frequent in Dutch (Booij, 2002).

CLIs can be longer than one word. Examples of multi-word CLIs include fixed expressions like *rode wijn* 'red wine' and combinations such as *trots op* 'proud of'. It is easy to identify two meaning-carrying elements in *rode wijn,* as they are separate words. Note that in this combination of adjective + noun, the adjective carries regular inflection (Dutch adjectives end in schwa, orthographically represented with –*e*, when they precede a non-neuter noun). This inflection is not fixed in this CLI, as becomes clear when the diminutive –*tje* is added to *wijn*: *een rood wijntje* 'a red wine-DIM'. The diminutive is used both as a quantifier and a term of endearment here. The resulting phrase means something like 'a nice glass of red wine'. For our discussion here, what is relevant is the form of the adjective *rood* 'red'. Since the diminutive suffix makes the resulting noun *wijntje* neuter, the inflection on the adjective changes correspondingly. If this expression behaves grammatically like any other adjective-noun combination, why then would we want to assume that it is a CLI? Because it has a specific meaning which cannot be deduced from its constituting parts: *rode wijn* is not a wine which is red in color, it is what Spaniards call *vino tinto*, coloured wine, in contrast to *witte wijn* 'white wine' and rosé. With *trots op* 'proud of' it is again not difficult to identify two meaning-carrying elements. The preposition *op* is used to lexically express what someone is proud of (see Chapter 4 for a description of the semantics and an experiment that focuses on this and other fixed adjective-preposition combinations).

The inventory of CLIs that speakers have stored in their constructicons is likely to differ from person to person: as a linguist, notions like *derivational morphology, universal grammar* and *formulaic language* have become fixed expressions for me, but I am aware that this may not be the case for normal speakers of English. Some CLIs are restricted to particular genres or jargon, as the examples in the previous sentence. Others mark someone as belonging to a particular (cultural, social or otherwise defined) community. For native speakers of Dutch, the three-word fixed expression *de grote* *vakantie* 'the big holiday' can only refer to the period in which schools are closed in summer. Used mainly with a definite article (a Google search on January 27, 2009 showed 2.240 hits for *een grote vakantie* versus 56.400 for *de grote vakantie*, near enough a 1:25 difference in distribution) both form and meaning are completely conventionalized.[7] Individual differences are a common observation in studies that focus on vocabulary; it is generally accepted that different people know different words to a different extent. The same thing must be assumed for CLIs.

The fact that two or more morphemes or words *can* be recognized does not mean that all speakers of a language *will* necessarily do this at each point in time. Most, if not all Dutch native speakers will happily concede that the word *burgemeester* 'mayor' consists of the two words *burger* 'citizen' and *meester* 'master' when asked if they can identify parts in the word. It is likely, however, that for many, this question will prompt them to realize this for the first time because they are explicitly asked to look at the word more closely. At this point, it is useful to emphasize that the explicit recognition of the complexity of a CLI, i.e. its internal structure, by every speaker of a language is not a defining criterion. The difference between speakers' knowledge (and representation) of CLIs and the linguists' perspective will be taken up in the discussion chapter at the end of this book.

1.3.2 The unity of CLIs

After focusing on the complexity of CLIs in the previous section, I now turn to its *unity*: the assumption that these strings of language are stored and processed as wholes in speakers' mental inventory of linguistic units, even though they contain two or more meaning-carrying elements. I discuss two important and at some points related reasons for making that assumption: non-compositionality of meaning and frequency of (co-)occurrence.

Firstly, all CLIs mentioned so far are to some degree non-compositional in meaning. Take for example the meaning of *eetbaar*, a deverbal adjective consisting of *eet* 'eat' and *–baar* 'can be V-ed'. Put together, the meaning

[7] The distribution of *grote vakantie* in the Corpus of Spoken Dutch (CGN, 10 million words in size) confirms this: of all 69 sequences, 54 are immediately preceded by *de* 'the' and 10 have no determiner in the NP. One occurs with *die* 'that', there are two occurrences with an added adjective *de vorige grote vakantie* 'the previous summer holiday', and only two have the indeterminate article *een* 'a'.

would be 'can be eaten', whereas a corpus search immediately shows that it is mainly used to convey that something is not poisonous. This restriction cannot be deduced from the verb, the affix or the construction in general. For that reason, speakers who use this complex word with the specific meaning must have it stored as a separate entry in their mental lexicon. Note that there still is a clear relation between the meanings of the parts and the complex form. The meaning of the complex item is just not completely compositional: it cannot fully be inferred from its morphological parts.

In combining different lexical items, their meanings must often be *accommodated* (cf. Langacker 1987:76). When *running* is applied to humans, it is a different kind of movement than the one horses or dogs make, but that doesn't prevent us from understanding both *my friends run ten miles in an hour and a half* and *that dog came running after me*. In our interpretation, we accommodate the type of movement to the moving agent. Whereas accommodation is common in complex expressions (which, to emphasize once more, are not necessarily difficult), this cannot account for the semantic specialization found with *eetbaar*: the two elements *eet* and *–baar* nor their combination forces a reading of food being non-poisonous. The same is true for *fietssleutel*. Although there are other N-*sleutel* compounds in which the left member specifies what can be opened or accessed by using this key (e.g. *huissleutel* 'house key', or *voordeursleutel* 'front door key') not all N-*sleutel* compounds have this meaning: *kruissleutel* 'cross-key' means four-way wrench.

If a CLI is non-compositional in meaning, it must be stored as a unit. Language users may also store larger strings of language for another reason: frequency of use. If a certain complex word or word combination occurs often enough, a memory trace may persist and over time this sequence will be stored as one unit with its own meaning, rather than always being composed from its constituent parts. There is psycholinguistic evidence that some frequent plural nouns in Dutch, which are both formally (in their morphological structure) and semantically quite transparent, are stored. Baayen, (1997) found that frequent plurals were recognized faster in a lexical decision task than less frequent plurals (controlling, of course, for their frequency of occurrence in the singular form). Such frequency effects can only be accounted for by a model which allows for the storage of regular plural forms. Views have changed over the past decade or two on this point: emphasis used to be placed much more on the notion of *economy of storage*. It requires less memory 'space' to store, say, a plural formation rule and a list of regular singular nouns, than to store a lot of plural forms separately. This view is clearly what Aronoff and Anshen (1998) express when they argue that "*[s]ince the actual sense of the [regularly complex] word does not diverge from its predicted sense, based on its parts and its morphological structure, there is no need for this word to be listed in the speaker/hearer's lexicon, for the morphological component of their grammar is able to process it entirely*" (Aronoff & Anshen

1998:239). Note that some form of rule, schema or generalization must always be assumed in order to account for the formation of novel plurals, which native speakers can do without any difficulty.

The scope of this chapter does not allow for an extensive discussion of the arguments for different models of the mental lexicon. In my opinion, the consensus has shifted from emphasizing economy of storage towards models that allow for redundant storage of information, with the advantage for speakers/hearers that these provide an *economy of processing*. These models may require more memory, but encoding or decoding a complex lexical item is no longer necessary if the whole form is stored as a unit. In computational modelling of language, a similar shift can be observed, with recent models such as Van den Bosch (2005) assuming massive storage. These stochastic models are memory-based and can be used to determine the likelihood of co-occurrence patterns (see also Chapter 4).

For the fixed collocations *rode wijn, trots op*, and *de grote vakantie* it may be possible to deduce their meaning, but the exact choice of words is conventional and partially opaque. Erman and Warren (2000) call this compositionality *in a strict sense*. Evidence for the conventionality of these combinations is the reported difficulty that second and foreign language learners have with them: they have to be rote-learned in order to be produced correctly. With these types of fixed expressions there is quite often a difference between the knowledge needed to understand them and what is necessary for their production. Fillmore, Kay and O'Connor (1988) capture this aspect in their typology of idioms through their distinction between *encoding* and *decoding* items.

Section 1.3 has introduced a definition of Complex Lexical Items: they are strings of language in which more than one meaning-carrying element can be recognized and which are likely candidates to be stored as units in people's linguistic repertoire. What it means exactly that they are both units and complex was discussed for morphologically complex words and larger lexical chunks. In my view, these two types of CLI are essentially the same, which is why they are treated together. Before I sketch a model of acquisition and (multiple) representation of CLIs, the next section is devoted to the theoretical framework within which the present research is carried out, and which has provided the motivation for that model.

1.4 Cognitive linguistics & Construction Grammar

1.4.1 Some basic tenets of Construction Grammar and Cognitive Grammar

The preceding section focused on completely specified complex lexical items. At some points, I touched upon the existence of more general schemas or (partially) underspecified generalizations. Examples of these include N-*sleutel* for a subgroup of Dutch compounds, or the even more abstract generalization that the large majority of Dutch compounds is right-headed: an XY is an X modifying a Y (a *huissleutel* 'house key' is a key of the type 'house'). Complex words with a derivational affix also pattern in partially schematic generalizations, as with V-BAAR, meaning 'can be V-ed'. In theories of language that postulate a strict dichotomy between lexicon and syntax, such patterns cannot be accommodated easily. They are not a problem, however, for Construction Grammar and Cognitive Grammar, the theories adopted here (Langacker, 1987; Goldberg, 1996, 2005; Croft, 2001, Boas, 2003; cf. Östman & Fried, 2005 for a comparison of different directions within the Construction Grammar framework; Schönefeld, 2006 discusses the notion of 'construction' in other linguistic frameworks). This section introduces some basic tenets of these theories in relation to complex lexical items. Construction Grammar (hereafter CxG) is influenced strongly by Cognitive Grammar. Whereas individual cognitive linguists and construction grammarians may emphasize different aspects of linguistic theory and disagree on some points, in my view the similarities and shared assumptions are much larger (see also Goldberg, 2006, chapter 10 and Langacker, 2009).

In the CxG framework, constructions are the starting point for linguistic research. They are defined as form-meaning pairings and constitute "*the primary objects of description*" (Goldberg 2006:220). A related tenet, which is essential to my research topic, is the conceptualization of lexicon and grammar as part of a continuum of constructions, an assumption already made by Langacker: "*(…) cognitive grammar, which posits for lexicon, morphology and syntax an array of symbolic units that range continuously along such parameters as specificity, entrenchment, and symbolic complexity*" (Langacker 1991:44). The 'lexical' part of the continuum contains completely specified constructions; fully underspecified patterns are found on the 'grammar' side of the continuum. Viewing language as an interconnected constellation of constructions, with the specificity of constructions falling on a continuum, entails that partially schematic patterns are a normal phenomenon, located somewhere in the middle of that continuum. This is explicated in Doğruöz and Backus (2009) who visualize the continuum in Figure 1.1, reproduced below. The figure describes the form aspects of the constituting elements in the sentence *it rains a lot in Holland*, but the same could be done for their meaning.

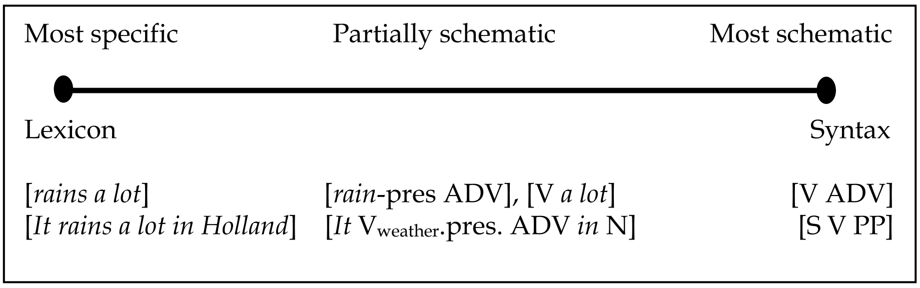


Figure 1.1: Specificity continuum (reproduced from Doğruöz and Backus, 2009)

Using the same continuum, it is equally straightforward to describe complex lexical items and their schematic patterns, as in Figure 1.2. From this figure it is clear that the lexically specified items and the patterns they instantiate only differ in their degree of specificity.

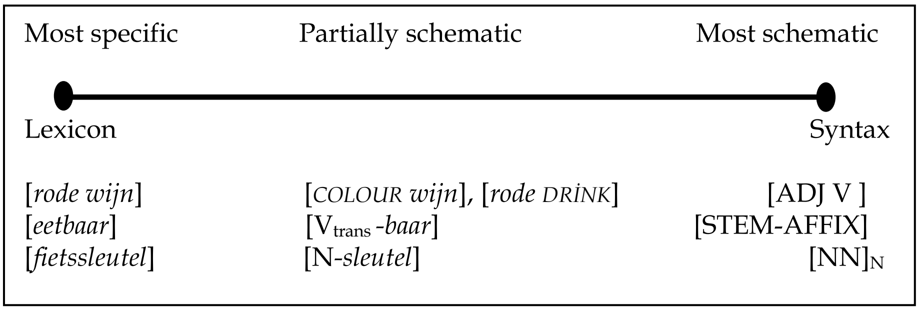


Figure 1.2: Specificity continuum for CLIs

In Figures 1.1 and 1.2 the more schematic patterns include notions like 'ADJ' and 'N' (short for adjective and noun).[8] To see if the distributional properties of the input that children receive is sufficient for categorizing words in grammatical classes, Mintz (2003) analyzed data from the CHILDES database, which consists of children's and (mainly) their caretakers' speech. He identified frequent lexically specific frames of two words, with one word intervening, like *I* ___ *it* or *put* ___ *in*. With just 45 such frames, the model he devised was able to correctly identify the word class of 91% of all types. This shows that an input-based 'detection' of such categories is possible. Whether children will actually develop general categories that are construction-independent is a question worthy of more research. One possibility, investigated by Lieven, Behrens, Speares and Tomasello (2003) is that children start out with lexically specific phrases and gradually develop more abstract representations. Dąbrowska and Lieven (2005) expanded upon this and found that they could account for the linguistic development with regard to question constructions in data from children between 2;0 and 3;0 with minor operations on existing patterns: juxtaposition and superimposition.

This research touches upon two important topics within the CxG framework: the combination of different constructions to form larger strings of language on the one hand, which is an interesting issue that will not be discussed further here (but cf. Lieven 2009), and the development of more abstract generalizations with the concomitant questions of how different levels of representation are related and what levels people use when they produce language. Language users are assumed to have lots of constructions stored, at both the specific and at schematic levels. In natural language acquisition, i.e. outside the classroom, this inventory of constructions, or constructicon, is acquired on the basis of what a language learner encounters

[8] Assuming that such grammatical categories exist independently from the constructions in which they occur is not acceptable for all people working in this framework. Croft (2001, 2005) explicitly argues against it: "*The parts qua instances of a syntactic category do not have an independent existence outside of the whole construction (or constructions) in which they play a role. That is, syntactic categories are defined in terms of the construction(s) in which they occur.*" (Croft 2001:48). He suggests that similar parts of different constructions, e.g. the verb in both the intransitive and the transitive construction, should be represented as related in a taxonomy, with the verb in both constructions as daughters of a superordinate called 'morphological verb'. Croft argues that this is the best way to capture both commonalities (e.g. in tense marking) and differences (e.g. in distribution) between the two constructions (Croft 2005:284).

and therefore always starts with specific representations of a form-meaning pair. Using general, not language-specific cognitive abilities, especially categorization, language learners may develop more schematic representations. Recognizing similarities between different instantiations, they build up more abstract patterns (Tomasello, 2003). In Section 1.5 I describe this development in more detail and argue that it results in representations at different levels of specificity.

Taking language acquisition as a starting point for the description of language users' constructicons and viewing linguistic development as essentially input- and usage-based, means that development of the constructicon must be bottom-up. Children start out with stored pairings of specific forms and meanings. Even as they come to recognize patterns of similarity and their constructicon expands to include abstract representations, a lot of linguistic knowledge is still stored at very specific levels. Extant specific representations are not assumed to be replaced once more general patterns have been discovered. The result of this development is a system which is far from economical in storage: much knowledge is stored for specific instantiations of a construction that could be inherited from a more abstract generalization. Croft (2005:276) signals that a usage-based approach to language is more or less incompatible with the assumption of default inheritance. Default inheritance entails that elements of form and meaning that are shared between different instantiations of a construction are stored at the more abstract levels rather than at the specific levels. Instantiations inherit these aspects from the more general schemas. In a usage-based approach it is assumed that abstract representations develop on the basis of observed commonalities between different instantiations. Therefore, these abstract representations must contain what is shared between the instantiations (e.g. that instantiations of the V-BAAR construction behave like adjectives, that is, they occur predicatively and attributively). A strict interpretation of default inheritance would then exclude that the same fact is also stored at the level of specific instantiations (e.g. that *eetbaar* is adjective-like in distribution). This interpretation is in fact untenable from a usage-based point of view, since it would mean that the information that was originally stored for these instantiations is somehow deleted once a more abstract, general representation has developed.

The conceptualization of linguistic knowledge as the mental inventory of constructions or constructicon means that storage of both lexically explicit forms *and* the schema they instantiate is not problematic. CxG assumes that *"[l]inguistic knowledge comprises vast numbers of constructions, a large proportion of which are "idiosyncratic" in relation to "normal" productive grammatical patterns."* (Goldberg 2006:220), thus shifting the balance from highly abstract rules to smaller patterns and specific instantiations compared with many other theories of language. Within the CxG framework, one of the questions that have received a lot of attention lately is to what extent a construction or family of constructions can best be described with abstract patterns or with more specific representations. Boas (2003) observes for the English resultative construction (NP V (NP) RESULT, e.g. *she drives me crazy, the pond froze solid*) that, although general patterns can and have been recognized (e.g. by Goldberg & Jackendoff, 2005), "*individual verbs lexically subcategorize for specific semantic and syntactic types of resultative phrases*" (Boas 2003:158). An example is the verb *drive*, which is usually combined with a negative mental state, expressed as an adjective (*crazy, mad, insane*). He argues that these collocational patterns mean that instantiations of the general resultative patterns are to a large extent conventionalized. The question then arises if the general patterns are 'real' for speakers of the language. Boas' solution is to propose that non-conventionalized instantiations of the resultative (*Kim sneezed the napkin off the table*) are produced utilizing existing 'mini-constructions' and contextual background information.

The question how psycholinguistically 'real' abstract generalizations are is related to the notion of productivity: a construction is deemed productive if it is possible to form new instantiations of a general or abstract pattern. For Barðdal (2008), who explicitly adopts a usage-based approach to argument structure constructions, it is "*the type frequency and the coherence of a schema [that] determine the actual level of schematicity at which the construction exists in the minds of speakers (...). This level of schematicity, i.e. a construction's highest level, also determines the construction's productivity.*" (Barðdal 2008:45). She seems to imply here that a construction exists at one level of abstractness only, but we know that in addition to any (partial) abstraction or schematic representation language users will have fixed expressions stored in their constructicon. In Section 1.5 I will argue for a model of multiple representations, where speakers have different levels of more or less abstract patterns available to them. Viewed in this way, the research question then shifts from which abstract generalizations exist in speakers' constructicons to when they use which levels of abstraction. This ties in with Boas' suggestion that a language user needs item-specific knowledge for correctly producing resultatives, but may make use of higher-level, more abstract generalizations when decoding resultatives, especially if these are non-conventionalized (Boas 2005, also cf. Boas 2008 for a further discussion of the different role for constructions in production and comprehension).

In sum, for construction grammarians language consists of constructions, which can be lexically specific or (partially) abstract. When people speak, they combine constructions; they use stored specific instantiations as well as abstract representations, for which they need to fill in the underspecified elements. Language acquisition is based on the input children get. They start out by learning specific form-meaning pairs, but on the basis of commonalities they may develop abstract schemas. As their linguistic repertoire increases, children build up a constellation of abstract constructions (NP V (NP) RESULT), low-level generalizations (*drive* + negative mental state) and fixed expressions (*she drives me crazy*).

1.4.2 CLIs in CxG

After discussing some of the basic tenets and issues in the theoretical framework, I now address how various types of complex lexical items can be described as constructions. Much research in construction grammar has focussed on argument structure constructions, both existing (e.g. the resultative, the ditransitive) and made-up constructions, e.g. the appearance construction constructed by Casenhiser and Goldberg (2005). Very often, they focus on verbs and events, either by analysing argument structure constructions, or by looking at specific verbs (e.g. Newman & Rice, 2006 on 'eat'). In contrast to these phenomena, most CLIs have the form of a constituent or a phrase rather than an utterance, full idioms being the exception. This does not exclude the possibility of analysing CLIs as constructions, but it does add a complicating factor: when one looks at actual use of CLIs in corpora, instantiations will always be found inside larger (argument structure) constructions, which may make it difficult to identify which elements of both form and meaning are really part of the CLI construction.

Aside from the scale of CLIs as compared to oft-researched constructions (phrasal vs. sentential), there seem to be two more differences in the distribution of CLIs and argument structure constructions. The first is that some CLIs contain specified elements that only occur inside a particular construction, such as affixes. In contrast, in the resultative construction the slot fillers are all schematic: a verbal phrase and arguments which a language user will come across in other contexts. Even for partially specific constructions like the LET ALONE construction (Fillmore, Kay & O'Connor, 1988), with its fixed lexical elements, those fixed elements may occur outside the construction as in (10), although they are more commonly found within it (11).

(10) the third and most important reason was that she must be **let alone**, to roam on the mountains.[9]

[9] www.tikkun.org/archive/backissues/tik0797/tik6.html, accessed March 2nd 2009.

(11) A: Did the kids get their breakfast on time this morning?
B: I barely got up in time to eat lunch, **let alone** cook breakfast.[10]

For morphological derivations, the affix is a very reliable indicator that a particular form is an instantiation of a particular CLI construction. This is only the case for this type of CLI, however: compounds often entirely consist of words (or stems) that can also be found in many other linguistic contexts (for a discussion of compounds and CxG see Booij, 2009). The same is true for collocations like *rode wijn* or *trots op* (see Section 1.3.1). Generalizations over instantiations of these kinds of CLIs are like those for argument structure constructions: all elements can be underspecified. For morphological derivations, the affix is always lexically specific, even in the abstract representations.[11] These are therefore more like constructions of the LET ALONE type, with the provision that *let alone* can be found outside the construction, while (most) affixes cannot; an important distinction, because it means that the affix will not be stored separately in the constructicon (see Section 1.5).

The third difference between argument structure constructions and CLIs is the (non)prototypicality of the most frequent type. Argument structure constructions and CLIs share the property of a skewed distribution: a small number of types account for a large proportion of all tokens. This Pareto distribution has been observed widely and lies at the heart of collostructional analysis (Stefanowitsch & Gries, 2003; Gries & Stefanowitsch, 2004). In collostructional analysis, the distributional properties of a construction and its instantiations are measured to identify, for example, which verbs are attracted to the resultative construction. That list of verbs can then be used to gain further insight into the semantics of the construction. The skewed distribution has been argued to be instrumental in the acquisition of argument structure constructions (Goldberg, 2006, chapter 4). For argument structure constructions, this is supported by the fact that the most frequent instantiation is prototypical in meaning (cf. N. Ellis

[10] Example taken from Fillmore, Kay and O'Connor (2003); example 15 in the original text.

[11] Some affixes have two (or more) variants, with the actual form depending on the phonological environment. The Turkish nominal plural marker is an example in case: since Turkish has vowel harmony, the plural affix either takes the form *–ler* (with stems containing front vowels) or *–lar* (for stems with back vowels). Since their function and meaning is the same, it is preferable to capture this in one construction, with the vowel in the affix underspecified: N-*lEr*

2005:320). This does not seem to be the case for derivations: the most frequent word in Dutch ending in *–baar* is *blijkbaar* 'apparently'. A large majority of words ending in *–baar* have a verbal stem and mean something like 'can be V-ed' as in *leesbaar*, 'can be read' or 'legible', but this is not the case for *blijkbaar*. Bybee (1995) suggests that morphologically complex frequent forms do not contribute to their pattern's productivity. From the point of view of acquisition, this leaves language learners with a difficult task to master: the types which are most frequent, and will often be encountered relatively early, do not provide them with reliable evidence with regard to the (partially abstract) pattern they instantiate. This issue will be taken up in more detail in Chapter 3, which deals with productivity.

The focus in CxG-research on larger constructions coincides with an emphasis on productivity and general patterns. Although Boas (2003, 2008) argues for so-called mini-constructions, which are specific to the verb in a construction (e.g. *drive* in the resultative construction), very little attention has been paid to completely fixed expressions. The reverse is true in research on formulaic language, collocations etc. (cf. Section 1.2.2). Other than the observation that most of these 'behave' grammatically, i.e. they follow regular patterns of word order, inflection and the like in a language, there is not much discussion of what it means that these larger lexical units are in fact also instantiations of more abstract patterns. Langacker's treatment of complex expressions (his term for any expression that consists of more than one component structure) is an exception: *"The familiarity of a complex expression does not blind us to its componentiality and render us unable to perceive the contribution of individual components. If this were so, the notion of a complex lexical item would be a contradiction in terms: the unit status characteristic of lexical items would entail their immediate and automatic loss of analyzability, removing any grounds for considering them to be complex; all fixed expressions would therefore constitute single morphemes, regardless of size or any resemblance to other units. (…) We need not assume that the component structures are accessed on every occasion when the composite structure is employed, or that when accessed they are necessarily activated at the same level of intensity as they are in a novel expression. However, only when the composite structure loses altogether its capacity to elicit the activation of its components can it be regarded as fully opaque and unanalyzable."* (Langacker 1987:461-2).

This quote captures a number of relevant reflections on CLIs, echoing my earlier critique of the definition of *morpheme equivalence units* by Wray (cf. Subection 1.2.1) that something can be a lexical unit yet still have internal structure. Langacker also suggests that the level of activation of internal structure may vary, depending on occasion of use. This issue will be expanded upon in Chapter 2, and fits well with the model of multiple representations that is introduced in the following section. With regard to the last sentence in the quote, it is unfortunate that Langacker does not specify whether he is talking about a development at the level of an individual speaker or of a speech community as a whole. The individual speaker interpretation is at odds with the bottom-up development that is assumed in an input-based approach to language acquisition. There, the starting point is that children try to attach a meaning to an expression and that they start with unanalysed form-meaning pairs (cf. Tomasello, 2003 for early child language acquisition and Wray, 2002 for differences in this respect between first and later second language acquisition). If the component structure of a complex expression is only acquired *after* the acquisition of an unanalysed whole, the development is towards increasing componential analysis rather than the loss of it.[12] If we read Langacker's comment on loss of activation of components as referring to a speech community, however, this mechanism reflects diachronic changes in productivity.

In sum, whereas some of the basic notions of construction grammar, such as the specificity continuum, the absence of a lexicon-syntax dichotomy and the concept of construction in general seem to lend themselves very well to the topic of CLIs (cf. Wray 2008:87), most research within the framework looks at more general patterns. In the following section, I introduce a model of multiple representations for CLIs which is based on CxG and cognitive linguistic assumptions.

1.5 Multiple representations of CLIs: The MultiRep model.

In this section, a path is sketched along which knowledge of a CLI develops. This path is not essentially different from the acquisition of any other part of the linguistic system: it starts with the mapping of a concrete form to a meaning. Based on the input a language user receives, this representation is extended in two important ways: as language learners hear or read this form in different contexts, they will link it to various other items. At the same time, they will encounter other items with similarities in form and meaning. On the basis of these, they may develop more general representations which are (at least partially) non-specific. This developmental path has been proposed, tested and observed by a number of people who investigate

[12] Of course it is possible that something starts out as an expression that is productively composed from different parts, but as you use this expression more often, it becomes fixed and may become stored as one unit.

language acquisition from a cognitive perspective (e.g. Tomasello, 2003 Lieven et al., 2003). It is also fully compatible with the basic tenets of Construction Grammar (Goldberg, 1995, 2006; Croft, 2001; Boas, 2003). In the figures illustrating the different levels of representation, I follow the conventions of Langacker (Langacker, 1987, 2008), whose analysis of morphologically complex items is not very different from mine, although it is not explicitly based on a developmental perspective, i.e. with acquisition as its starting point. Taken together, the different steps in development constitute a model of knowledge with multiple representations and links between items which I call the MultiRep model.[13] This is also strongly inspired by the work of Bybee (1995) and Dąbrowska (2009). After proposing the developmental path, I look back at the morphological data introduced in subsection 1.2 and discuss how a model with multiple levels of representation can account for these.

[13] Shortly after writing this, I came across Bod's review of Constructions at Work (Goldberg 2006), where he asks 'where is the model? (…) There is no precise definition of (i) the notion of a productive unit in CxG, (ii) the way productive units are acquired step by step from incoming input utterances" (Bod 2009: 130). Whereas what is proposed here may not be formal enough to satisfy Bod, the explicit aim of this book is to give a template for the process of language acquisition that results in productivity of constructions. The notion of productivity is the topic of Chapter 3.

1.5.1 The representation of a CLI in speakers' constructions: A developmental path

The first step in language acquisition is a holophrase: the form is an unanalyzed chunk, with one meaning. This is true for CLIs as well,[14] and is represented in Figure 1.3 below.

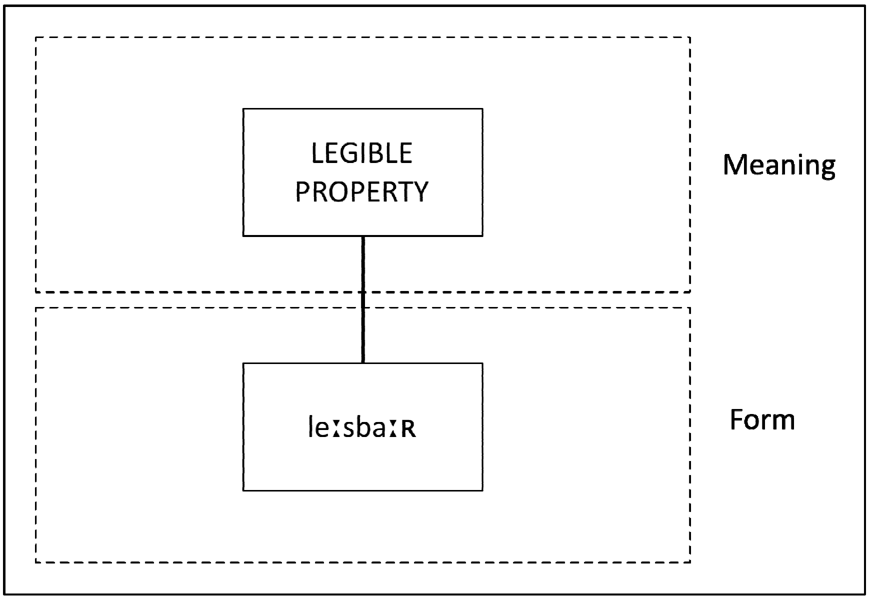


Figure 1.3: Representation of (internally unanalyzed) CLI

In this and the following figures, the conventions as introduced by Langacker (1987, 2008) for representations in cognitive grammar are applied. The boxes represent units, with vertical lines indicating symbolic relationships. For each representation, two levels are distinguished: a form level (the lower box or boxes, with phonological representations) and a

[14] At least initially. Once partially schematic representations have been acquired, it is possible that a new form is immediately recognized as a type of this pattern.

meaning level (the upper box or boxes, with semantic representations in capitals).

As children hear the linguistic unit in Figure 1.3 (*leesbaar* 'legible') more often, they may find that it occurs regularly with the same types of entities. It is not possible to predict the amount of encounters a child needs to recognize the recurrent patterns that are necessary for the extension of the representation. It is very likely that this quantity of input varies for different children and depends on other contextual factors.

Based on the input they get, children may infer that this particular unit is mainly used to describe a property of script (handwriting, fonts, smudgy number plates etc.) or of a written message (municipal announcements, their own essays…). This information becomes part of the representation, as represented in Figure 1.4.

It is at this stage that children will learn information about the elaboration sites of a linguistic unit (cf. Langacker 1987) and about which other linguistic units co-occur with it. In the figures below, schematic units are expressed with '…'. When > is used, this means that linear precedence is assumed. No ordering is presupposed in all other cases when boxes co-occur horizontally, i.e. when more than one form/meaning is part of one lexical representation. This ordering will be determined in actual utterances by the larger constructions (phrases, argument structure constructions etc.) with which they are combined.

Note that the phonological part of the entity this property is ascribed to is left unspecified. The representation given here is one of a partially specific construction: one element is lexically specific while another is phonologically unspecified. Different lexical units may fill this unspecified slot, as long as they match with the semantic notion given in the top right box in Figure 1.4. No ordering is assumed, as this depends on the larger structure, e.g. the attributive or predicative use of *leesbaar*.

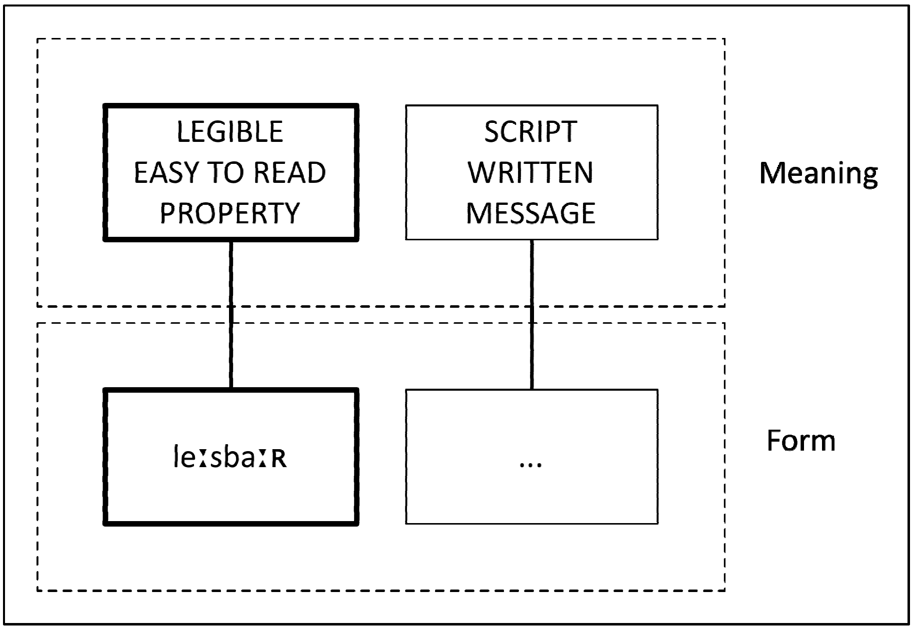


Figure 1.4: Representation of distributional knowledge for *leesbaar*

At the same time, children may develop representations of specific instances of the schema given in Figure 1.4 for collocations they encounter frequently, such as *leesbaar handschrift* 'legible handwriting'. The representation of this collocation is given in Figure 1.5. Again, no ordering is assumed, as this collocation may be found both in the form of a complex noun phrase as in (12) or in a copular construction (13).

(12) Hij heeft een leesbaar handschrift
He has a legiblehandwriting

(13) Haar handschrift is leesbaar
Her handwritingis legible

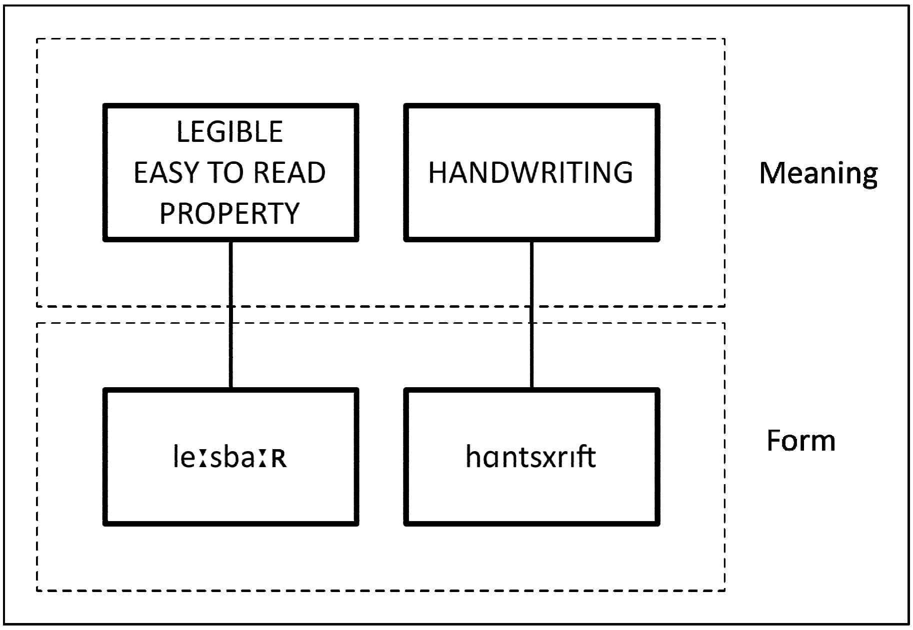


Figure 1.5: Representation for the collocation *leesbaar handschrift*

These examples show that CLIs do not occur as isolated units, but in meaningful discourse. So far, I have looked at the representations of *leesbaar* as if it were acquired independently from any existing linguistic knowledge. Of course this is a gross simplification. By the time children learn *leesbaar*, they already know hundreds if not thousands of other words. Psycholinguistic experiments have provided evidence for the fact that there are lots of links between lexical representations. An impressive body of work on priming effects (among many others cf. Feldman, 2000; McNamara, 2005) has shown that words which are either similar in form or in meaning (i.e. they share part of their distribution) are linked in some way. Having just seen a word like *bug* means that you recognize *fly* faster than if you had not.

As children encounter the same unit (in our case *leesbaar*) more often, they will not only extend their representation to include common patterns of combination, but may also recognize that it is similar in both form and meaning to other units e.g. *lezen* (to read). This makes it possible that they come to recognize the internal structure of the hitherto unanalyzed unit. The verb *lezen* shares elements in both the phonological and the semantic units with *leesbaar* and possibly also collocates, frame elements etc. It is a reasonable assumption that links between different representations exist on the basis of shared elements in meaning and/or form. Bybee (1995) suggests the same: *"Words entered in the lexicon are related to other words via sets of lexical connections between identical and similar phonological and semantic features. These* *connections among items have the effect of yielding an internal morphological analysis of complex words,(...) their morphological structure emerges from the connections they make with other words in the lexicon"* (Bybee 1995:428-9). The existence of links at multiple levels may lead to the analysis of *leesbaar* as consisting of two elements. This is depicted in Figure 1.6.

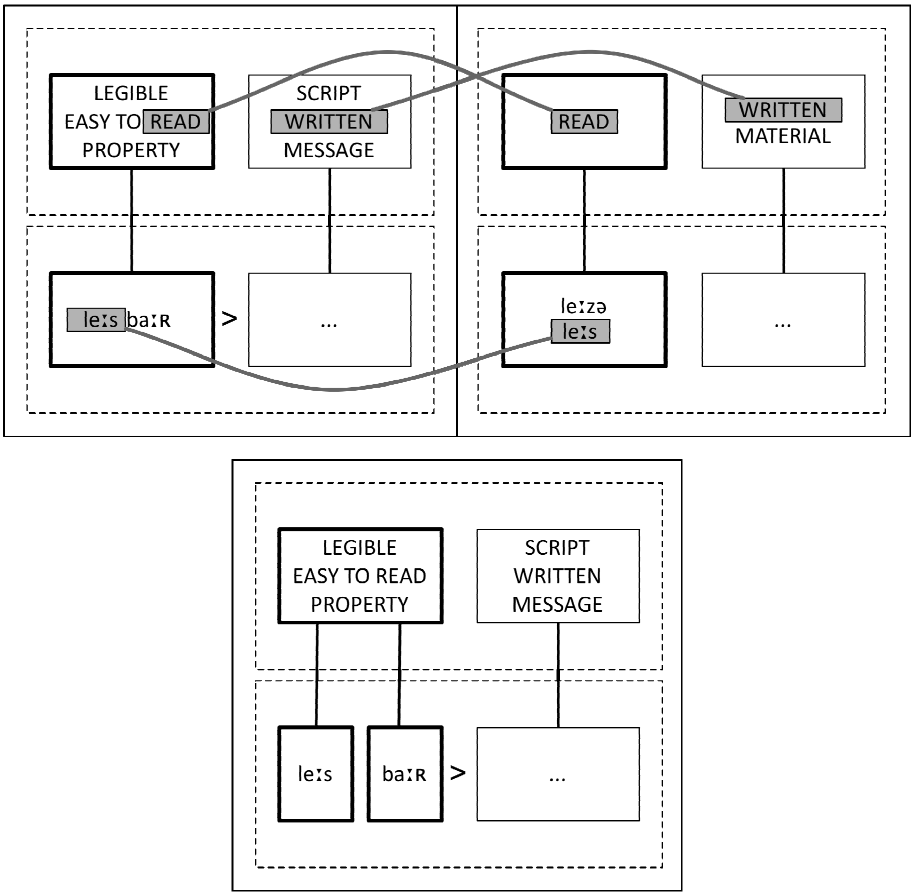


Figure 1.6: Representation of the initial morphemic analysis of *leesbaar* based on similarities with *lezen*, through links at the level of form, meaning and distributional properties. The resulting morphemic analysis is depicted in the bottom half of the representation.

As more and more links between *leesbaar* (left top half) and *lezen* (right top half) are made on the basis of shared elements in input, the representation of *leesbaar* is extended to include a first morphemic analysis (bottom half). At this point it is important to emphasize that the original representation is not 'overwritten' or deleted. Rather, a new level is added.

This is essentially the same process that happened between the representations in Figures 1.3 and 1.4. There too, I do not wish to suggest that the second representation replaces the first one.

The starting point for the morphemic analysis here are the links with the lexical representation for *lezen*. Note that this initial morphemic analysis does not yet include a clear representation of the semantic contribution of both morphemic elements. For this to develop, a child needs to encounter other lexical units with *–baar*. These will have similarities in both form (the actual morpheme) and meaning (some notion of potentiality) to *leesbaar*. Over time, this will allow for the development of a representation in which the semantic contribution of both morphemes is present, as depicted in Figure 1.7 (collocates and other stored information are not presented in this figure for reasons of clarity).

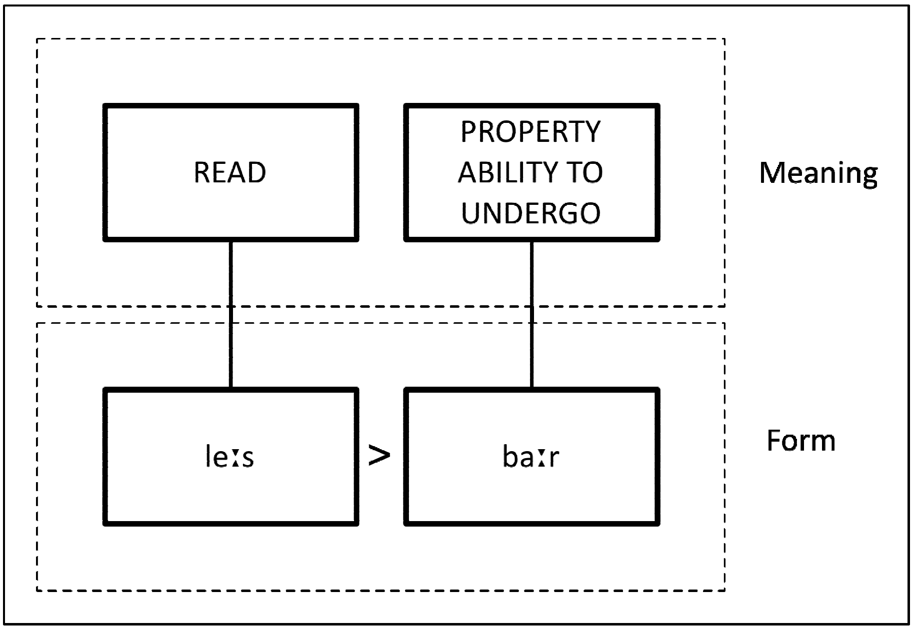


Figure 1.7: Representation of extended morphological analysis of *leesbaar*

Finally, on the basis of this representation of *leesbaar* and other lexical units with the same affix and similar meanings, a partially schematic representation may develop of the *–baar* construction (Figure 1.8). This contains the phonemic and semantic representation of *–baar*, a maximally schematic phonemic representation of the item *–baar* is combined with, and an underspecified semantic representation for this item, which includes the similarities between the items on which this schematic representation is based. If *-baar* occurs in very different contexts, children may well develop more than one (semi)specific representation or mini-construction (cf. Boas, 2008). These different representations are of course linked through the strong overlap at least on the form side.

It is this final representation that speakers of a language can use when they encounter lexical items with this affix that are new to them and when they productively form novel instantiations of the construction. The semantics of *leesbaar* include the notion PROPERTY. Through this, it can be linked to a more general representation of adjectives. In this way even very general and abstract categories like basic word class can be part of the same system of representations. There is no essential difference between concrete lexical representations and (much) more general ones, other than the fact that not all elements are equally specific.

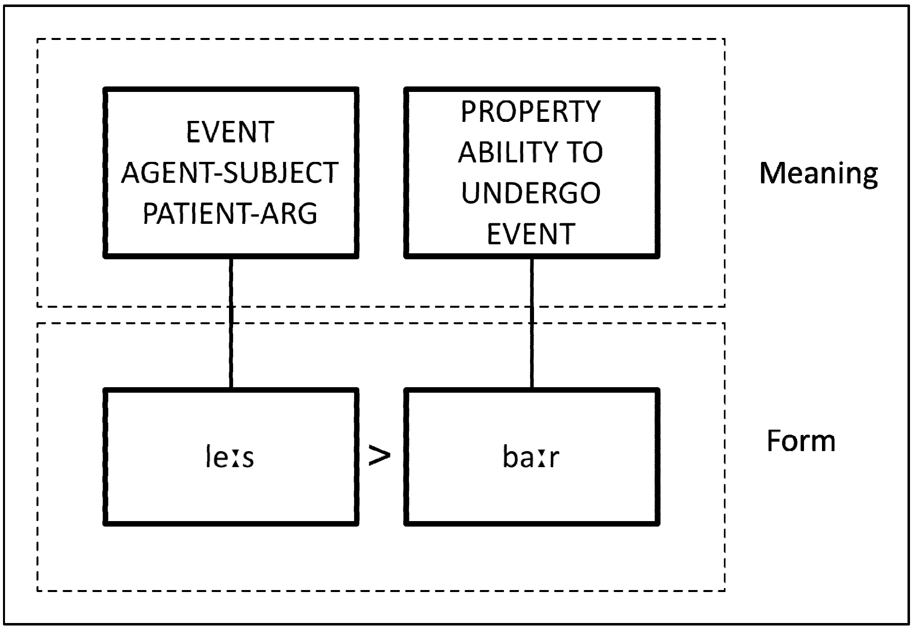


Figure 1.8: Partially schematic representation of *-baar*

The representation in Figure 1.8 seems similar to but should not be confused with analyses in which an affix is seen as a schematized adjective, noun etc. Langacker (1987:334) seems to assume this where he states that "[*t]hus –ful is analyzed as a schematic adjective, –er as a schematic noun (...) and so forth*". He then proposes a representation of an affix including schematic elements in its phonological pole. In my opinion, this should not be called the representation of the affix, but one of the partially schematic construction that has the affix as a (or the only) phonologically fixed element. I claim that it is not the suffix *–baar* which is a schematized adjective, but the partially schematic construction. An adjective is an instantiation of this construction; the affix is not. Note that there is no representation for *–baar* outside of complex representations (as in Figures 1.6-1.8). No such separate representation is assumed because *–baar*, in contrast to instantiations like *leesbaar*, does not occur outside of these restricted contexts. Langacker himself is not completely consistent in this regard: although he states explicitly that he views the affix as a schematic noun (see the quote above), earlier he suggests that "*the structures to which unit status can be attributed involve more than just the elements with explicit segmental manifestation. The instances just cited are more adequately given as (…) [STEM-ful-ly], where the entities represented by capital letters are substructures characterized only schematically*" (Langacker 1987:314). It is this suggestion that I follow here.

Langacker uses an example from Classical Nahuatl *neki* (want), which occurs both as a verbal stem (14) and as a suffix (15).

(14) ni-k-neki ni-k-ita-s
I-it-want I-it-see-FUT *I want to see it.*

(15) ni-k-i-ta-s-neki
I-it-see-FUT-want I want to see it.

The difference between these two distributions he conveys through two representations, one of which is partially schematic. My model would not be very different apart from the names for the representations: one (the verbal stem function) can be described in a fully specified lexical representation such as Figure 1.3 (which can then be used to fill slots in complex constructions) and the other takes the form of a partially schematic representation as in Figure 1.8.

For all constructions acquired up to the level of partially schematic representations, a language user ends up with multiple representations for the same structure (i.e. lexically specific and more abstract). These multiple levels are linked to each other through their shared aspects of meanings and form. Langacker, whose analysis of CLIs is very similar to the one presented here, summarizes this pictorially in one figure (Langacker 1987:479, Figure 12.10), containing both the abstraction and the specific instantiation. To a large extent, the difference with the current proposal (distinct, but linked representations) is a matter of choice rather than contrasting insights. Psycholinguistic research seems to indicate, however, that language users access different levels of representation depending on the linguistic task they execute. It is easier to account for that fact in a model that depicts these different levels as separate, albeit linked representations.

1.5.2 Advantages of the MultiRep model

1.5.2.1 Acquisition

The development described for *leesbaar* in the previous section provides a path along which all constructions may be acquired. This path is compatible with observations made by Tomasello, Lieven (Lieven et al., 2003; Tomasello, 2003) and many other linguists who look at language acquisition from a cognitive perspective. Tomasello's description of the acquisition of verbal templates is essentially the same as the developmental path described here: children start out with concrete lexical items (e.g. *throw*) for which they develop a specific argument structure (in the case of *throw* there is a *throwing person* and a *thrown object*). As and when they are exposed to more instantiations of different verbs and develop more of these specific templates, this will lead to the generation of a more abstract schema including a notion like Agent-Subject and Patient-Object. In Tomasello's model, as in the one introduced here, representations initially are specific and their generation is input-based. On the basis of similarities between different instantiations, more abstract representations will develop. Tomasello and Brooks (1999) point out that this development is driven by general cognitive abilities: "*As children learn an ever-increasing number of complex constructions they are also extracting commonalities among these structures and so moving toward a more abstract form of linguistic competence. They presumably do this in the same way that they form schemas, scripts, and categories in other cognitive domains, that is, by extracting commonalities of both form and function.*" (Tomasello & Brooks 1999:179). There is an abundance of evidence that children are able to form categories and see patterns in their linguistic input (cf. Saffran, Aslin & Newport, 1996 and Mintz, 2003 for phonological patterns and grammatical categories, respectively).

Research investigating children's proficiency in using morphologically complex words tends to show a clear developmental pattern with gradual growth in adult-like performance (e.g. Carlisle, 1988; Anglin, 1993, cf. subsection 1.1.6) rather than all-or-nothing scores. This finding, repeated in experiments I describe in Chapter 2, seems to point out that more abstract levels of representation develop gradually and may not always be accessible: recognition or the correct use of an affix in one context by a child does not guarantee similar linguistic behaviour with another lexical item.

Thus, different levels of representation co-exist, with links between more and less specified representations. This may mean that the storage of information is far from economical. Models which assume full inheritance need rather less storage space because they assume that as much information as possible is stored in general schemas. It entails that in those models the lexically specific representation of *leesbaar* does not have to include the notion PROPERTY: the existence of a more general schema with *–baar* already includes that aspect of meaning and use. This type of full or complete inheritance models can be distinguished from 'normal' inheritance, which allows for inheritance without excluding the possibility of subregularities. Goldberg (1995) describes Construction Grammar as using normal inheritance with *real copying*: "*each construction is fully specified, but is redundant to the degree that information is inherited from (i.e. shared with) dominating constructions.*" (Goldberg 1995:74). The levels of representation I sketch in this chapter have not been discussed in terms of 'dominating' and mother or daughter constructions, but there is a clear hierarchy with regard to the degree of explicitness of the different levels. In this, I follow Goldberg's suggestions, because full inheritance is not possible from the point of view of acquisition, as has become clear in the sketch of development made in this section.

1.5.2.2 Dynamicity and variation in representation

In subsection 1.5.1, I described the development of the representation of a CLI. For the particular example there, *leesbaar*, I assumed that on the basis of encountering this lexical item in different contexts as well as hearing and reading instantiations of both *lezen* and other *–baar* words, a morphemic analysis takes places which ultimately results in the development of a partially specific schema for *–baar* (Figure 8). Not all CLIs will undergo such an analysis, and the levels of representation need not be the same for all speakers of a language. Someone who does not have a representation as given in Figure 6, i.e. with both morphemes present, may still be a fluent speaker of the language and a completely native-like user of the lexical item *leesbaar*. In fact, speakers who do have a representation as in Figure 1.8 will also have representations like Figure 1.5 in their constructicon. It may not be possible to distinguish between these speakers in terms of their production and understanding of this lexical item. It is only when novel forms with *–baar* come into play that a representation without morphemic boundaries cannot suffice to decode these forms.

The research in the Chomskyan framework described in Section 1.1, which focuses on highly productive processes, basically equates morphology with representations at the most abstract, general levels. Halle and Marantz (1994) assume that morphology operates before positions are lexically filled. Reformulated in terms of my multiple levels of representation model, it seems that what Halle and Marantz see as morphology, is only the most general representational level. Since the actual morphological features are stored with the specific lexical items, more specific representations are part of the lexicon. I have argued, however, that the different levels of representation are linked because they are part of a developmental path, whereas Halle and Marantz see these levels as two essentially different modules of the linguistic system.

Different levels of representation will co-exist but they will not all be equally active all of the time. We know from priming studies that speakers can recognize words more quickly if they have been exposed recently to lexical items that are linked to this word. The results of Duñabeitia et al. (2008) show that priming also occurs for partially abstract patterns with affixes. Priming effects are generally described in terms of activation: the use of a specific lexical item partially activates other items it is linked to. This then results in shorter recognition times for those items. Such activation patterns are very useful for speakers: we have a lot of lexical items in our inventory. If all items were activated to the same degree, it would be difficult to decide for speakers which item they are hearing. In language use, we find that the use of one lexical item can be a strong predictor for the occurrence of another; some items co-occur very frequently. If one of these items is recognized, it makes a lot of sense to already partially activate the other, as it is very likely to occur next.

Different activation levels are assumed to exist independently from priming effects. In lexical decision tasks, it is generally found that recognition times are strongly correlated with frequency, with more frequent words being recognized much faster. This too makes sense from the perspective of a language user: it is a reasonable hypothesis that a word which has occurred frequently in the past is more likely to occur in the direct future than another word that has not occurred as often. With regard to the different representational levels discussed in this section, this may mean that the lexically specific level is usually activated more than the underspecified ones, since these can be mapped directly to the input. Dąbrowska (2008) cites psycholinguistic evidence from which she concludes "*that mental grammars are organized redundantly (e.g., that low-level schemas and specific exemplars co-exist with more general rules), and that speakers prefer low-level generalizations.*" (Dąbrowska 2008:934). This is completely in line with the current proposal. An open question, deserving of more empirical attention, is in which contexts people are more likely to use the more general patterns. The experiments reported on in the following chapters try to shed more light on this.

The debate about processing regular and not so regular forms (e.g. in the dual route approach) can also be reformulated in light of the suggestion that people have multiple levels of representation: processes that can be described with a very general rule, applicable to a lot of instantiations, are more likely to be processed with general, more abstract representations. Subregularities, as observed by Albright (2002) and Dąbrowska (2008) are captured in more specific representations. Frequent forms are stored as completely specific units, regardless of whether they are regular (i.e., the complexity can be described in a more general representation) or not. This means that there is no need to postulate two very different mechanisms, *contra* the dual route model. Our hypothesis is that people will make use of the representation that maps most easily on the input they receive. Often, this will result in the activation of very specific representations, because these require less processing of the input from the complex and abstract representation to a mapping of the concrete form to meaning.

As this model is inherently dynamic, individual variation as well as changing linguistic 'behaviour' from one person through time is to be expected: activation levels change over time and new representations can always be generated on the basis of new input. Since representations are based on the input people get, and this input will be different for each person, we may expect large differences in people's linguistic knowledge. Such individual variation is generally accepted at the lexically specific level (differences in vocabulary size) and with regard to lects that people speak, something clearly related to the input speakers get. However, letting go of the distinction between lexical knowledge and syntactic knowledge means that variation is expected to exist at all levels. This is not acknowledged often in linguistic research, although there are some exceptions (Chipere, 2003; Dąbrowska & Street, 2006). [15]

Different people may have different levels of representation available to them, but this does not impede mutual understanding as lexical items can be accessed from different levels. Change over time is also a phenomenon which is to be predicted. After a number of years, many of the more frequent lexical items will have established links and representations. Their basic activation level (outside of specific discourse contexts) may also be relatively stable. Adding new senses or collocates to existing representations or new lexical items altogether can be accounted for in the same way as earlier acquisition. It is also understandable that learning new items at a later age is more difficult: many of the representations that are in the inventory of a speaker are deeply entrenched. These firmly established patterns are quite robust, compared with 'newcomer' words or patterns.

[15] In theories that take syntax to be innate, variation in syntactic linguistic knowledge would equate with genetic deficiencies or malfunctioning of the acquisition device. In the present work, I assume that language is learned like other -complex- cultural skills, for which it is generally accepted that not everyone is equally proficient.

1.5.2.3 The Constructicon

In the model described here, there is no essential difference between the acquisition and representation of concrete lexical items and more abstract and productive schemas. All of these are described in terms of form and meaning and can be acquired on the basis of linguistic input. As such, the resulting inventory of linguistic units will include both lexically specific items (words or larger units) and partially or completely abstract (i.e. underspecified) constructions. This was already articulated clearly by Langacker (1987) for morphologically complex expressions: "[*t]o the extent that the expression has unit status, all these structures and relationships [semantic, phonological and symbolic, MM] are part of a speaker's knowledge of linguistic convention.*" (Langacker 1987:347). As Achard and Niemeier put it: "*The goal of a usage-based-model is not to achieve mathematical elegance, but to depict the complexity of language use. Consequently, it is composed of an eclectic array of expressions at different levels of complexity, abstraction, and generality. Individual lexical items cohabitate in the system with idioms, conventionalized idiosyncratic collocations, and fully productive grammatical constructions.*" (Achard & Niemeier 2004:5). A number of the different levels Achard and Niemeier mention can be identified in the representations given in Figures 1.3 through 1.8: an individual lexical item (Figure 1.3) a conventionalized collocation (Figure 1.5) and a productive construction (Figure 1.8). By assuming that lexically specific, partially schematic constructions and highly underspecified patterns are stored in the same way, with the only difference being the level of specificity, the dichotomy between lexicon and syntax is clearly abandoned. I argue that this is a desirable aspect of the model, as it allows for small-scale generalizations, something which Construction Grammar has placed a lot of emphasis on. Various people (Goldberg, 1995, 2006; Boas, 2003, 2008) have made it clear that people have representations of such constructions, which range from patterns without any fixed lexical element (e.g. the resultative construction) to partially specific ones (e.g. the V-ING TIME AWAY construction, the LET ALONE construction). For all of these, the meaning of the whole construction cannot be deduced from the combination of lexical items. At the same time, speakers may have representations of specific instantiations of these constructions (she sneezed the foam off the cappuccino, the sooner the better, dancing the night away). In my opinion, these facts can only be accounted for in a model that incorporates lexically specific representations, partially schematic and fully abstract ones in one system, in which these different levels are linked: a constructicon.

The representations also include information about collocations, arguments and other distributional aspects of the items. From corpus studies and psycholinguistic experiments it has become clear that language users have a substantial amount of information stored about distributional properties including various frequency aspects (e.g. collocational patterns between different lexical items, collocations of lexical items and constructions and associations of lexical items with certain registers or genres etc.). The question what information must be included in the representations is taken up at the end of this book (Chapter 5).

1.5.2.4 The role of frequency in the storage of CLIs

The model introduced in this section assumes that for many CLIs various levels of representation co-exist. These different levels -lexically specific unanalyzed unit, specific morphemic analysis and partially schematic representation- can all be accessed in language production and reception. They are also linked to each other, which means that use of one level will (slightly) activate more concrete and more abstract levels. The assumption that there are such links and that both the links and the different levels may be activated to a higher or lower degree provides a simple but convincing explanation for a whole range of observations with regard to the storage of CLIs and their frequency.

In our model, Bybee's hypothesis that frequent regular complex forms do not contribute to the productivity of the pattern would be interpreted as weak links between the lexically specific representation and the partially schematic construction of which it is an instantiation. For CLIs at the opposite end of the frequency scale, the reverse is true: if speakers want to generate a novel complex form, they must use the partially schematic representation or productive schema. Although the present model has no problems in incorporating the relative independence of frequent types from a productive construction, this leaves open two important issues. First of all, it begs the question of learnability: if frequent types are often stored as unanalyzed chunks, how do children ever get to the development of more abstract and general representations? Secondly, there is a need for quantification: how frequent must a form be to be stored independently and how many different types do speakers need to encounter before they develop a more general representation? The answer to these questions is probably that it depends -on the speaker, on the linguistic context etc. The multiplicity of relevant and possibly confounding variables here, however, must not deter linguists from trying to tackle this issue, as it is central to the notion of productivity (cf. Chapter 3).

The phenomenon of *semantic drift*, where an instantiation of a productive construction becomes semantically opaque, can be accounted for quite easily: when the word *dirty*, for example, was first coined, the productive construction N-y, meaning 'property of being similar in some respect to N' was used. As speakers of English started saying it more often, some may have stored it as a morphologically complex unit. At this point, there would still be strong links with the productive construction and with the lexical representation of *dirt*, but this need not remain the case. Both within and across speakers, these links can become weaker. In the case of *dirty*, phonological accommodation between stem and affix results in a reduction of the similarities in form between the independent word and the stem in the first part of the complex word.

In the MultiRep model effects of *relative frequency* (Hay & Baayen, 2002) would be explained in the form of a difference in activation level. If a CLI contains a morpheme or word which is very frequent, the link to the representation of that word or morpheme will be very influential as that element of the CLI will be strongly activated on a regular basis. When a particular CLI is relatively frequent compared to its morphemic elements, as is the case with *preside* and *president* (a Google search on Feb. 16 2009 showed 2.5 million hits in English for *preside* and 366 million hits for *president*), this activation is a lot weaker. Effects of relative frequency show once more that linguistic units are stored in an interdependent, structured network.

Finally, the model accommodates the *morphological family size effect*: a word like *werk* will be linked to many other words (*bewerken, werkuren, werkster* etc.) which will help in reaching a criterion activation level necessary for the recognition of the word. The fact that it is a type effect indicates that it is the existence of a link *as such*, which is the facilitator. It does not seem to matter how frequent each of the linked words are (in which case it would have been a *token* effect). For a more extensive discussion about the *family size effect*, see Section 2.4, which reports on a lexical decision task done with children.

In this section, I described the developmental path along which knowledge of a CLI may grow based on the input a language learner receives. This developmental path results in multiple representations as the model for completely acquired linguistic items and structures. I showed that such a model can accommodate very different research findings in the fields of morphology and larger lexical units.

1.6 Summary and preview

In this chapter, a model of multiple representations was introduced to reflect people's knowledge of complex lexical items (Section 1.5). This model is based on an input-based approach to language acquisition, in line with work of cognitive linguists such as Tomasello (2003), Diessel (2004) and Dąbrowska (2008): the basic assumption is that language learning consists of mappings of forms to meaning. As language learners encounter more examples of a form-meaning pair in different contexts, they extend their representations to include collocations, common patterns, etc. Hearing and reading other examples which are similar in form and meaning provides them with evidence to analyse a pairing as internally complex, i.e. containing elements that are meaningful, and to develop more general, partially abstract representations.

The model that was outlined draws on insights in cognitive linguistics and construction grammar, in work by Langacker (1987, 1991), Bybee (1995) and Goldberg (1995, 2006), among others. Section 1.4 reviewed some of the work that has been carried out over the last fifteen years within these frameworks. Of particular importance is the assumption that constructions exist at different levels: more concrete or specific as well as underspecified or abstract patterns. Whereas construction grammarians have mainly focussed on argument structure constructions, their insights are equally applicable to complex lexical items. A second vital point, at the heart of cognitive linguistics, is the belief that children acquire language using general cognitive mechanisms such as categorization and shared attention. Taking this assumption seriously means that one cannot postulate representations of linguistic knowledge that are unlearnable. It is with this in mind that the model is based on the development of knowledge which in turn depends on the input children get. Thirdly, and in contrast to some other theories of language, these approaches do not stipulate a dichotomy between syntax or rules on the one hand and lexicon or fixed elements on the other. The concomitant existence of constructions at different levels of specificity entails that there is a continuum from specific to more abstract and general representations.

The model was introduced to describe the knowledge of complex lexical items that people may have, although it is not limited to such items. In Section 1.3 (p. 24), complex lexical items were defined as *"strings of language in which more than one meaning-carrying element can be recognized yet are very likely candidates to be stored as units in people's linguistic repertoire."* Both the complexity and the unit aspects of this definition are discussed in more detail in Section 1.3, using a number of examples. From these examples it is clear that CLIs may consist of just one word or a larger combination of words. As such, they span the traditional field of morphology (see Section 1.1) and fixed expressions (Section 1.2).

After introducing the model, placing it in the context of a theoretical framework and outlining the linguistic phenomenon in this first chapter, it is now time to test the model and look at more concrete examples of CLIs. Chapters 2 through 4 each contain descriptions of experimental tasks performed by children aged 9-12. Research focussing on the acquisition of derivational morphology has shown that knowledge about these constructions seems to develop around these ages. In each case, the constructions tested are Dutch, as the research was carried out in the

Netherlands. For many of the participating children, this is not the only language they are acquiring. This is a reflection of the diversity in language backgrounds in today's Dutch classrooms. Depending on the experiment discussed, the results are compared to frequency measures, adult task performance and corpus analyses.

The next chapter describes three experiments I carried out. These experiments were designed to investigate what levels of representations children have at their disposal when they are asked to do different linguistic tasks. The experiments include online and offline tasks and are set up to measure different aspects of children's language proficiency. For each experiment, I identify what type of knowledge can be used and interpret their performance in terms of the MultiRep model. The main topic that is addressed, therefore, is how different tasks trigger the use of different levels of representation.

Chapter 3 is a case study, where two constructions are investigated: a deverbal adjectival construction and a modal infinitive construction. Semantically, these two constructions are highly similar. The research question in this study is whether we can find evidence that the constructions are productive for individual speakers. To answer this question, both corpus data and experimental data are used. The focus on productivity implies, that this chapter seeks to find evidence for the use of an abstract level of representation for these two constructions. Since it is unclear whether this level of representation is part of the linguistic repertoire of adults, the performance of children on an acceptability task of novel instantiations is compared with adult speakers' responses to the same task.

Chapter 4 reports on another experiment, designed to investigate the unit status of two-word CLIs like *trots op* 'proud of'. The results of this experiment are compared to the predictions of a computational model, which assigns a predictability value of the next word to any three-word combination on the basis of co-occurrence patterns. I discuss how the computational data match up with the observed linguistic behaviour and in what ways this can be incorporated in the model. In contrast to the preceding chapters, this case study does not centre not on abstract levels of representation, but on finding evidence for the storage of specific CLIs. It also discusses what elements might be part of the representations.

Finally, Chapter 5 provides a summary and discussion of the data. I conclude with some comments on the relevance of this model for linguistic theory and make suggestions for further research.

Chapter 2:

Investigating children's knowledge of CLIs[1]

2.0 Introduction

This chapter deals with the question what children know about Complex Lexical Items and the patterns they instantiate, by looking at data from three different experiments in which morphologically complex words play an important role. When children have acquired the most frequent words or basic vocabulary, many of the subsequent items they encounter are morphologically complex. For Dutch, it has been shown that words in texts for children from grade three upwards tend to be morphologically more complex and are longer than words in texts for younger children (Wauters, Tellings, Van Bon & Van Haaften 2003). Green et al. (2003) estimate that for English approximately 60% of new words acquired by school-age children are morphologically complex with clear internal structure. In this respect Dutch is typologically similar to English, so it is reasonable to assume that children who learn Dutch face a comparable task. Trying to answer the question what children know about such items and patterns means that you need to find ways to measure this knowledge. It turns out that this is not a straightforward issue: children do not perform equally well on different types of tasks that are set up to measure some aspect of CLI knowledge. Each task places its own demands on children in terms of memory, awareness, experience etc. and as we will see, different linguistic tasks may trigger activation of different levels of representation.

The main part of this chapter consists of reports on three experiments: a word formation task, a definition task and a lexical decision task. These tasks each trigger the use of different levels of representation as defined by the MultiRep model. A word formation task (Section 2.2) may be performed using either lexically specific representations for the target word or

[1] A significantly shorter version of this chapter in Dutch has appeared as Wat weten kinderen nou eigenlijk helemaal van derivationele morfologie? [What DO children know about derivational morphology?] in the Sixth Anéla Conference Proceedings, 2009. That contribution was co-authored by Anne Vermeer. The results of the lexical decision task (cf. Section 2.4 in this chapter) were also reported on in Mos (2006) Complexe woorden in het mentale lexicon van kinderen. [Complex words in children's mental lexicon]. We are grateful to the audiences at the 2006 Anéla Junior Researchers' Day in Utrecht and the 2009 Sixth Anéla Conference in Kerkrade for their comments.

(partially) abstract representations of the affix constructions that are tested. For a definition task (Section 2.3) children need to make use of abstract representations, because they are dealing with unknown words for which they do not have any lexically specific representations. The lexical decision task (Section 2.4) test whether there is an effect of morphological family size in children's reaction times. The presence of such an effect can only be explained if we assume lexically specific representations with internal structure, and links between items that are similar in form and meaning. Before we describe these experiments in detail, the following section focuses on different kinds of knowledge, data and tasks. In the final part of this chapter, the experimental results are discussed in light of the MultiRep model that was introduced in Chapter 1.

2.1 Measuring knowledge of CLIs

Within linguistics, traditionally a distinction has been made between competence and performance, with performance not always fully reflecting a speaker's competence. If someone tries to convey their feelings about an emotionally loaded subject, while making dinner and with the television blearing in the background, they may not be able to formulate complex sentences, monitor the choice of words and be fluent at the same time. N. Ellis (2006) defines the distinction as follows: "*Competence and performance both emerge from the dynamic system that is the frequency-tuned conspiracy of memorized exemplars of use of these constructions, with competence being the integrated sum of prior usage and performance being its dynamic contextualized activation.*" (N. Ellis 2006:100). Clearly, part of the variation in language performance is due to contextual factors, other cognitive tasks that compete for attention etc. The problem here is that we cannot look at competence directly: we can only observe linguistic behaviour and have to draw conclusions about people's skills from that. Fortunately, it is possible to compare performance in different situations and on varying tasks. By taking into account what kinds of (non-linguistic) demands are placed on speakers in the execution of each task, an approach that combines different sources of evidence for linguistic knowledge can approximate a fuller picture of linguistic competence.

With regard to knowledge about complex lexical items, the previous chapter argued that speakers have multiple levels of representation at their disposal. In addition to having lots of specific items stored, with information about usage, collocational patterns etc., they may also have representations of generalizations over the specific types. These are (partially) underspecified schemas containing the same information as the lexically specific items, but with slots that can be filled by combining the pattern with other items. Different levels of representation co-exist, which means that for many sequences there is more than one way they can be produced or decoded: in using unkind, a speaker may retrieve the whole word from storage as a unit, or combine a partially specific schema UN-ADJ = negation-property with the stored unit kind.

While they are using language,[2] speakers can employ whichever representation is most easily retrieved at that point in time. Generally speaking, because specific representations match closely with the actual linguistic output, these are preferred in language processing over more abstract representations that require (more) (de)composition (cf. Dąbrowska 2008:934). But on some occasions, the underspecified pattern will be drawn upon. Speakers need to resort to schematic representations when there is no direct match for the sequence, i.e. when they have no stored representation that is exactly the same as what they need to express or hear. This is always the case when they hear or want to produce a novel combination of lexical items and patterns: instantiations of a productive pattern (cf. Chapter 3). Whereas word sequences are often fixed combinations, few utterances will be completely stored. At this level, productivity seems very common, e.g. in argument structure constructions.

More general patterns may also be employed because that pattern has already been activated in previous use. An example that this happens in speech, is the constructional priming effect: speakers are more likely to use a given pattern if they have just heard that pattern, even when it contained different lexical items. This effect has been shown in the production of ditransitive or prepositional object utterances (Kaschak, Loney, & Borreggine, 2006), for second language learners (McDonough, 2006) and even cross-linguistically, with the activation of a pattern in one language influencing the selection of the equivalent in another (Salamoura & Williams, 2006). Language users do not control which levels of representation are activated at any one moment. This activation is determined by short-term and long-term contextual factors, such as frequency and recency.

[2] At this point, we make no distinction between production and reception, and between oral and written language. Since these place very different demands on the language user, they are bound to influence the levels of representation that are activated and will be used.

Overall frequency of a sequence or pattern has been shown to influence lexical decision times. This factor is commonly incorporated as a variable or controlled for in such tasks, viz. Allegre and Gordon (1993) for experiments in English, Bertram, Schreuder and Baayen's (2000) experiments in Dutch, and Lehtonen et al. (2006) for studies on bilingual speakers. This effect can only be accounted for if we assume that frequent items are activated more easily. Alongside these baseline differences in activation level caused by variation in frequency, there is an abundance of psycholinguistic evidence that recent occurrence also influences activation levels. Priming studies, which investigate the effects of previous exposure on reaction times, show that activation is elevated most strongly in case of full repetition, i.e. when the prime is identical to the stimulus that participants react to (cf. Schwanenflugel & Gaviskà 2005:1740-1741 for an overview). Semantically related primes (cat-dog), and morphologically related primes (baker-bake) also speed up recognition. Some researchers have attempted to tease apart effects of form (orthography) and meaning (semantics) overlap (e.g. Feldman, 2000; Raveh, 2002; Diependaele, Sandra & Grainger, 2005). They find that morphological priming can be reliably distinguished from simple orthographic similarity between prime and stimulus, but only marginally from semantic similarity. Embedded words can also inhibit responses: Bowers, Davis and Hanley (2005) found, for example, that rejecting *hatch* as a piece of clothing was slower than rejecting the same word as a body part. They suggest that this is due to interference from the sequence *hat*. Drews and Zwitserlood (1995) observed a negative influence on reaction time of primes that were orthographically but not semantically related to the stimulus (e.g. Dutch *kerst-kers* 'Christmas'-'cherry'). All these studies show that activation is influenced by experience, both accumulated over a long time (general frequency effects) and recent occurrence (short-term priming effects).

In spite of their influence on language production and reception, activation levels are not something speakers are usually aware of. Even when they are conscious about what they know ('explicit knowledge', see below), being able to formulate and define this knowledge is a separate, though not independent, skill. This means that when speakers are unable to define a pattern, this does not entail that they cannot use it. Conversely, speakers may conform to other native speakers in their use and definition of specific lexical items but that does not prove that they also have the generalization over those items stored in a schematic representation. The ability of a speaker to use items that are instantiations of a pattern is not proof of the psycholinguistic reality of the underlying generalization for that speaker. The conclusion we must draw for linguistic research is that the simple observation of instantiations is an insufficient basis for answering the question whether particular rules or generalizations are really part of speakers' constructicons.

In language acquisition research, accurate use of a pattern (or 'rule') is often taken as evidence that this pattern has been acquired. There has been some discussion about how many instances or what percentage of instances must be correct before a pattern can be argued to have been acquired (e.g. Brown, 1973), but in essence any convention (five different types, 90% correct etc.) is arbitrary, and the underlying question about knowledge of the generalization can never be answered if researchers restrict themselves to natural production: any sequence that is produced may have been retrieved as a stored unit.

One way out of this limitation is the use of pseudo words in experimental research. For the investigation of the acquisition of (mainly inflectional) morphology, Berko (1958) famously introduced the *wug*-test. She showed children a picture of a creature and told them that this was a *wug*. On the next picture, there were two of these little animals. The participating children were then asked what they saw. Children responding *wugs*, it was argued, showed that they knew the rule for the English plural: they could only have arrived at this form by using the schematic representation and combining this with the newly learned item *wug*.

This innovation in research design was certainly important, but Berko's results also pose some problems, as was already mentioned in the first chapter (Section 1.1.5): children performed much better with real words than with pseudo words. It is not easy to accommodate this with her interpretation that children "*operate with clearly delimited morphological rules*" (Berko 1958:269). At the very least, this means that the children she tested did not employ these rules for all words in the task. Even if they have stored the generalizations that underlie the instantiations, somehow that level of representation was not activated sufficiently for all test items.

A second important point to be made about Berko's -and others'- pseudo word experiments is this: how do we know whether these tasks reflect what language users do in normal life? In contrast to adult speakers, young children encounter new, unknown lexical items on a regular basis. One could, therefore, argue that the use of pseudo words in experiments for children does not make the experiments any less natural. For young children, it has been suggested that spontaneous speech data show that they are conservative in their language production: a large majority of what they say can be traced back directly to what they have heard (Dąbrowska & Lieven, 2005). In experiments, this finding has been corroborated with regard to the type of argument structure with which children combine novel verbs (Brooks & Tomasello, 1999). This conservative approach can be very useful in delimiting the hypothesis space. If children have to assume all the time that novel combinations are possible, it becomes unfeasible to determine which combinations can actually be used. This may very well be a reason why the children in Berko's task performed less adult-like on pseudo words than on real words: they are reluctant to produce something they have never heard before. Even if it is normal for young children to hear new things, they tend to be reluctant in using these new items in any linguistic context other than the one they have heard them in. In other words: if they hear the new word *wug*, they are much more likely to use the word *wug* again than to form a plural with it.

Data from another experiment indicate that children as old as ten are still not as proficient as adults in using novel derivations. Smith and Nicoladis (2000) provided their participants with a pseudo word with a normal English derivational affix, e.g. *prastful*, *propanity*, and asked them to form a sentence with this word. The responses reflect whether participants were aware of the word class restrictions of the affix. The answers of (highly educated) adult participants showed clear knowledge of the functions of the suffixes and their semantic subtleties, but for children this was much less the case. As with Berko's experiment, however, we may ask ourselves whether the ability to use a novel word correctly in a sentence based on its morphological structure is a skill that a language user employs very often in everyday speech.

In sum, measuring knowledge concerning CLIs is complicated, especially if we want to find out whether speakers have abstract representations, i.e. knowledge of generalizations that underlie instantiations. The remainder of this first part of the chapter is devoted to a discussion of different kinds of evidence -spontaneous and elicited language use in Section 2.1.1, and online and offline tasks in 2.1.3- and to the distinction between implicit and explicit knowledge (Section 2.1.2).

2.1.1 Natural and elicited data

An obvious source of information about linguistic competence of speakers is their language use: if speakers use a particular CLI, this is clear evidence that they have some representation of this sequence and/or the pattern that it instantiates. With regard to research questions concerning knowledge about CLIs, there are two main drawbacks of spontaneous speech corpora. The first of these has to do with frequency: if the linguistic phenomenon that you are interested in is not particularly frequent, you might need a very large corpus of spontaneous speech before you have an adequate number of instances. The advent of electronically available corpora has, of course, facilitated access to large amounts of data, but these are still mainly based in written language, and if they are not (e.g. the Corpus of Spoken Dutch, CGN), they may still not be of sufficient size for some phenomena. CGN contains around ten million tokens, and an in-depth study of two constructions showed that this is probably too little to find attestations of fixed but infrequent instances of a pattern (cf. Chapter 3). If you want to look at the development of knowledge, i.e. language acquisition, you would need samples that are much denser than are usually collected. Tomasello and Stahl (2004) calculated that the usual samples of one or two hours a week are insufficient for reliable observations about acquisition order and development for all but the most frequent constructions.

An alternative to high-density sampling in language acquisition research are so-called diary studies, where the caretakers of a child write down any new forms they notice. Nicoladis (2005), who studied the acquisition of complex deverbal words, is a prototypical example of such studies: the author looked at the language her own son produced and "*noted any relevant forms he produced in her presence as well as the apparent meaning, whenever practical*" (Nicoladis 2005:426). Although this poses its problems (absence of sampling when the mother is not present, the subjective notion of *relevant* forms etc.), with this method a lot of data can be gathered. In the case of Nicoladis' study, the diary is claimed to represent at least 5500 waking hours. Her son's bilingual English-French language development data suggest that "*children may be using a variety of cues from lexical items in the input that may play a role in development of the ability to coin novel lexical constructions. These cues may include the frequency of lexical constructions, the meaning of the bound morpheme, and the order of the elements in a variety of lexical items*" (Nicoladis 2005:441).

Aside from the frequency and sample size problem, there is also the issue of interpreting attestations: what does it mean that a speaker uses a particular CLI? When someone uses a specific CLI, it is impossible to determine whether they have this sequence stored as a unit in their constructicon or have used an abstract representation of the pattern and combined this with a lexical item. An example will clarify this distinction: if someone states 'I am very unhappy with the results' the word *unhappy* could simply be a unit in their inventory, but it could also be assembled on the spot by combining the UN+ADJ construction and HAPPY. In other words: the occurrence of a sequence that conforms to a general pattern (e.g. *unhappy* is an instantiation of the UN+ADJ construction) does not tell us whether the person that used this sequence has an abstract representation of that pattern.

These two limitations do not mean that spontaneous speech corpora should be discarded as a valuable source of linguistic data. In fact, such corpora are very informative about the distribution of types over patterns, and recent developments in methodological tools have made it possible to extract more of this information (e.g. in collostructional analyses, Stefanowitsch & Gries, 2003), to the point where they even provide evidence about "*significantly absent structures*", i.e. the non-occurrence of sequences (Stefanowitsch 2006:62). Explicitly marked creations of a new instantiation are, admittedly, rare, but they are also irrefutable proof of productive use of a pattern, as in "*It's a brilliant start, but it'll be brilliant if it were slightly more customisable* (*if that's a word*)"[3]. In sum, spontaneous speech corpora are an important tool for linguistic research, but they are insufficient to determine if speakers have abstract representations of patterns, and very large corpora may be necessary for the study of infrequent CLIs.

If the aim is to find out something about speakers' knowledge regarding a linguistic phenomenon, designing and executing experiments offers a lot of advantages over the study of spontaneous speech corpora. It is possible to contrast the influence of a specific variable, or to control for variation caused by one. One can also make sure that there are enough data points to perform statistical analyses over them. Doing experimental research also provides you with the opportunity to select what language users you want to do the experiment with and to find out more about them.

In analyzing the responses participants give, however, there is one important question one needs to keep in mind: to what extent does this experimental task measure what a language user does in real life? That this is a difficult issue, is illustrated by Krott, Gagné and Nicoladis (2009), who investigated children's ability to interpret novel compounds (e.g. *birthday room*). They opted to present their test items in isolation "*in order to study whether children can arrive at a meaning without contextual clues*" (Krott et al. 2009:89, footnote 2). Although earlier research has shown that context helps to identify the most likely relation between the elements in a compound, of which there are many, e.g. PART_OF, MADE_OF, HAS, etc., they explicitly

[3] https://addons.mozilla.org/nl/firefox/reviews/display/4258?nor=true&show=10&page=31, accessed on June 12, 2009.

chose to investigate what relations are the likely set without context. Even if this is a valid research question, it is debatable whether this study looks at natural language processing if you define 'natural' as 'in everyday life'. Obviously, providing a context to each compound you want to test means that you are adding a lot of information which is different for each test item. If children then vary in their responses on different items, it is no longer possible to determine the cause of these differences, since these may lay in either the context or the test item itself. Krott and her colleagues found that the relations mentioned by the young monolingual English participants (aged 4;9 - 5;8) pointed at a bottom-up acquisition of compounds. Children start by learning individual compounds as wholes, then develop partially abstract patterns with one specific constituent, and eventually acquire an abstract pattern. Interestingly, the authors observed that the five-year olds that participated in this experiment seemed to use more item-specific knowledge (especially right-constituent families, i.e. items that had the same second constituent, like *summer house* and *town house*) than the adults. Adults showed more sensitivity to overall relation frequency in their definitions of the compounds: their answers reflected how frequent different types of relation such as PART_OF and HAS are. According to Krott et al., this is caused by the fact that there are many different possible relations between the two elements in a compound, none of which has a clear general preference. Another interpretation, more in line with the MultiRep model (Chapter 1), is that whereas children rely more on (partially) specific representations, adults' responses reflect the use of an abstract schema.

Krott's study is explorative in nature: there are no 'good' or 'bad' responses. The fact that differences were found between adults' and children's responses, however, can either be caused by a development in linguistic competence, or be attributed to a development in task strategy. Perhaps the difference between the participating children and adults lies not so much in linguistic knowledge, but in test-taking abilities and experience. The closer the set-up of an experiment is to everyday-life language use, the more probable it becomes that you actually observe 'natural' linguistic behaviour. Such a set-up might come at a cost: adding a context to the novel compounds presented in the experiment just discussed would have made it difficult to tease apart the effect of that context from the actual compound itself.

Experimental tasks using pseudo words are by definition unnatural: they contain elements that are not part of the language. Using pseudo words means that you are certain to avoid participants' familiarity with part of your experimental input, however. Smith and Nicoladis (2000) and Windsor (1994) utilized this in their experiments, in which children and adults were given pseudo words with real affixes and were asked to provide a sentence context for these items. Whether the observed proficiency can be interpreted as evidence that the participants draw on the cues that suffixes provide about word class in natural language use, however, is questionable.

Table 2.1: Value and limitations of different sources of linguistic data

	Behaviour	Evidence for	But not for
Spontaneous language use	Correct[4] use of instantiations	Knowledge representation of instantiations	Representation of the underlying generalization
	Novel forms, innovations	Productive use of schematic representation	Use of schematic representation in non-novel forms
Elicited language use	Correct responses on real word tasks	Knowledge representation of instantiations	Representation of the underlying generalization
	Correct responses on pseudo word tasks	Ability to employ strategy for correct responses	Use of schematic representations in real-life language speech

Both experimental tasks and spontaneous speech data have their limitations, as is summarized in Table 2.1 above. It is for this reason, that we have opted for an approach that combines different experimental tasks, based on distributional information from corpora. By comparing and contrasting linguistic behaviour in different situations, we hope to achieve more insight into the knowledge that people have about CLIs and the levels of representation they use.

2.1.2 Implicit and explicit knowledge

Until now, we have discussed knowledge of CLIs in terms of levels of representation that differ in terms of specificity: there is storage of specific sequences (*unhappy*) alongside representations of patterns with

[4] 'Correct' use refers to use that conforms to adult native speakers' conventions.

underspecified elements (UN-ADJ). In the preceding sections, we argued that speakers use different levels of representation depending on the demands they face, and that evidence for specific levels of representation (i.e. correct use of specific instantiations) does not entail that a schema has also been acquired. It is necessary to complicate the issue still further and introduce the notions of *explicit* and *implicit knowledge*.

The difference between explicit and implicit knowledge has been defined clearly by Hulstijn (2005) who says that *"[e]xplicit and implicit knowledge differ in the extent to which one has or has not (respectively) an awareness of the regularities underlying the information one has knowledge of, and to what extent one can or cannot verbalize these regularities"* (Hulstijn 2005:130). Many traditional experimental tasks in linguistics, examples of which are the word formation task and the definition task reported below, require explicit knowledge. If someone were to perform poorly on such a task, it would be premature to interpret this performance as absence of knowledge per se. It might be the case that there is a representation, yet not available as explicit knowledge. By allowing for knowledge to be present, but unavailable, we can interpret variation in performance. Taking the results of Berko's wug-test as an example, the fact that the participating children produced some but not all correct plural forms for novel nouns can be interpreted as a lack of explicit knowledge. If the participating children had had explicit knowledge about plural formation, like adult native speakers, their performance would have been at ceiling, with all forms correct. This was not the case, which means that Berko's interpretation of the results must be modified: children's responses to the test items show that they have some knowledge of the patterns that are tested, but also reveal that they do not have explicit knowledge available that they can use on all test items.

Experimental tasks can be designed to make it likely that participants will make use of either their explicit or implicit knowledge. R. Ellis (2006) suggests that *"to investigate (...) implicit knowledge it will be necessary to employ tasks that tap into what learners intuitively feel to be correct, that are time-pressured, that call for a primary focus on meaning and that make no call on the learners' metalinguistic knowledge. In contrast, to investigate learning difficulty in terms of explicit knowledge, the tasks employed should encourage learners to respond using 'rules', should be performed without any time pressure, should call for a primary focus on form, and should invite the use of metalinguistic knowledge."* (R.

Ellis 2006:434).[5] In setting up the experiments described below, we tried to operationalize these differences such that the various tasks elicited both types of knowledge. It is important to note here that characteristics of tasks in which people mainly use implicit knowledge are more like natural language use: in everyday speech, we are usually under time pressure and our primary focus is on communication. Information exchange and intention reading rather than correct use are our main concerns in conducting a conversation.

Metalinguistic knowledge is our explicit knowledge about how language is structured. Explicit attention to language in schools fosters a growing metalinguistic awareness. Grammar instruction provides children with labels for categories ('noun', 'verb') which they might have been able to group items into before. Some abilities related to language, such as defining words and longer items in terms of semantic categories and distinctive characteristics (an X is a Y that Z's) may be at least partially 'school-induced' (cf. Snow et al., 1991; Kurland & Snow, 1997). Metalinguistic awareness and the ability to define words have been found to be correlated for both children and adults (Benelli et al., 2006). For CLIs it has been suggested that there is such a thing as *morphological awareness*. Learning to read and write is related to awareness of morphemes: children who do well on morphological skills or awareness tasks have a larger vocabulary (McBride-Chang et al. 2005) and outperform their peers in reading (Carlisle & Nomanbhoy, 1993; Singson, Mahony & Mann, 2000; Nagy et al., 2003; Deacon & Kirby, 2004).

Experimental tasks differ in the extent to which they require metalinguistic knowledge to execute them, but also in whether it is necessary to put this knowledge into words while doing the task. A definition task (cf. experiment 2 below), for example, consists of verbalizing this knowledge, whereas an acceptability task (cf. Chapter 3) does not. Deciding how acceptable an expression is, is a metalinguistic job, but since you do not need to express your considerations in words, it might be less taxing than a definition task. This difference, we argue below, explains some of the apparently contrasting results of the experiments reported on in this chapter. Before turning to a description of these experiments, it is important to discuss the essential difference between *online* and *offline* experiments.

[5] The related issue of explicit and implicit learning is not discussed here. Different authors have argued either that explicit teaching cannot lead to implicit knowledge (most notably Krashen's distinction between learning and acquisition, Krashen ,1982) or that it can (e.g. N. Ellis, 2005).

2.1.3 Online and offline tasks

Online experimental tasks measure language processing. They are aimed at the collection of data of real-life language use as it happens, but within an (often highly) controlled environment. Measurements tend to focus on speed rather than 'correctness'. An example of an online experiment is a self-paced reading task. In this type of task, participants are asked to read sentences about which they usually have to answer questions (to make sure that they try to understand the content of the sentence). While they read, their eye-movements are traced with eye-tracking devices. The eye-tracking data provide information about the units people use when reading (saccades), elements that are difficult (longer reading times, regression of fixation) etc. Lexical decision tasks (cf. experiment 3) are also viewed as online tasks, although it is important to recognize that the task itself (deciding whether a string of letters constitutes an existing word) is not natural in the sense that it is not something that speakers do in natural language processing (cf. Gilquin & Gries, 2009, who rank linguistic data in terms of naturalness of production and collection, on which a lexical decision task falls rather close to the 'unnatural' pole). Recalling R. Ellis' characterization of tasks that tap into either explicit or implicit knowledge, it is clear that online tasks mainly measure implicit knowledge: these are time-pressured, focused on meaning and do not require metalinguistic knowledge.

Most traditional tasks for measuring language proficiency are offline tasks. They allow participants to 'step outside' the language processing and reflect on their responses. This means that they can use the generalizations or rules that they have available to them as explicit knowledge in order to arrive at correct responses. Compared to online tasks, the focus is more on form and metalinguistic knowledge aids in performing well on offline tasks. In executing an offline task, participants may make up generalizations on the spot. In case of offline tasks about CLIs such as the definition task described below, such *ad hoc* "*morphological problem solving*" (Anglin 1993:129) can lead to correct responses, but this may well be a sign of good task-related skills rather than the use of schematic representations.

In sum, participants mainly use implicit knowledge for online tasks and can employ explicit knowledge in their performance on offline tasks. Although this is not a one-to-one match between task and knowledge type, this means that it is possible to trace back differences in performance between online and offline tasks to differences between implicit and explicit knowledge. The three experiments described in the remainder of this chapter show that the distinctions made so far -between spontaneous and elicited data, online and offline tasks, and implicit and explicit knowledge- provide a valuable basis for the analysis of variation in performance.

2.2 Word formation task

The first experiment has a long tradition in language acquisition research, going back to Berko's (1958) seminal work. Using a range of elicitation techniques, a number of researchers have tried to investigate children's proficiency with regard to morphological affixes by presenting them with a monomorphemic word, of which they then had them form a derived or inflected form. Carlisle (1988) provided her participants (monolingual fourth- to sixth-graders) with a word, followed by a sentence with a gap. The gap had to be filled, using a single word derivation from the word given at the beginning (e.g. *farm. My uncle is a ... (farmer)*). She found a clear developmental pattern in children's performance, with older children outperforming younger ones. There is evidence that this development starts before fourth grade: Deacon and Bryant (2005), who did a very similar experiment, observed the same pattern for second to fourth grade pupils. Development continues at least until eighth grade. Windsor (1994) tested older children's ability to form derived words and saw children's performance approaching that of adults only by grade eight. Nagy, Diakidoy and Anderson (1993), who used a similar design but with multiple choice format, found that even among high-school pupils *"most students did not do as well on derivative items as on stem items, indicating that knowledge of even common English suffixes was not complete"* (Nagy et al. 1993:168).

Various researchers have used this experimental design with pseudo words. Lewis and Windsor (1996) did this by first introducing these pseudo words with pictures in a context, thus providing a referent and clear indication of its word class (e.g. *this is a FID*). The children that participated in their experiment (fourth- to eighth-graders) were able to correctly form some novel derivations. They did not find a developmental pattern, possibly because of their small sample size (20 children in total). Nunes, Bryant and Bindman (1997) also used pseudo words and investigated whether children showed morphological awareness in their spelling of the derived word. They observed a developmental pattern, which they formalized in a Stage Model that includes progressively more awareness of generalizations as well as irregularities and exceptions.

All of these experiments require participants to produce the morphologically complex form. Derwing (1976) and Mahony, Singson and Mann (2000), among others, asked children whether a derived form (e.g. *teacher*) 'comes from' a monomorphemic form (*teach*). When the same children who did the production task were also confronted with a recognition task (e.g. Lewis & Windsor, 1996; Carlisle, 1988), they performed better on the recognition task. In discussing the results of her experiments, Carlisle suggests that *"[a]wareness of the morphological relatedness of words and the ability to analyze morphemic structure may depend on combined features of phonological and semantic similarities and associations, on linguistic sophistication, and even on the specific characteristics of the language tasks used to assess this ability."* (Carlisle 1988:249). This points to the fact that participants in this task make use of their explicit knowledge. The fact that this is an offline task allows them to reflect on the correct form.

With the present experiment, we try to answer the question to what extent children are able to manipulate words morphologically in such a way that it fits a given sentence context. This experimental, offline task measures explicit knowledge.

2.2.1 Participants

Sixty-nine fourth-graders (mean age 9;4) from three different primary schools in Amsterdam participated in the experiment. In each school, all the children in fourth grade ('groep 6' in the Dutch school system) took part on a voluntary basis. The schools were chosen to reflect the demographic variety in the present school population in the Netherlands, especially in the larger cities. In a questionnaire, more than half of the participating children reported using another language than Dutch on a regular basis in the home environment (N = 38). There is not one main minority language: different varieties of Arabic, with 14 speakers, were mentioned most often, but 18 other languages were listed. Because of this diversity in other language or languages spoken at home (four children said they spoke two other languages), all children who said they regularly used another language than Dutch were grouped as bilinguals, and the remaining 32 participants as monolinguals. All children were born in the Netherlands and had attended school there ever since they were of school age. Table 2.2 illustrates the distribution of the pupils over the three schools.

Table 2.2: Participants word formation task per school

School	Nr of children	Monolingual	Bilingual	Mean age
1	27	25	2	9;2
2	21	4	17	9;6
3	22	3	19	9;5
Total	70	32	38	9;4

2.2.2 Test items

The test consisted of 14 sentences with 15 affixes in total, i.e. one test item combined two affix constructions (ON– and –BAAR). Five different frequent affix constructions were part of three test items each: the prefix ON– and suffixes –EREN, –ER, –HEID and –BAAR. Table 2.3 contains an example for each affix, also listing an English near-equivalent.

Table 2.3: Affixes in the word formation task

Affix	Stem	Derivation	English equivalent
on-	Aardig (kind)	Onaardig (unkind)	un-
-eren	Fantasie (fantasy)	Fantaseren (fantasize)	-ize
-er	Inbreken (break in)	Inbreker (burglar)	-er (ag. noun)
-heid	Duidelijk (clear)	Duidelijkheid (clarity)	-ity / -ness
-baar	Lezen (to read)	Leesbaar (legible)	-able

For each test item, the stem occurred at least five times in the Schrooten and Vermeer corpus (Schrooten & Vermeer 1994), which indicates that these are words that children are likely to have encountered in school and will therefore be familiar to them.

2.2.3 Procedure

The format of the task is a basic fill-the-gap procedure: the test items consist of a single monomorphemic word, followed by a sentence with a gap in it. The participants had to fill in the gap, using the word at the beginning (see

Figure 2.1 for an example and Appendix 1 for a complete list of all test items).

Fantasie Ze zegt dat haar vader drie auto's heeft, maar volgens mij zit ze maar een beetje te .... (fantaseren). Fantasy She says that her father three cars has, but according_to me sits she but a little to (fantasize) "She says that her father owns three cars, but I think she's just making that up"

Figure 2.1: Example test item word formation task

Prior to the task, the researcher introduced herself and the project in each class, explaining that she wanted to find out what kinds of words are difficult for children. This formulation was chosen so that it seemed like the test items where evaluated, rather than the children, which meant that there was no such thing as poor performance on the tasks. The task started with a short oral introduction, after which all children did the test individually. Using two examples with different affixes, it was explained that they had to change the word at the beginning of each sentence so that it fit in the gap. To make sure that the children needed to look at the stem they were given when completing the task items, a small pilot study had been done, in which the same sentences were given without the monomorphemic word. For all test items, participants came up with other possible words to fill in the gap, e.g. in the *fantaseren* example these included *liegen* 'lie' or *kletsen* 'chat'. None of the participants in the pilot study took part in the main task.

The children were allowed to take as much time as they needed; the task took approximately ten minutes. To reduce test anxiety, they were told that they would not be graded for their work.

2.2.4 Variables

Three factors that might cause variation in performance between participants were taken into account: the *school* they attend, speaking more than one *language*, and their *vocabulary size*. With regard to the test items, differences can be due to the *affix* in the complex word, the *frequency* of the *derivation* and the *frequency* of the *stem*.

Individual variables

The participating schools differ in various respects: they are located in different parts of the city and attract children with a different socio-economic background (especially school 1 compared to schools 2 and 3), and they also adopt different teaching methods, do not all use the same text book and, of course, have different teachers. As it is impossible to control for all of these variables, it was decided to take these as a given, but it was verified that none of the teachers had paid explicit attention to derivational affixes in the week prior to the experiment.

The language background of the participating children was very diverse. All children filled out a short questionnaire with questions about their language use: they were asked to rate on a 1 to 5 scale how often they used another language than Dutch with their parents, their siblings or their friends. Nearly all children who reported ever using another language (N = 38) mostly did so with their parents (mean score = 3.26) and to a lesser extent with their siblings (mean score = 2.35) but hardly ever with their friends (mean score = 1.21), which might be due to the multilingual and –cultural environment these children grow up in: many of their friends will speak other languages at home. The picture that emerged from the questionnaire for bilingual children was that for most of them the distribution of languages was the same: they often talk to their parents in the other language and with their siblings in either Dutch or the other language. Because of these similarities, all children who reported speaking more than one language were classified as bilinguals.[6]

A third individual variable that may well influence performance on this task is the participants' vocabulary size (cf. other experiments, e.g. Freyd & Baron, 1982; White, Power & White, 1989). A large vocabulary can aid

[6] We are aware that this is an oversimplification. Within the group of bilinguals, some children will have started to learn Dutch at the same time as the other language. For the current research project, it was not feasible to conduct an in-depth investigation of the children's language background. This would have to include measures of proficiency in their native language, something which was practically impossible because of the variety of languages spoken. We chose not to exclude bilingual children for two reasons. Firstly, if speaking two languages influences knowledge and/or awareness of morphological structure, this should be visible regardless of the specific languages spoken. Secondly, the sample of children who participated is representative of the current primary school population in Amsterdam. Figures of the Central Statistics Bureau (CBS) indicate that currently (2005-2009), 48% of all primary school children in the capital belong to a 'cultural minority'. These minorities include most migrant citizens, such as Turks and Moroccans, but also some largely monolingual Dutch groups such as Moluccans, while not counting those who come from West-European countries (e.g. Germany and France). Although it does not give us an exact match with all likely bilingual pupils, these figures indicate that bilingualism is very common among children in Amsterdam.

performance on this task in two ways. Firstly, children with a large vocabulary are more likely to have stored representations of the complex words that they need to fill in. Secondly, a child who knows more words in general is likely to also know more words with each particular affix. This in turn facilitates the development of generalizations over the different words with the same affix, resulting in a mental representation of a construction with that affix (as is argued in the description of the MultiRep model in Chapter 1). That construction may then be used to form the derivations in this task. Either way, by retrieving a stored representation of the derivation or by using the stored generalization, a large vocabulary will boost performance on this task. If children mainly rely on the stored specific lexical items, the frequency of these items should correlate strongly with performance scores. If it is the case that children use their stored generalizations, the frequency of the derivations will not be a strong predictor.

To assess children's vocabulary size, the receptive word meaning task in the TAK (Taaltoets Allochtone Kinderen Bovenbouw, Language Test for Migrant Children grades 4 - 6, Verhoeven & Vermeer, 1993) was administered. This is a 50-item multiple choice task, with each item containing a word in a sentence context, followed by four definitions of the word from which they must choose one. This test has been standardized for children in grades four through six and takes approximately half an hour to complete. The test items of the TAK include mono- and bimorphemic single words (e.g. *duf*, 'dull', *opvolgen*, 'succeed, follow up') and multi-word lexical items (*beneden peil*, 'below the mark', *in ongenade vallen*, 'fall into disgrace').

Within-participant variables

The test items varied on the *affix* they contained, the frequency of the *stem* and the *derivation*. Five different affix constructions occurred three times each. The frequency of the stems (the monomorphemic word that was given) and the derivations was determined using the Corpus of Spoken Dutch (Corpus Gesproken Nederlands, CGN). This is a 10 million-word corpus of contemporary Dutch, with different (spoken) genres, ranging from telephone conversations to speeches, and with samples from different Dutch-speaking regions, including Flanders (Belgium). Table 2.4 contains the frequency data for each stem and derivation, based on lemma searches (see Appendix 1 for translations of all test items). This means that different inflectional forms (plural for nouns, tense- and person-marked forms for verbs) are included. By varying the relative frequency of the stem and derivation, it is also possible to ascertain the influence of this factor. No pseudo words were included in this task. These items would constitute an extreme case of relative frequency, with the frequency of the stem being zero.

Table 2.4: Frequency data for test items word formation task

Item	Stem	Frequency stem	Frequency derivation
Kopiëren	Kopie	115	177
Inbreker	Inbreken	45	18
Fantaseren	Fantasie	68	23
(Dromen)vanger	Vangen	327	96
Adviseren	Advies	269	117
Onaardig	Aardig	1014	29
Leesbaar	Lezen	5539	20
Ongeduldig	Geduld(ig)	117	46
Verkoudheid	Verkouden	67	36
Duidelijkheid	Duidelijk	2319	135
Onhoorbaar	Horen	7494	16
Hoorbaar	Horen	7494	47
Bijter	Bijten	168	3
Hardheid	Hard	1817	7
Herhaalbaar	herhalen	361	1

2.2.5 Results word formation task

All responses were simply scored as either correct or incorrect. One item contained two suffixes. This was scored twice: once for the first, and once for the second affix. The task was reliable (Cronbachs α = .85).

Participant-based variables

The participants’ mean score on the 15-item task was 8.38 (sd 3.77). At first glance, it seems that the monolingual children (mean score 10.00, sd 3.42) outperformed their bilingual classmates (mean score 6.97, sd 3.52). The two histograms in Figure 2.2 below for the monolingual participants (grey bars) and the bilingual participants (black bars) point at the existence of two different groups in terms of performance on this task.

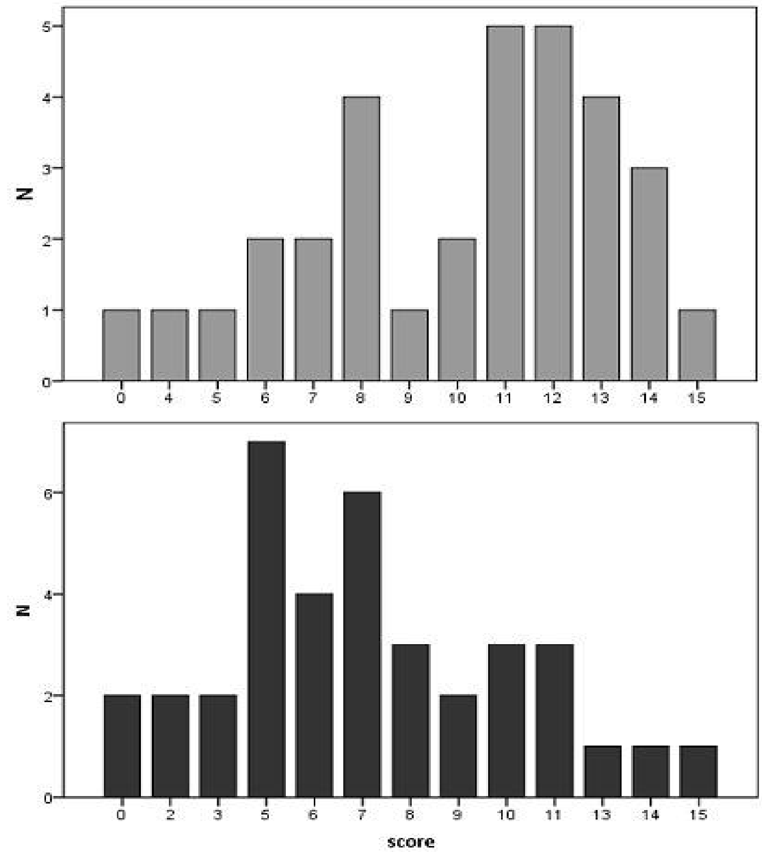


Figure 2.2: distribution of scores on word formation task for monolingual (grey bars) and bilingual participants (black bars)

The monolingual children also scored much higher than their bilingual peers on the vocabulary size task (TAK): their mean score on the 50-item TAK test was 34.00 (sd 7.66) while bilingual children scored 22.97 (sd 7.09) on average. As was outlined in the previous section of this chapter, there are reasons to assume that vocabulary size influences performance on this task. For that reason, an analysis of variance was done with vocabulary size as a co-variate. The apparent difference in scores on the word formation task turns out to be entirely attributable to the difference in vocabulary size between both groups. While the difference between the two groups in TAK scores is highly significant ($F_{(1,67)} = 39.55$, $p < .001$), the distinction between bilingual and monolingual children is not ($F_{(1,67)} = .00$, $p = .99$). The performance on both tasks (TAK and word formation) are strongly correlated (r = .70). Taken together, these figures suggest that performance on the word formation task is strongly influenced by vocabulary size. Bilingualism by and in itself does not affect performance (positively or negatively).

The third individual variable that was identified is the school each child attended. As for the language background of the participants, there are large differences in performance from school to school. Again, though, these differences are no longer significant when the effect of individual pupils' vocabulary size is taken into account. With the TAK results as a co-variate, the school variable is no longer significant ($F_{(1,66)}$ = .23, p = .63).

The strong correlation between scores on the word formation task and the vocabulary size measure makes it likely that knowing a lot of lexical items is beneficial to performance on the task. As was suggested above, a large vocabulary can help in two ways: it aids in developing abstract representations for affix constructions, and it means that you are more likely to know the words in the task. A closer look at the performance on each test item will assist in deciding which of these (or both) apply in this situation.

Item-based variables

Performance on the test items ranged from 11 to 65 children providing the correct answer per item (with 69 participants). If the participating children were able to always use abstract representations, correlations on test items with the same affix construction would have to be higher than the correlation between different items. This is not the case: correlations between items with different affixes (N = 89) are .257 (se .015) on average, and .333 (se .036) between items with the same affix (N = 15). This difference is not significant in a 2 samples Kolmogorov-Smirnov test (Z =.980, p = .29), selected because of the small sample size. An Anova-analysis also shows that scores on the five different affix constructions do not differ significantly ($F_{(4,10)}$ = 1.76, p = .21). This indicates either that children did not consistently rely on abstract representations, or that they all did for all five constructions. The latter explanation has to be ruled out, however, because scores were far from at ceiling.

The overall scores show that not all children consistently used an abstract representation. This does not exclude the possibility that some children always provide the correct answer, while others consistently fail to do so. A closer look at the distribution of scores over each affix construction serves to distinguish between this all-or-nothing hypothesis and a more gradual development. Each affix construction was tested with three items, meaning that each child scored between 0 and 3 correct on a construction. If children either do or do not know the rule, the majority of scores should either be 0 or 3. If development of the abstract representation is a gradual phenomenon, scores should be distributed more evenly. This last hypothesis is tested in a χ^2 test, contrasting the observed score distribution with the expected distribution, given the overall percentage of correct responses for that affix construction. Table 2.5 summarizes the expected and observed distributions. The expected distributions are calculated as the binomial probability given the total score for each affix construction. Figure 2.3 below contains bar plots reflecting the distribution of scores for the constructions.

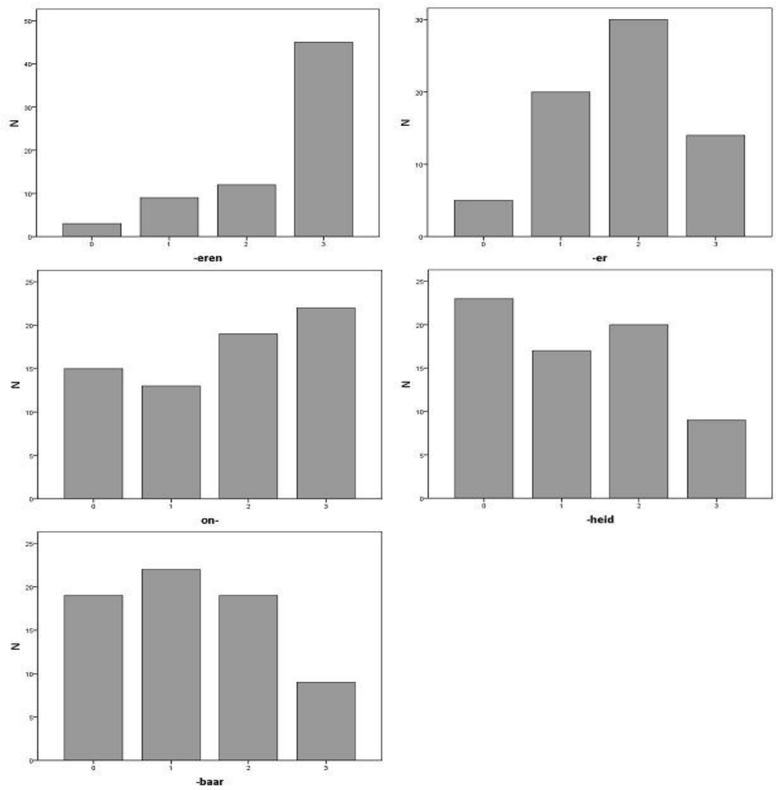


Figure 2.3: Distribution of children's scores for each affix construction. N = nr. of participants

Table 2.5: Summary of observed (O) distribution of scores and expected (E) distribution given p = percentage correct responses for that affix construction

	-eren (p = .81)		-er (p= .59)		on- (p= .57)		-baar (p= .42)		-heid (p= .41)	
	O	E	O	E	O	E	O	E	O	E
0	3	.7	5	4.9	23	14.5	15	5.5	19	13.8
1	9	6.2	20	20.9	17	29.7	13	22.1	22	29.0
2	12	25.	30	28.6	20	20.0	19	29.0	19	21.4
3	45	36.	14	14.6	9	4.8	22	12.4	9	4.8

The observed frequencies do not differ significantly from the expected values for the –ER and the –BAAR affix constructions. The differences are significant for the other three affix constructions. With regard to –EREN ($x^2_{(3)}$ = 18.10, p < .001) the observed frequencies are in line with a ceiling effect for a majority of the children. For the ON– and –HEID ($x^2_{(3)}$= 30.84 and 14.01, p < .001 and < .01 respectively), the difference between the observed and the expected frequencies is very similar: the observed number of 0 scores is larger than expected, and fewer children get full marks (score = 3) than expected. The all-or-nothing hypothesis cannot explain this distribution. It is, however in line with the expectation that people will use lexically specific representations whenever possible. Many children may have been able to solve at least one item with a particular affix in this way. Therefore, the distribution is more compatible with a view that sees the emergence of an abstract representation as a gradual development.

Looking at the relation between frequency of occurrence and item scores, the correlation between the frequency of the derivation and task performance is reasonably strong (r = .54). No such relation is found with the frequency of the stem (r = -.23). This does not mean that the stem frequency is irrelevant for item scores. The relative frequency of the items, calculated as a percentage (frequency derivation / stem frequency x 100) shows a stronger correlation with performance than the simple frequency of the derivation does (r = .65). Table 2.6 contains relative frequency data and the percentage of correct responses for each test item.

The strong correlation between relative frequency and scores on items means that performance was highest for those derived items that occur frequently compared to their stems. With regard to morphologically complex words, Hay and Baayen (2002) argue that especially those words that have a high relative frequency are likely candidates to be stored separately, and to not have any morphological structure in their mental representations. This ties in with Bybee's suggestion that highly frequent forms do not contribute to a pattern's productivity (Bybee, 1995). The fact that it is these words that participants perform best on, seems to indicate that they indeed do not use abstract representations in executing the task, but rather retrieve specific complex items from their constructicon. How this specific form is activated is a question these data cannot answer.

Table 2.6: Correct responses and relative frequency test items word formation task

Item	Stem	Affix	Correct responses (%)	Relative Frequency (%)
Kopiëren	Kopie	-eren	94	153.91
Inbreker	Inbreken	-er	80	40.00
Fantaseren	Fantasie	-eren	78	33.82
(Dromen)vanger	Vangen	-er	72	29.36
Adviseren	Advies	-eren	71	43.49
Onaardig	Aardig	on-	71	2.86
Leesbaar	Lezen	-baar	68	.36
Ongeduldig	Geduld(ig)	on-	56	39.32
Verkoudheid	Verkouden	-heid	54	53.73
Duidelijkheid	Duidelijk	-heid	49	5.82
Onhoorbaar	Horen	on-	42	.21
Hoorbaar	Horen	-baar	42	.63
Bijter	Bijten	-er	25	1.79
Hardheid	Hard	-heid	19	.39
Herhaalbaar	herhalen	-baar	16	.28

It can be argued that a correct response to at least one item with a particular affix suffices to indicate that a child is able to manipulate that affix. For that reason, a simple variable was created for each affix indicating for each participant whether they had provided minimally one correct answer (e.g. for the –ER construction, the response on at least on the the three items *inbreker* 'burglar', *vanger* 'catcher' and *bijter* 'biter' had to be correct). For two affix constructions, nearly all children were able to supply the correct answer at least once: the –EREN construction (66/69) and the –ER construction (64/69). For the other three affix constructions, the number is slightly lower: 54, 50 and 46 for the ON–, –BAAR and –HEID constructions respectively. This means that for each of the five constructions tested, at least two thirds of the participants were able to manipulate the affix successfully once or more.

The answer to the research question whether children are able to manipulate words morphologically in such a way that it fits a given sentence context has to be affirmative, but it is clearly not the case that they can do this for any instance of a given affix construction. Performance strongly depends on the relative frequency of the test items, suggesting that the explicit knowledge children have to draw upon mainly relates to specific representations of complex items.

2.3 Definition task

Like the word formation task, a definition task is a classic experiment that has been used many times to investigate children's linguistic proficiency, especially their vocabulary skills. With regard to morphologically complex words, a definition task can shed light on the question to what extent children make use of this morphological structure in coming up with a definition for the item. If they are aware of an affix construction, and are able to put its meaning into words, they can define morphologically complex words that are new to them. If they are not familiar with the affix construction, or perhaps unable to formulate its meaning, they may still realize that there is morphological structure, for instance because they recognize the stem. A large majority of complex words contains a stem that is more frequent than the derived form. Clearly, awareness of morphological structure is a condition *sine qua non*: if you are not given any context, it is impossible to fully define an unknown morphologically complex word if you do not appreciate that it contains meaningful elements.

Rubin (1988) made use of this in an experiment in which she showed children complex words and asked them whether they recognized something in that word (*is there a little word in ... that means something like ....?*). This test very much focuses on the *stem* in complex words. For spelling, research has shown that children do better on separate monomorphemic words than on that same letter string when it is the stem in a morphologically complex word (Kemp, 2006). This suggests that children are not able to 'transfer' their knowledge about the spelling of the word to the complex item, which in turn means that the complex item is stored separately at least at some level of representation.

Larsen and Nippold (2007) tested sixth-graders on their ability to define morphologically complex words. They found that there was a lot of individual variation in skills. The researchers provided 'dynamic assessment' in eliciting responses: they gave progressively more helpful prompts, including an explicit reference to the possibility that there might be 'smaller parts' in the word. Scores on each item were determined by the amount of prompts a child needed to give the answer. In my view, this interpretation has some problems: if a child only provides a definition after being told to look for smaller parts in the word, this does not necessarily mean that they could not have given an answer without that prompt. An absence of a response might be due to other issues, e.g. insecurity or lack of motivation.

In the current experiment, children were confronted with very complex words and asked to tell the researcher what they thought this word meant. Their responses provide an answer to the question whether, in the absence of any linguistic context (e.g. a story or a conversation), children are able to deduce the meaning of a novel complex word from their knowledge of its stem and the affix construction(s) it instantiates.

2.3.1 Participants

The same 69 fourth-graders that participated in the word formation task also did the definition task (see Section 2.2 for details). Due to recording problems, the responses of 22 participants were lost. Of the remaining 47 children, 29 are monolingual and 18 are bilingual.

2.3.2 Test items

The test consisted of 14 words, ranging from two to five morphemes (mean length 3.1) and nine to seventeen letters in length (mean length 12.1). Four out of the five affix constructions tested in the word formation task (the ON–, –ER, –HEID and –BAAR constructions) each occurred four times in the test items, with –EREN occurring three times. Five items instantiated two affix constructions. None of the test items occurred in the Schrooten and Vermeer corpus (Schrooten & Vermeer, 1994), indicating that these are not frequent words in the linguistic input children get. In the corpus of spoken Dutch (CGN), six of the words occurred at least once, but none more than six times. Because of the choice to use real rather than pseudo words, it was impossible to exclude the possibility that one or more words are known by some of the children tested.[7] All test items, glossed and with translations, are listed in Appendix 2.

2.3.3 Procedure

This task was administered individually, with the children and the researcher in a separate room from the rest of the class. To reduce test anxiety, the participants first filled out a short questionnaire about their language background. The researcher then explained that she had a number of 'big' and sometimes 'strange' words, that the child might or might not know. She wondered whether the child could figure out what the words might mean. She then presented the words, one by one, by saying the word out loud and placing a card on the table in front of them with the word printed on it in a large font (Times New Roman, size 48). The order in which the words were presented was randomized by shuffling the pack of cards for each child. With each word, the researcher asked "what do you think that X means" or a variant of that question, always choosing a formulation that would not reveal the word class of the word, e.g. in "what is an X". If children said that they didn't know or did not say anything, she prompted by asking whether there was a part of the word that the children recognized. If that too failed to elicit a response, the researcher moved on to the next word. She regularly gave compliments to the children ("that is a good example", "you're probably right", "well done") to encourage them.

The children's responses were recorded using a small tape recorder which was placed on the table in front of the researcher and the child. The researcher told the children that she would record the answers so that she would not forget anything the child said. None of the participants seemed inhibited by the recording apparatus. The task took approximately five minutes.

[7] A few responses by children indicated that they were in fact familiar with a specific word: one child replied to the word herintreder, 'returnee to the workforce' dat is mijn moeder "that's my mother", who had recently started a job after being a stay-at-home mother for a number of years.

2.3.4 Variables

As in the word formation task the *language* background and the *vocabulary size* of the participants were taken to be relevant individual variables. With regard to their language background, children were classified as either bilingual or monolingual. Their vocabulary size was assessed with the receptive word meaning task in the TAK-test (Verhoeven & Vermeer, 1993). See Section 2.2.4 above for more details.

Since this task tests whether children are able to identify and define certain affix constructions, the main item-based variable is which affix occurs in it (ON–, –EREN, –ER, –HEID and –BAAR). For those test items that combined more than one of the five affix constructions investigated, these were analyzed separately, i.e. for *adresseerbaar* (addressable), which contains both *–baar* and *–eren*, the responses were coded twice: once for the –BAAR construction and once for the –EREN construction.

Analysis

The children's responses were coded on a four point scale, based on the extent to which they showed the ability to identify and define the affix construction. Table 2.7 contains the coding categories and examples.

Table 2.7: Response coding definition task

Points	Response types	Example
0	None, unintelligible	*Weet ik niet*, 'I don't know'
1	Incorrect decomposition	*Maxi-meren, dat er te veel meren in een stad zijn* 'maxi-mize, that there are too many lakes in a city' (nb: the word *meren* literally translates as 'lakes')
	Sound associations	Onaangedaan-onaangenaam 'unmoved-unpleasant'
2	Naming or paraphrasing the stem	Smalheid*: smal weet ik wel, dat is iets dat heel dun is* 'narrow I know, that is something that is very thin'
3	Naming or paraphrasing the affix	Onaangedaan: *denk dat het betekent dat je iets niet hebt gedaan (...), dan staat er een 'on', betekent dat je het niet hebt gedaan.* 'think that it means that you haven't done something (...) then it says 'un', means that you have not done it'.

Note that in the coding procedure the emphasis was not so much on whether the child gave a correct definition of the word, but rather the amount of morphological awareness that was apparent in the response. To give another example, the following answer received full marks, although for adult native speakers of Dutch the test item in question means something else. In her definition, this child showed recognition of the stem and the semantics of the –BAAR construction:

> Wraakbaar: "bijvoorbeeld iemand heeft met jou een grap uitgehaald en dan wil je wraak nemen, en die grap is dan wraakbaar, die kan je ook terugdoen`'
> Objectionable (literally: revenge-able): 'for instance someone has pulled a prank on you and then you want to return the favor, and the joke is *revengable*, you can in fact do it back to them'. (bilingual nine-year old girl)

In addition to the score on a four-point scale, all responses were also coded simply for reference to the affix construction that was tested, i.e. only those responses that received a '3' on the scale were counted. This second score is a measure of the extent to which children showed the ability to provide a definition for the affix construction.

2.3.5 Results word definition task

Participant-based variables

With 19 test items scored on a 0 to 3 scale, the maximum score was 57. On average, children scored 37.85 (sd 6.95). This means that they got 66% of the maximum points. When only references to the affix construction were counted (maximum score = 19), the average is 6.34 (sd 3.44), or only 33% of the points. This shows that, whereas children were likely to give a response that showed some awareness of morphological structure, it was much less likely that this response included a reference to the affix construction.

Performance on the definition task correlated with the vocabulary size measure (r = .49). The difference between monolingual and bilingual children (mean scores 38.45 (7.11) and 36.89 (6.76) respectively) is not significant ($t_{(46)}$ = .74, p = .46), and differences between individual children are large. It was clear in administering the task that some children felt much more comfortable than others to give formulating a definition a try when they were unsure.

Item-based variables

There were considerable differences in performance on different affix constructions. Table 2.8 ranks them in order of the percentage of responses that contained an explicit reference to the affix (i.e. the '3' scores in the analysis).

Table 2.8: Percentage of responses on definition task containing reference to the affix.

Affix construction	Percentage responses with reference to affix
-er	65.3
on-	47.9
-eren	31.9
-baar	22.3
-heid	15.4
Total	33.4

These affix constructions vary on a number of points: one (ON–) is a prefix while the other four are suffixes. With two constructions, the resulting word is an adjective (ON– and –BAAR), one forms a verb (–EREN) and two are used to make a noun (–ER and –HEID). They also differ semantically, in the abstractness of their meaning. It is relatively easy to refer to the meaning of some constructions, such as the agentive –ER construction ("someone who..."). The rank order of the affix constructions, based on the responses, seems to correspond rather well with the degree of 'abstractness' of each construction. The –HEID construction (close in meaning to English *–ness*) forms abstract nouns, which are probably most difficult to define. Since only five different affix constructions were tested and these differ in various ways, we cannot conclusively state that the differences were caused by one of these variables. This is, however, a point that deserves further research.

In contrast to the word formation task, performance on the definition task was clearly influenced by the specific affix construction that was tested in each item. We argue, however, that this affix effect is no clear sign that the children who participated in this task all have abstract representations of these constructions in their constructicons. There are at least two reasons why this interpretation would be premature. Firstly, lexically referring to the meaning of the affix constructions is more difficult for some than for others: any mention of 'not' was analyzed as a reference to ON- for items with that affix, but for –BAAR children had to say that something 'could be done' or 'was possible'. Although this is probably related to differences in difficulty of the constructions, it is not the same. If a child is aware of both ON– and –BAAR he or she may still score more points on the ON– items because it is easier to formulate its meaning. In this respect, this task differs crucially from the word formation task: for that task, performing correctly on each item was equally difficult.

Secondly, if we postulate that some children do have an abstract representation for some affix constructions, which is especially likely for –ER and ON–, we are still faced with the problem of variation between items within the same participant: correlations on scores between different items did not differ significantly between items with the same or different affix constructions. Correlations between items with different affixes (N = 144) are .077 (se .018) on average, and .162 (se .033) between items with the same affix (N = 27). This difference is not significant in an independent samples *t*-test: $t_{(169)} = 1.91$, p = n.s. In other words, children who referred to the affix in one item with ON– did not necessarily do the same for another item with the same suffix. To account for this, we must assume that the abstract representation is not always available, or that the representation contains a smaller generalization: one that captures some items, but not others.

2.4 Lexical decision task: the *Family Size effect*

Both experiments described in the preceding sections tap into explicit knowledge that children have about complex words. There are other tasks, however, that allow us to investigate the role of morphological structure online: while lexical items are being processed. One such experiment is the lexical decision task, a format that was chosen for the present research question. Lexical decision tasks are widely employed, because they offer a number of advantages: it is a relatively quick way to gather a reasonable amount of data, the task is simple in design and allows for the manipulation of numerous aspects of words. Providing an overview of research using lexical decision tasks falls outside the scope of this chapter, but see Rastle (2007) for an overview.

In a lexical decision task, participants see a string of letters, for which they must decide as quickly as possible whether it represents an existing word. Both the response (yes or no) and the reaction time (in milliseconds) are recorded. The speed with which a word is recognized or rejected depends on a large number of factors. One obvious aspect is the length of words: the longer the string of letters, the longer the reaction time (hereafter RT). Similarly, RTs are influenced by word frequency: words that occur frequently are recognized faster (cf. Balota, 1990 for a discussion of the role frequency and meaning play in responses to lexical decision tasks). This observation indicates that frequent words have a higher baseline activation level; the threshold that is necessary to recognize a string as an existing word is reached more quickly for frequent words. Frequency may be seen as a reflection of entrenchment: words that are encountered often, will be more entrenched in memory. The measure for word frequency, which is, crucially, always based on a corpus, usually includes not only the frequency of the word that is given as a test item (e.g. *werk* 'work'), but also its inflectional variants (e.g. *werkje* 'work-DIM'). For experimental evidence that the lemma frequency is a better predictor of RTs in these experiments than the simple frequency of the form itself see Baayen, Dijkstra, and Schreuder (1997) and Schreuder and Baayen (1997). For English and Dutch, the inflectional forms include only a small set (for Dutch nouns, for instance, there are only the plural and diminutive forms). Research has shown that such a *lemma frequency measure* provides a better indicator for RT than the frequency of the test item alone. In many lexical decision tasks, test items are matched or controlled for the two variables mentioned here (length in letters and lemma frequency). Given the variation in RTs that is caused by these factors, another variable is manipulated to investigate its effect on recognition.

In the experiment we report on here, the main variable of interest is the *morphological family size*. This notion was introduced by Schreuder and Baayen (1997). The family size (hereafter FS) is the number of derivations and compounds a word occurs in. For Dutch, the word with the largest FS is *werk* 'work'. In Celex, a 40 millon word corpus of written Dutch, this word has more than 500 family members, ranging from *werkster* 'cleaning lady', literally: 'work-female agentive noun' to *huiswerk* 'homework' (Baayen, Piepenbrock & Gulikers, 1995). Baayen and Schreuder found that, *ceteris paribus*, words with a large FS are recognized faster than words with a small FS. Note that this is a *type* effect: RTs are not so much influenced by the frequency of the family members, but by the number of different family members.

Since family size was first recognized as a factor that influences recognition, the variable has been explored further. De Jong et al. (2000) replicated the finding that RTs are not influenced significantly by the frequency of family members, but by the number of derivations and compounds in which a word occurs. De Jong and her colleagues looked at the family members in somewhat more detail, by contrasting different counts for the family size measure using the Celex corpus. One obvious question is whether highly infrequent family members contribute to word recognition and therefore improve reaction times. On this point, De Jong reports that even removing the lowest frequency band of family members (with fewer than 10 occurrences in the 42 million word corpus) results in less variance explained by the FS measure. The type-based nature of this effect in combination with the observation that low-frequency family members do contribute to the effect indicate that the FS effect is not based in string familiarity in itself. Other experiments and more detailed analyses of the family size measure (Bertram, Baayen & Schreuder, 2000; Feldman & Pastizzo, 2003) have revealed that it does matter whether family members are semantically transparent: in an experiment with test items ending in the Dutch affix *–heid* (roughly comparable with English *–ness*), an FS effect was observed, but analyses showed that this effect seemed to be caused by semantically transparent family members only. This has some interesting consequences for models of the mental lexicon. For semantically opaque derivations and compounds, we must assume that they are stored as lexical items, to account for their interpretability: it is impossible to derive their exact meaning from their constituent parts. For semantically transparent derivations and compounds it could be hypothesized that they do not have a separate lexical entry in the mental lexicon, because they are decomposed (e.g. in the affix-stripping model of Taft & Forster, 1975). The finding that these words contribute to the recognition of the stem as an existing word indicates that they are in fact stored, with some kind of link to the stem in the word.

Over the past decade, family size effects have been investigated for languages with richer morphology than English and Dutch (e.g. Finnish, Lehtonen & Laine, 2003 and Hebrew, Feldman et al., 2004). For compounds, more detailed analyses have revealed that there are different effects for the frequency of both the left and the right constituent. Kuperman (2008) investigates this with a combination of corpus and experimental eye-tracking and lexical decision data. In all these experiments, the family members that are included in the FS measure are derivations and compounds. One follow-up study by De Jong, Schreuder and Baayen (2003) looked at the effect of adding a small context to test items. They found that adding the comparative suffix *–er* to monomorphemic adjectival test items had a strong effect: in comparison with a presentation of the adjective alone, the non-adjectival family members no longer contributed to the family size effect. When the modifier *heel* 'very' was added, only colour compounds and intensified adjectives (the scale-focusing adjectives) were relevant (De Jong, Schreuder & Baayen 2003:84). Clearly, this means that a disambiguating context reduces the number of different lexical items that are activated. It also raises the question why only complex words are counted as family members, and larger lexical units are not. In a view that recognizes the existence of multi-word conventional pairings and fixed expressions as items that are likely to be stored in speakers' constructicons, excluding these in the family size count seems an omission. This point will be taken up in the discussion section of this chapter.

Participants in lexical decision tasks are typically undergraduate students, who take part for course credit or a small reward. In the present experiment, the participants are children in grade 4 (mean age 9:4). Reports on lexical decision tasks done by children are scarce. Leong (1989) found that his fourth- to sixth-graders (mean age 9;11 for the youngest group) understood the task and were able to execute it. There were clear developmental patterns in de RTs for grade and reading levels. Similarly to adults, the children were slower to reject a non-word than to accept a word; yes-responses have a shorter RT than no-responses. This is not surprising: the decision to accept a letter string as a word can be made as soon as it is recognized. For a no-response, an absence of recognition is necessary. In these lexical decision tasks, all non-words are phonotactically legal. This means that they *could* be an existing word. This is done in order to make sure that a string is not rejected on the basis of its letter combinations alone.

Martens and De Jong (2006) also used a lexical decision task for their investigation of the effect of word length for dyslectic and normal reading children. In their research design, they departed from the usual set-up by asking participants to respond verbally instead of pressing one of two buttons, because they "*assumed that this would be easier for the (young) children than to remember which button to push*" (Martens & De Jong 2006:143). The ten-year-old Dutch dyslexics showed a much larger effect of word length in their RTs than their normal reading peers. For all groups, however, reaction times were considerably slower than is commonly observed for adults (mean reaction times for non-dyslexics was >1100 milliseconds where responses of 500-700 ms. are often found in experiments with young adults).

Bosman and De Groot (1996) asked even younger Dutch children, pupils in first grade (no mean age given) to do a lexical decision task. The task turned out to be comprehensible for children who had only recently begun to read: 94% of all words were correctly recognized. With this percentage so far above performance at chance level, it is clear that the participating children knew what was expected from them. Although the responses were similar to adults' in accuracy, again their RTs were a lot slower: the mean RT of correct yes-responses was 2177 ms., with words ranging from three to five letters. Rejecting control misspellings and pseudo-homophone misspellings took considerably longer (3886 and 3734 ms. respectively). Bosman and De Groot suggest that this is caused by the lack in reading experience these children have: at this point in life, they are likely to encounter new written words nearly every day. For that reason, seeing an orthographic pattern that they do not recognize does not trigger a strong rejection.

In sum, the lexical decision task has been used successfully in research on children's linguistic abilities. Experiments investigating the effects of morphological family size have focused on (monolingual) adult speakers. The existence of an FS effect has implications for our model of speakers' constructicons. If derivations and compounds contribute to the recognition of a monomorphemic word that is a stem in these complex words, then we must assume that there are links between the stored complex words and the monomorphemic word. The fact that this is a *type* and not a *token* effect seems to exclude the possibility that the complex words are not stored separately, but decomposed when they are encountered. In a decompositional model, such as Taft and Forster's affix stripping model (1975, see also Section 1.1.2) one would expect full frequency effects, i.e. a token effect.

In the present experiment, we test whether the reaction times by children (mean age 9;4) show a family size effect similar to that observed for adults. The lexical decision task is an online task, which directly taps in to the processing of words. It is a time-pressured task, and, especially in comparison with the word definition task discussed in the previous section, requires little metalinguistic awareness. Knowledge of the concept of 'words', which is a metalinguistic concept, is of course a prerequisite for good performance on this task. As such, it is likely to measure implicit knowledge participants have.

2.4.1 Participants

The same 69 fourth-graders that participated in the word formation task also did the definition task (see Section 2.2 for details). All children had normal or corrected to normal eyesight.

2.4.2 Test items

The nominal test items in De Jong et al. (2000) experiment 1 were the starting point for this experiment. In De Jong's design, twenty words with a large family size and twenty words with a small family size were selected using Celex (Baayen, Piepenbrock & Gulikers, 1995). The two groups of test items were matched for lemma frequency and length. The test items did not include long words (mean length 5.0 letters, with a range of 3 to 7 letters) and all items were monomorphemic. Celex is a corpus of written Dutch, composed of fiction and non-fiction books for adults. Since the input that children receive is rather different, we contrasted these counts with those from another corpus, *Woordwerken* (Schrooten & Vermeer, 1994). *Woordwerken*, with nearly 1.8 million tokens, is considerably smaller than the 42-million word corpus Celex. Its main attraction is that it consists of texts in children's books, textbooks used in primary schools and verbal interaction between teachers and pupils. In this way, it really reflects what children are likely to encounter in school.

All tokens in Celex are tagged for their family size, but for *Woordwerken* family size had to be counted manually. Initially, this was done for the forty original test items used by De Jong et al. (2000). On the basis of a comparison between the family size counts from each corpus, a number of test items had to be replaced. In each case, the new test item was of identical length and similar lemma frequency in Celex as the original word. For De Jong's test items, the correlation between the two family size counts, one based on Celex and the other on *Woordwerken*, was only .58 (Spearman's rho). The counts were recoded into Z-scores to make them directly comparable. There were five words for which the family size in the two corpora differed more than one Z-score. These words were replaced. One of the original test items, *lies* ('groin area') did not occur in *Woordwerken* at all, and was replaced for that reason. A few test items in the large FS size group had a smaller FS in *Woordwerken* than some of the words in the small FS size group. To avoid this overlap, four more items were replaced. In total, ten of the original forty items were discarded. The correlation between the two family size counts was now satisfactory (r = .87). The adaptations did not influence the equality of the two groups of test items in terms of lemma frequency or length. To these two groups of items, 40 phonotactically legal non-words were added, equal in length to the test items. In sum, the final test items were 20 words with a large family size, 20 words with a small family size and 40 non-words. The test items and their frequency counts are listed in Appendix 3.

2.4.3 Procedure

The experiment was made using E-Prime (Schneider, Eschmann, & Zuccolotto, 2002). A response box with two buttons was built especially for this task. The participating children executed this task individually, on a Dell Latitude D600 Laptop, in a quiet room. There were two versions of the experiment, differing only in one respect: whether a yes-response required pressing the left or right button. All children were asked which hand they used for writing, and the version of the experiment was selected accordingly, so that the yes-button always corresponded with the preferential hand.

The whole experiment took approximately five minutes. First, the children read a brief explanation on the computer screen. This introduction contained a simple description of the task. The children were told that they would have to determine as quickly and as accurately as they could whether the letters they were about to see in the middle of the screen formed a word or not. If they thought they saw a word, they had to press the button with the green sticker on it; the no-button had a red sticker on it. Before entering the main experiment, the participants did a short practice session (10 items). Only if they responded correctly to at least 7 of these could they proceed with the main session.

Each stimulus was preceded by an asterisk for 500 ms., both in the practice and in the main task. The stimulus disappeared from the screen as soon as the participant pressed one of the buttons or after three seconds, if none of the buttons had been pressed.

2.4.4 Variables

As with the previous experiments, there are two important individual variables: children’s language background (monolingual or bilingual) and their vocabulary size (see Section 2.2 for more details). With regard to the test items, we identified two length measures and two frequency variables: test items differed in number of letters and syllables, and in lemma frequency as well as family size. Our main interest in the present context is the effect of morphological family size. FS was included in the design as a dichotomous variable: test items had either a large or a small FS. This is, of course, a simplification, but we followed the research design as reported in De Jong et al. (2000). In order to be able to reliably observe the effects of FS, the other three item variables were matched between the two groups of test items. The mean length in letters (5.00 (sd 1.17) for large FS and 5.05 (1.15) for small FS) and syllables (1.55 (0.61) for large FS and 1.75 (.79) for small FS) were not significantly different for the two groups. Similarly, mean lemma frequency, which includes token counts of all inflectional variants of a word, was not significantly different for large FS and small FS test items (frequency per million words in Celex was 35.70 (27.65) and 38.10 (35.51) respectively).

Analysis

An initial item analysis showed that for eleven items performance was below 70%. These items included three non-words, seven small FS items and one large FS item, which were removed from all further analyses. The mean frequency in Celex for the two groups was still not significantly different (37,37 (27,36) per million words for large FS items and 39,85 (37,10) for small FS items).

Two children scored less than 70% correct on this task. Their responses were excluded from the analyses.[8] All incorrect responses and reaction times under 300 ms. were also discarded and RTs slower than two standard deviations above a participant's mean were not included, to reduce the effects of outliers.

2.4.5 Results lexical decision task

Table 2.9 below lists the mean reaction times for all remaining responses. The mean RT for non-words, i.e. a 'no' response, is much slower than that for

[8] This cut-off point is, of course, arbitrary. It seems that there is no clear convention in this respect. Percentages are reported ranging from 10% incorrect (Schreuder & Baayen, 1997, with adult participants) through 20% incorrect (Davis, Castles & Iakovidis, 1998, exp. 1, adult participants), 30% incorrect (Castles, Davis & Letcher, 1999, mean age participants 7;10) and up to 40% incorrect (Davis et al., 1998, exp. 4, mean age participants 9;8). With a 50% score indicating performance at chance level, the decision was made to exclude participants with less than 70% correct.

words. This is a common observation in lexical decision tasks (see above). An analysis of variance shows that the 33 ms. difference between small FS test items and large FS words is significant ($F_{(1,1968)} = 7.05$, $p < .01$). On average, words with a large family size are recognized faster than words with a small family size which are matched for lemma frequency and length.[9]

Table 2.9: Mean reaction times on correct responses lexical decision task

Stimulus type	N (correct responses)	Mean reaction time (ms.)	SD (ms.)
Small FS	779	1171	276
Large FS	1191	1138	262
Non-words	1924	1402	327

The children who participated were much slower than the adults taking part in De Jong's experiment (whose mean RT was 521 ms. (48) for small FS items and 502 (51) for large FS items). Moreover, the correlation between the mean reaction times per stimulus for both experiments is very low: $r = .12$, which is far from significant ($p = .59$). This indicates that the factors determining speed of recognition are not the same for adults and children. It has been suggested that children do not come across the same words frequently as adults do (Van der Werf, Hootsen & Vermeer, 2008). If that is really the case, this might explain the relative differences in recognition times for adults and children here: words that are frequent for children are possibly not very frequent for adults and vice versa. The *Woordwerken* corpus

[9] This experiment was originally reported on in Mos (2006). In recent years, more sophisticated statistical tools are becoming more common in analyses of these kinds of data (e.g. Kuperman, 2008; Baayen, Davidson & Bates ,2008). In these analyses, family size is not reduced to a dichotomous variable, but entered as the continuous variable it really is. Mixed-effect models also allow for the incorporation of multiple variable simultaneously. Unfortunately, using mixed-effects models with subject and item as random effects, the significant FS effect in our data does not hold up. The length of test items shows an inhibitory effect on RT (longer words equal slower RTs) and lemma frequency speeds up recognition. An effect of presentation order suggests a fatigue effect, but since all items were randomized for each participant, this need not affect other variables. Neither of the two family size measures (based on Celex and Woordwerken respectively) nor a Celex-based token count of family frequency are predictive of RTs. The absence of replication of the Anova-based significance is likely to be due to the fact that mixed-models are a more stringent technique: the possibility to consider multiple factors simultaneously obviously comes at a price. We are grateful to Victor Kuperman for his help with the analyses and interpretation thereof.

is too small to investigate this hypothesis in more detail, but this is clearly a point that would benefit from further research.

Performance on a separately measured vocabulary size task coincides with speed of recognition: children with a large vocabulary on average have shorter RTs ($F_{(1,66)}$ =10.91, p < .01). This means that children with high scores on the vocabulary size test are more adult-like in their reaction times, possibly because they know more words, or because they have more deeply entrenched lexical representations of items. In an input-based view of language acquisition, these two elements go hand in hand. As children get more input, they will encounter new words, increasing their vocabulary size, and they will observe words they already know again, causing deeper entrenchment of the existing lexical representation. This correlation seems indicative of a path of development. This observation means that knowing more words aids in the recognition of a stimulus, whereas the reverse could also be hypothesized: there are potentially more candidates that need to be checked (in the case of a 'no' response) or distracting items in the inventory. Clearly, the knowledge is structured in such a way that more in this case means less; a larger inventory equals less time needed to achieve recognition. The performance on this task cannot provide a complete picture, but it is plausible that entrenchment of lexical representations also plays an important role here.

Contrary to the effect on overall speed of recognition, there is no relation between vocabulary size and the FS effect ($F_{(1,66)}$=.55, p =.46). It is not the case that children with a large vocabulary showed a bigger (or smaller) effect of family size in their responses. This can be taken to indicate that the family size effect is there regardless of the number of words these children know. Obviously, since only children from one age group were tested, it is not possible to say whether the family size effect is present for children with a significantly smaller linguistic repertoire. Similarly to the experiments described above, the differences in RT between monolingual and bilingual children were not significant when effects of vocabulary size were controlled for ($F_{(1,65)}$ = .75, p = .39).

In sum, the FS effect as measured by De Jong et al. (2000) is replicated for children aged 9;4, although their mean RTs are considerably slower. The children take approximately twice as long as adults in responding, which may be tied to their smaller linguistic repertoire, encompassing both vocabulary size and level of entrenchment: performance on a separate vocabulary test was significantly related to speed of recognition. Replicating the effect found for adults indicates that the way in which knowledge of lexical items is structured is not essentially different for children this age.

The simple fact that the existence of morphologically complex items aids in recognizing the monomorphemic base means that there must be some sort of link between these items. Models of lexical processing that assume full decomposition of complex words (e.g. Taft & Forster, 1975) cannot accommodate why this should be a *type* effect. Full listing models (e.g. Butterworth, 1983), on the other hand, would have a hard time explaining the occurrence of the FS effect. For the FS effect to be available, the stem *werk* 'work' in a derivation like *werkster* 'cleaning lady' has to be identifiable to some degree: *werkster* can only speed up the recognition of *werk* if the two are linked. In other experiments, the effect of a morphologically related prime has reliably been distinguished from the effect of pure orthographic overlap (Diependaele, Sandra & Grainger, 2005). It seems, however, that shared aspects of meaning do play a role, as semantically transparent family members contribute more to the effect (Bertram, Baayen & Schreuder, 2000). For a language user, semantically opaque family members are really the same thing as words with no more than an orthographic overlap, although etymologically the words are related.

At this point, it is useful to point out that analyses of structure that a linguist makes need not always correspond to normal language users' implicit linguistic knowledge, which is what the MultiRep model that was introduced in the last section of Chapter 1 aims to reflect. This point will be taken up in the discussion chapter. In the last section of the current chapter, the results of the experiments described above are interpreted in light of the MultiRep model.

2.5 Summary of results and discussion of experimental results in the MultiRep model

Taking a usage-based approach to language acquisition, the MultiRep model that was introduced in Chapter 1 suggests that acquisition of a construction starts with learning some instantiations of the pattern. At first, children store these instantiations as simple form-meaning pairings without any internal structure. When children encounter various instantiations of the same pattern, with similarities in form and in meaning, this allows them to categorize these instantiations as belonging to one group. The similarities then form the basis on which (partially) abstract representations can develop. The question now is whether the knowledge as measured with the word formation task, the definition task and the lexical decision task is compatible with this model. Additionally, it is interesting to identify what level(s) of representation are accessed in each task, and how this is related to the notions discussed in the first section of Chapter 2: online vs. offline tasks, explicit vs. implicit knowledge and metalinguistic awareness.

Word formation task

In the first task, the word formation task, the participating children had to fill in a gap in a sentence with a word that contained a given stem. Performance on the task showed considerable variation, both at the level of participants and at the level of items. Children's scores on the task were strongly correlated with their linguistic repertoire as measured by a vocabulary size task (r = .70). The percentage of correct responses for each test item was correlated most strongly with the relative frequency of the derived word form (r = .65): derivations that are relatively frequent compared to the token count of their stem were most likely to be correct. With the mean score at 56% (8.38 out of 15 items), the results of this task show that the participants do not have full command over the skills that are necessary yet. On the other hand, for each of the five affix constructions that were tested at least two thirds of the children were able to provide the correct derivation for one or more of the items. Taken together, these numbers are indicative of an ongoing development. This development is gradual and is not compatible with a dichotomous view of morpheme acquisition: it is simply not the case that the children who participated were either able or unable to manipulate each affix construction. Even if they have an abstract representation, it is clearly not available with all test items. The development may therefore be apparent in the degree of availability of the abstract representation.

Within the MultiRep model, the gradual performance indicates that not all children can access the abstract level of representation through which derivations can be formed productively for all test items. At the same time, they do show the ability to form a derivation some of the time. In our view, there are three possible though not equally likely explanations for this type of performance, each of which is compatible with the model. First, children may not have any abstract representations at all, and come up with the correct answer because the monomorphemic word that is given and the correct response are both stored as lexical items and linked with each other. In Chapter 1 it was argued that this is the stored information on which basis abstract representations are formed. The fact that the relative frequency of the items is significantly correlated with performance makes this a likely explanation: frequent derivations will be stored as separate lexical items. For this task, this link is sufficient for successful performance (but not for other tasks, as will be argued below). The distribution of scores seems to indicate that children prefer to make use of specific representations when they perform this task (Dąbrowska, 2008 argues that this is a general preference). This preference is probably not a conscious choice.

Second, the extant abstract representation may not include the whole construction a linguist might identify, but rather a subset. This ties in well with the notion of conservative and bottom-up acquisition and development of abstract representations (e.g. Tomasello, 2003). If this were the case, however, it should be possible to define the subset(s) semantically, e.g. for the –ER construction performance would be nearly at ceiling for referents denoting a human being, but not for other agentive nouns. Unfortunately, the small number of items per affix construction (3) makes it impossible to determine if such sub-schemas exist: with the *–er* affix only one item occurred which referred to a human being. An additional hurdle is provided by the fact that, if the input-based nature of language acquisition is taken seriously, different children might develop different subschemas. The amount of data points is too small to perform individual analyses. A thorough investigation of this issue requires significantly more dense data collection.

A third possibility is that there is an abstract representation, but this is not used or available at all times during the task. If the participants do have general abstract representations for the affix constructions that were tested, it is definitely not the case that they are completely available for use in an explicit, offline task such as the word formation task, as they are for adults.

As this is an offline task, children have time to reflect on their reaction and can use their explicit knowledge. This does not mean their performance is an optimal reflection of their linguistic repertoire: the preference for a solution through specific representations does not always lead to a correct answer. The observation that a conscious effort does not necessarily result in good performance is found elsewhere too: if you consciously try to close and open an eyelid, this takes much longer than the normal blink of an eye. Similarly, if you explicitly consider the steps that are necessary for a gear change, chances of choking your car are higher than when you employ an automatized routine.

In sum, it is clear that the participants had some awareness of morphological structure and of the meaning of the affix constructions that were tested. It is also certain that they did not have abstract representations that they can use consciously and at will for all items. The correlation with overall vocabulary size indicates that knowledge about these constructions is still developing for the age group that was tested. For more insights into this development, longitudinal data would be necessary.

Word definition task

Performance on the second experiment, the definition task, presents a rather similar picture to that on the first task: here too, children demonstrated that they are able to perform well on some items, but not on all. Again, the notion of acquisition as an all-or-nothing switch must be rejected for the affix constructions that were tested. As with the previous task, performance was correlated with vocabulary size (r = .49). The number of correct responses for the different affix constructions varied considerably. Because only five constructions were tested, and these differed on a number of aspects (prefix vs. suffix, word class, frequency etc.) it is not possible to determine the cause of these differences. One notable point of difference is the formulation necessary to provide a correct response: whereas for the ON–construction (comparable with English *un–*) any mention of negation is sufficient to indicate that you understand its meaning, defining –HEID (*–ness*) requires more complex formulations.

For this task, in which words were presented without any linguistic context, participants could only rely on their explicit knowledge. Responding requires a high degree of metalinguistic awareness. Since the test items were chosen in such a way that children would almost certainly not have come across them, the only route to a correct response was by recognizing them as instantiations of an abstract pattern. Identifying morphemes and defining them involves the use of explicit knowledge. Given these high demands, correct responses are unmistakable evidence that a child has an abstract representation of a certain affix. What seems puzzling, then, is the observation that a correct response on one instantiation of an affix construction does not predict performance on other instantiations of the same pattern. This seems to be corroborating evidence for the second or third scenario that was discussed for the word formation task: the successful definitions show that children do have some abstract representation, but it is not available for all items, or it does not capture all possible instantiations.

For the definition task participants have to rely heavily on metalinguistic awareness: performance on it does not reflect normal language processing. In Gilquin and Gries' (2009) scale of naturalness, however, it is classified as a relatively natural task. Giving a definition of a word is something that people sometimes do, and in that sense it is natural in comparison with, for example, the lexical decision task that requires participants to decide if a string of letters is in fact a word, something that does not occur outside of experimental settings. The naturalness of a task, as defined by Gilquin and Gries, is unrelated to the online-offline distinction or the notions of explicit and implicit knowledge. Gilquin and Gries focus on what speakers may or may not do with language when they are not involved in experimental tasks, which may include giving definitions for words, although to the best of our knowledge it has not been investigated how frequent this linguistic act really is. An aspect that is clearly non-natural is the absence of context. Regardless of how (un)natural the task is, we take the position here that in the execution of this task, people have to employ their linguistic repertoire. As such it provides the researcher with valuable information about their skills, but not about how people go about using language outside the experimental setting.

In sum, the results of the word definition task show that children do have some abstract representations for the affix constructions that were tested. These abstract representations, however, either did not capture all instantiations or were not available for all items. The results of the experiment also do not provide any evidence about the application of these abstract representations in normal language use.

Lexical decision task

In contrast to the other two experiments, the lexical decision task was an online experiment, tapping in to the implicit knowledge participants have. Test items in this task either had a small morphological family size, i.e. the word occurs in only a few derivations and/or compounds, or a large family size. The main result was a replication of De Jong et al. (2000): participants were significantly faster in recognizing large FS items than small FS items. In the current experiment, the participants were children (mean age 9:4) while young adults took part in De Jong’s experiment. The family size effect was unrelated to vocabulary size, indicating that there are no qualitative differences between larger (adults, children with a large vocabulary size) and smaller (children with a small vocabulary) constructicons. These data therefore suggest that the way in which linguistic knowledge is structured does not change in fundamental ways as the repertoire develops, other than that it grows. There are large quantitative differences: adults are twice as fast as children in recognizing words, and children who score well on the vocabulary test are faster than children who perform poorly.

The family size effect can only be accounted for with a model in which words that are morphologically related are either linked to each other and/or in which the internal structure of complex words is part of the representation of these words. The MultiRep model assumes that both of these are the case: all lexical items are linked to other items on the basis of co-occurrence patterns, and form and meaning similarities. The internal structure of a complex item may also be part of the representation (see Figure 1.7 in Chapter 1). The effect can therefore easily be accounted for within the model. The model also has no problems accounting for the additional finding that semantically transparent family members contribute more to the FS effect than members that are etymologically related but synchronically are unrelated to the original stem in meaning. The assumption is that links are formed because of similarities in meaning and form. Semantically opaque morphological relations have a weaker connection: for a speaker, the only relation is form-based.

Conventionally, the only family members that are included in the family size counts are derivations and compounds. In Chapter 1 it has been argued extensively that larger units will also be stored, if they are frequent and/or have meaning aspects that are particular to the specific sequence rather than its constituting elements. It stands to reason, therefore, that these stored sequences also contribute to the family size effect, if they are either linked to the monomorphemic word or have an internal structure in their representation. *Werk* 'work', which already has the largest family size of all Dutch words, probably has quite a few more family members, such as *aan het werk zijn* 'working', literally: 'being on the work' *hard werken* 'work hard' and *goed werk leveren* 'doing a good job', literally: 'deliver good work'. In currently available counts, these CLIs are part of the token count for *werk*, but because they are fixed combinations, they must also be elements in Dutch speakers' contructicons. Identifying which CLIs to include in such a count remains a problem (cf. the discussion in Wray 2002) but that does not mean it should not be attempted. One helpful aid can come from co-occurrence data in corpora. These can be a starting point for conventions and quantifications of multi-word CLIs.

In sum, these experimental data are compatible with the model suggested in the previous chapter, but only if we assume that participants use different levels of representation for different tasks. It seems that the participants tend to access representations that are *as specific as possible* to solve the task. This may very well be a general strategy, motivated by a need to reduce processing costs. The MultiRep model presupposes massive storage (no full inheritance, co-existence of different levels of representation, storage of frequent regular forms, storage of frequent patterns, distributional patterns etc.). It may be that this is only workable if speakers try to make use of specific representations as much as they can.

In this chapter, the CLIs and patterns they instantiate that were investigated all belong to a specific subtype: morphological derivations. The definition given in Chapter 1, that they are *"strings of language in which more than one meaning-carrying element can be recognized yet which are very likely candidates to be stored as units in people's linguistic repertoires"* (p. 27) is not restricted to one-word CLIs. For that reason, Chapters 3 and 4 also discuss multi-word CLIs. Chapter 3 focuses on the notion of productivity, while the main question in Chapter 4 is when a multi-word CLI is processed as one unit.

Chapter 3:

Productive CLIs: A case study of the V-BAAR construction and the IS TE V construction [1]

3.0 Introduction

This chapter focuses on the issue of productivity: what is productivity, and how can we determine if a certain pattern is productive? We address this general question by looking at two constructions in detail. Productivity is of central concern in linguistics, we argue (based on, among others, Langacker, 2008; Boas, 2008:130; Barðdal, 2008; Wray, 2008). The overview given in Chapter 1 shows that morphological productivity has been studied from a theoretical linguistic background as well as in processing-related studies and in applied linguistics. The current chapter consists of a more detailed investigation than the experiments described in Chapter 2. In this chapter, two Dutch more or less synonymous constructions are analyzed. The first part of this chapter defines what a productive construction is, and discusses which characteristics of a construction influence its productivity. In derivational morphology, productivity is the term used to denote the degree to which a particular morpheme is used to form new words. Theoretical approaches such as Construction Grammar (e.g. Boas, 2003; Goldberg, 2006) remove any reason to limit this term to such a narrow domain. After all, grammatical constructions also differ from each other in the degree to which they are put to use by speakers in order to build new utterances, some being very 'productive', such as the Transitive Clause Construction (e.g. *I read Dostoyevski over Christmas*); others being much more limited in scope, e.g. the oft-discussed TIME AWAY construction (e.g. *I idled my Christmas break away*, cf. Jackendoff, 1997). It is argued here that constructional productivity should

[1] This chapter is an extended version of *Islands of (im)productivity in corpus data and acceptability judgments: Contrasting two potentiality constructions in Dutch*, a paper presented at the 2008 DGKL conference and submitted to be published in the forthcoming book *Converging Evidence in Cognitive Linguistics*, edited by Schönefeld. That chapter is co-authored by Ad Backus. We are grateful to the audiences at the 2006 Cognitive Linguistics Day in Leuven, Belgium, the 2007 Morfologiedagen in Amsterdam, the 2008 Formulaic Language Research Network Conference in Liverpool, UK, and the International Conference of the German Cognitive Linguistics Association in Leipzig, Germany, for their comments on parts of the research presented in this chapter.

be thought of as a reflection of speakers' abstract representations of that construction.

The second section of this chapter is a corpus study, with a distributional analysis that results in a description of the form and semantics of each construction. The productivity of the constructions is then tested experimentally in a *magnitude estimation* task (Section 3.3). In the final section, the value of corpus and of experimental data with regard to determining productivity is discussed, and outcomes are related to the MultiRep model that was introduced in Chapter 1.

In sum, this chapter introduces a definition of productivity based on speakers' representations; it discusses findings about two semantically similar but formally different constructions from a corpus analysis and from an acceptability task, and relates these findings to the MultiRep model. The research question can therefore be formulated as: how productive are the V-BAAR construction and the IS TE V construction for speakers of Dutch?

3.1 Definition of a productive construction

One of the essential characteristics of language is that it is both conventional and creative: speakers constantly recruit chunks they have stored in memory, in their linguistic repertoire, but they use them in novel ways and combine them in ever-changing combinations in order to deal with the conceptual and communicative demands of everyday life. This means that, depending on speakers' needs and on the available units and representations in their constructicon, they can sometimes use a stored instantiation of a pattern, but at other points combine a (partially) schematic representation with stored lexical items, e.g. UN+ADJ and HAPPY. This last process is what we call productive use of a construction.

Within the field of cognitive linguistics, recent definitions of productivity have focused on different aspects of this notion. Langacker (2008) says that productivity is subsumed by regularity. For him, productivity *"pertains to a schema's degree of accessibility for the sanction of new expressions"* (Langacker 2008:244). This is closely related to our own interpretation, albeit that Langacker formulates his definition without mentioning the speaker to whom this schema is accessible. In contrast to Langacker, Barðdal (2008) suggests that regularity is actually an aspect of productivity, along with generality and extensibility. In our view, generality is a separate, if not independent variable: an abstract representation may describe general or more specific productive processes. Barðdal's notion of extensibility is what we view as productivity here, again with the addition that we take the viewpoint of speakers. This leads to the following definition of a productive construction:

A construction is productive if speakers have in their constructicon a (partially) abstract representation or template, with at least one element that is not lexically specific.

Speakers can use this representation to form new instantiations of the construction (in generative theories of languages, this is referred to as licensing). An example of a partially abstract representation would be *FRUIT-JAM*, where the first element is underspecified: the name of any kind of fruit can be inserted here. A language user may have come across *strawberry jam*, *apricot jam* and *blueberry jam*. On the basis of the observed similarities in form and meaning, the partially abstract representation *FRUIT-jam* can come about. The language user can then apply this representation to form *açai jam*, for instance.

Underspecified elements in an abstract representation are often called *slots*. These are the points at which a speaker of a language can insert lexical units, resulting in different instantiations of the pattern. A slot can be filled by a smaller unit, but not by any unit: the slot fillers have to be compatible with the construction. Part of the representation of the construction is an underspecified representation for the slot. In the case of *FRUIT-jam* the commonality between *strawberry jam*, *apricot jam* and *blueberry jam* is that the first element is a fruit. This semantic aspect is part of the representation of *FRUIT-jam*. In forming a new instantiation of the pattern, a speaker has to combine the partially abstract representation with a lexical item that matches the underspecified element. In this simple case, this is a word that is similar to *strawberry*, *apricot* and *blueberry*, i.e. a fruit. What is specified about the slot fillers may be more or less specific, ranging from basic word class (e.g. plural –*s* requires a count noun) to more defined categories (e.g. V-ING TIME AWAY, where the TIME slot has to be filled by a phrase expressing a period of time).

Saying that a construction is productive if speakers have an abstract representation in their constructicon does not answer the question *which* expressions attested in a corpus were formed productively: the researcher is now faced with the new question of what it means that speakers have an abstract representation in their constructicon. The simple occurrence of different instantiations of a pattern is not sufficient evidence, because speakers can have these instantiations stored separately, without any abstract representations. Research has provided evidence that speakers store a lot of fixed units that could also be produced through the productive use of a construction. An example of such evidence is the observation of surface frequency effects for regular (i.e. conforming to the general pattern) morphologically complex words in lexical decision tasks. In lexical decision tasks, one of the variables affecting speed of recognition is the frequency of a word: the more frequent the word is, the faster it is recognized as a word. If a complex word is decomposed into its morphemes in language processing, the relevant frequency measures for recognition should be the frequency of the morphemes, rather than the complex item. Allegre and Gordon (1999) found frequency effects for the surface form, i.e. the complex word, in the recognition of regular plural nouns. These frequency effects were restricted to those plurals which where relatively frequent (more than 6 tokens per 1 million words), suggesting that infrequent forms are decomposed. For derivational morphology, Bertram, Schreuder and Baayen (2000) observed surface frequency effects for complex words with various derivational affixes in Dutch (*–heid*, *–te*, and agentive *–er*, (near) equivalents of English *–ness*, *–th* and *–er* respectively). These findings suggest that even for very frequent patterns which are both semantically and in form quite transparent speakers will have instantiations stored. This means that the occurrence of different instantiations of a pattern is insufficient to label this pattern as productive. The fact that some instantiations are stored, does not exclude the possibility that the pattern is productive.

At the surface, it is not possible to see whether a specific expression is formed through the productive use of a construction or has been retrieved as an unanalyzed whole: the plural form *houses* can be the result of [[*house*] *–s*] or [*houses*]. An individual speaker may have stored the form, but also have the abstract representation. As was argued before, in a usage-based and bottom-up view of language acquisition, the burden of proof is on the abstract pattern: the default assumption is that forms are stored (see Section 1.5.2.2). In sum, in order to determine whether a given construction is productive, it is not enough to observe forms are semantically and formally transparent instantiations of that pattern. Psycholinguistic evidence shows that such transparent or regular forms are often stored, especially when they are frequent.

Given the definition of a productive construction above, how then do we establish whether a construction is productive? The only way we can gain insight in the representations speakers have of a particular construction and its instantiations is through actual linguistic behavior, as the usage-based approach indicates, which is adopted throughout this book. Below, we argue that both corpus analysis and experimental research are useful sources of data.[2] In this chapter, both methods are employed to investigate two constructions in contemporary Dutch. Before discussing these constructions, the following section discusses some characteristics of productive constructions.

3.1.1 Characteristics of a productive construction

Given the premise that the development of abstract representations such as FRUIT-JAM is dependent on the input a speaker gets, it is reasonable to assume that characteristics of the input will influence the likelihood of such an abstract representation to come into existence. In this subsection, three elements are described in more detail: the frequency of items and the pattern they instantiate, the salience of the items and the pattern, and the decompositionality of items.

3.1.1.1 Frequency

A first important factor in the analyses of linguistic behavior is *frequency*. The occurrence of different types is a *sine qua non* for the development of abstract representations: speakers can only build up a schematic representation of a construction on the basis of different instances. If they only hear *dancing the night away*, they have no reason to assume that other utterances can be formed using a pattern of the type *V-ING TIME AWAY*. Encountering a large number of different types of a construction, i.e. many different instantiations of the same pattern, means that the similarities and differences between these types allow a speaker to develop a more detailed characterization of that construction. Meaning aspects that are shared between different types can be analyzed as part of the construction's meaning, and consistent characteristics of otherwise changing parts of the construction are stored as aspects of the lexically underspecified element (e.g. the phrase occurring between *V-ing* and *away* can take many forms, but is always some expression of a period of time).

[2] This sentiment is reflected in the recent special issue of Corpus Linguistics and Linguistic Theory (5.1, 2009), edited by Gilquin and Gries and in the forthcoming ICLC/CorpLing 2009 Volumes, edited by Gries and Divjak.

For this reason a construction with a high type frequency, many *different* instantiations, is more likely to be productive than one with a low type frequency. With regards to token frequency, the picture becomes a little bit more complicated. Bybee (1995) claims that high token frequency forms, with the same instantiation occurring often, will not contribute to the productivity of a pattern, because they *"can be acquired without forming relations with other items, and without undergoing internal analysis"* (Bybee 1995:434). In other words, types with a high token frequency will not easily be recognized as instantiations of a particular construction. This raises an important point: in saying that (type) frequency of a construction is relevant for its productivity, there is an often implicit assumption that instantiations are in fact perceived as such. Potentially, this leads to a circular argument: if we assume that in order for speakers to build up an abstract representation of a construction, they must first be able to recognize an expression as an instance of said construction, where does this recognition come from if not from meeting the construction in many different types? Avoiding this circularity requires a diachronic approach. As a starting point, consider that making an abstract representation can only take place on the basis of perceived similarities. Such perception of similarities crucially depends on some form of generalization, requiring a schematic representation. In the model that was introduced in Chapter 1, this is dealt with in terms of links between lexical items. As children come across different instantiations of a lexical item, they expand their representations of this item. This includes co-occurrence patterns and more detailed semantics. Taking up the earlier example of *apricot jam* and *blueberry jam*, both these lexical items will have *bread* and *sandwich* as likely collocates. The meaning of both includes notions like *fruit, spread* and *sweet*. Because both lexical items have these semantic aspects, they will become linked. The overlap in form provided by *jam* occasions another link. Together these links make the common pattern in these lexical items more salient. This can serve as the basis for a partially abstract representation like *FRUIT jam*. Note that the generalization is now the <u>result</u> of the similarities, and that these are acquired through exposure.

It has been shown that infants 'count' and recognize recurrent patterns; babies react differently to sequences they have heard more often than less frequent concatenations of syllables or unheard syllables (e.g. Saffran, Aslin & Newport, 1996). Distributional analysis of language input has been argued to be an important tool children use for word segmentation (Saffran, 2001), word class (Gómez & Gerken, 2000; Mintz, 2003, 2005) and many other language processes (for a recent overview, cf. McMurray & Hollich, 2009). These studies mainly looked at syllable structures, but we may extend this general ability in sound perception to more complex and less transparent recurrent patterns. A usage-based perspective on language acquisition assigns a central role to such pattern recognition.

Having said that, the formation of an abstract representation may not always take place. Linguists of different theoretical persuasions tend to try to find highly abstract and general rules or representations. An example is Goldberg's (2006, chapter 8) analysis of subject-auxiliary inversion (SAI). In this analysis, she claims that there is an underlying commonality beneath a large number of seemingly rather different functions of the pattern: the main attribute of SAI is "non-positive" (Goldberg 2006:171). SAI is a deviation from the prototypical sentence, with *''the prototypical sentence being a declarative, positive assertion with predicate-focus information structure."* (Goldberg 2006:168). This analysis has been strongly criticized, both by researchers who provide examples of SAI which do not contain the aspects listed in the definition (Borsley & Newmayer, 2009) and by others who question the need and indeed the possibility to formulate one general pattern or function when the actual distribution patterns of specific functions are very different from each other (Croft 2009). At this point it is useful to underline the difference between patterns a linguist may detect and the psycholinguistic reality of these patterns: even if (diachronically or synchronically) one all-encompassing construction can be identified, this need not correspond with speakers' representations in their constructicon (this point will be taken up in Chapter 5). Whether strongly dissimilar instantiations of different uses of subject-auxiliary inversion (e.g. Wh-questions and comparatives) should be counted as tokens contributing to the frequency count of the general pattern is therefore debatable. (cf. also discussions in Croft, 2001 and Boas, 2003).

The V-BAAR construction and the IS TE V construction that are the topic of analysis in this chapter show comparable frequency patterns, both in types and in tokens (see Sections 3.2.1.2 and 3.2.2.2). The type-token distributions are skewed: the two constructions each have a small number of types that occur very frequently and a larger number of infrequent types. The implications of this distribution will be discussed in the last section of this chapter.

Setting aside the question what counts as an instantiation of a pattern for now, in what follows, frequency alone will not be seen as sufficient evidence for the productivity of a construction. In addition, frequency is not the only factor that determines whether an abstract representation will arise: in the remainder of this section, we briefly discuss two other relevant factors: *salience* and *decompositionality*. In the literature on (first and second) language acquisition, these have been identified as related to the speed with which generalized patterns are acquired. These factors may then also be hypothesized to be relevant for the productivity of a construction: people are more likely to store abstract representations for constructions with many decompositional and salient instantiations that cause the commonalities to be more easily recognizable.

3.1.1.2 Salience

Salience, like productivity, is based in the language user: something is only salient *in the eyes (ears) of someone*. Salience is the perception of something as standing out in contrast to its surroundings, the *figure* to a *ground* (cf. Langacker 1987:120). This figure-ground organization is present at all levels of linguistic structure: in a given conversation an event is salient, and in an intonational contour there is usually a specific syllable that receives the main stress. Talmy (1996) uses the term *attentional salience*, emphasizing thus that salience attracts a language user's attention to a linguistic element. According to Talmy there are many linguistic devices for the setting of attentional salience, including "*the presence versus the absence of overt language material (...), positioning at certain sentence locations instead of other locations, head versus non-head constituency within a construction, degree of morphological autonomy, solo expression versus joint conflation, phonological length, and degree of stress*" (Talmy 1996:237, footnote 2). This list illustrates that salience is a relative notion and is always contextually defined.

Whether a particular form strikes the attention of a speaker/listener is to a very large degree dependent on the linguistic material surrounding this form. In her *Graded Salience Hypothesis*, Giora (2003) states that *"more salient meanings -coded meanings foremost on our mind due to conventionality, frequency, familiarity, or prototypicality- are accessed faster than and reach sufficient levels of activation before less salient ones."* (Giora 2003:10). This psycholinguistic view implies that a linguistic structure's salience is the coded result of a speaker's encounters with that structure. Note that the reasons that are mentioned why a structure would be coded "foremost on our mind" can also be interpreted to be context-dependent (defined by Giora as both linguistic and extra-linguistic): conventions are, to a certain extent, register-specific and frequency can include recent occurrences. Including such temporary factors allows the hypothesis to make correct predictions about the influence of discourse topic and to explain some priming effects (e.g. why it is easier to recognize the word *syrup* after you have heard *maple* than if you have not).

Salience, thus, is a context-dependent notion. It is possible, however, to identify variables that make a form-meaning pairing more or less salient. In a meta-study on the acquisition order of grammatical morphemes in English as a second language, Goldschneider and DeKeyser (2001) defined five predictors that together explained 71% of the variance in accuracy scores in 12 studies: frequency of a given morpheme, phonological salience, semantic complexity, morphophonological regularity and syntactic category. These five characteristics measure different aspects of salience; they address the ease of identification of the phonological form or forms and the distribution of the morpheme in its context.

The first characteristic, frequency, was discussed in the previous section. Goldschneider and DeKeyser predict that morphemes that occur more frequently in the input a learner hears will be acquired earlier than those that do not occur often. Secondly, they scored the phonological salience of morphemes by counting the number of phones and determining the "relative sonority" of the morpheme as well as determining whether it has a vowel (i.e. if it has syllabicity). The third variable, semantic complexity, takes into account how many meanings a particular form can express. Note that for this variable, a negative correlation is expected between complexity and acquisition order: forms with few meanings are hypothesized to be acquired early. Morphophonological regularity, the fourth determinant, deals with the number of different surface forms in which a morpheme may occur (e.g. plural –*s* can be realized as [s], [z] or [əz], depending on the phonological environment). Finally, syntactic category refers to the types of stems a morpheme can attach to. Goldschneider and DeKeyser distinguish between lexical and functional as well as bound and free stems. Formulating it slightly differently, Seidenberg and Gonnerman (2000) address some of the same aspects when they state that *"units might be perceptually salient not because there is a distinct morphological level but because they happen to make consistent orthographic, phonological and semantic contributions to different words."* (Seidenberg & Gonnerman 2000:355).

A large caveat is in order, however, before we extrapolate these results directly to the constructions investigated in this chapter. Goldschneider and DeKeyser looked at studies investigating the acquisition of *grammatical morphemes*, i.e. verbal inflection etc. It is not self-evident that the analysis can be extended to all constructions. However, the structures of grammatical morphemes and partially specific constructions are not that different: there is a specific element (the grammatical morpheme, the –*baar* affix or the *is te* sequence) that occurs in combination with various lexical elements (stems of particular word classes for the grammatical morphemes, verbal stems and infinitives for the V–BAAR and IS TE V constructions). The language learners' task in each case is to identify the commonalities in form and meaning between different instances of each pattern, given the variation provided by the slot fillers (the stems or verbs).

Given these similarities, it is a useful exercise to look at how the constructions fare on each of the characteristics proposed by Goldschneider and DeKeyser, leaving out the already discussed element of frequency. The notion of phonological salience does not serve to distinguish between the two constructions: both of them are salient in this respect. The absence of syllabicity may be restricted by and large to bound morphemes, so this variable cannot be applied easily to larger constructions.[3] The semantic complexity of the fixed elements in the two constructions differs quite dramatically: the affix *–baar* is an extremely reliable indicator of the V–BAAR construction; there are hardly any words that end in *–baar* yet are not instances of the construction. The story is rather different for the IS TE V construction: the only constant element, *te*, is found in many other contexts, both pre-infinitival (see examples 1 and 2) and in completely different contexts (3).

(1) Ik zit **te** wachten
I sit **to** wait
"I'm waiting"

(2) Zij is leuk om **te** zien
She is nice for **to** see
"She is pretty"

[3] Bybee (2000, 2002) observes that phonological reduction of complex lexical items, both morphologically complex words and larger units, often coincides with frequency of that complex item. The past tense affix *–ed* in English verbs, for instance, has a longer duration in infrequent forms like *kneaded* than it does in *needed*. Bybee takes this to be evidence that frequent complex words are stored and produced as simplex wholes, whereas infrequent forms are produced compositionally. In an article that focuses on *don't*, she observes that the vowel is only reduced to a schwa in contexts that occur frequently: when it is preceded by *I* and followed one of a short list of verbs (including *know, care, want, think* and *have*). Again, the frequency of the sequence coincides with phonological reduction (see also Jurafsky et al., 2001).

(3) Je bent **te** goed voor deze wereld
You are **too** good for this world

In a usage-based approach to language acquisition, it is straightforward that this complexity may influence how easy it is to identify instances of the construction as such and therefore the speed with which the construction is acquired. In this context, salience means that a specific instance can be spotted as such, and the IS TE V pattern seems to be less salient than the V–BAAR construction.

With regard to morphophonological regularity, again the two constructions are quite similar. Instances of V–BAAR words occur in two forms, depending on whether they are used attributively of predicatively. Dutch has some, if only very little, adjectival inflection. Adjectives occur in two surface forms: the bare form as in *groen* (green) or *drinkbaar* (drink-able) and with a schwa (-e) ending (*groene*, *drinkbare*). Because of spelling conventions, this means that instantiations of the V–BAAR construction end in either –*baar* or –*bare*.[4] There is no such difference in forms for IS TE V, but if we restrict ourselves to predicative use, there is only one form for V–BAAR as well.

The syntactic category variable identified by Goldschneider and DeKeyser is the second for which V–BAAR and IS TE V are different: –*baar* is a morpheme that attaches to verbal stems, whereas the slot in the IS TE V construction is a whole word, i.e. the verbal infinitive. It is not completely clear which is more beneficial in terms of salience: on the one hand, the verb can be recognized easier in the IS TE V construction, as it is a separate word (at least in writing this is likely to be an advantage), but on the other hand, the stem-affix pattern is very frequent in Dutch, and the verbal stem occurs in many other derivations, compounds etc.

In sum, it seems that the V–BAAR pattern is more salient, mainly because the fixed element is a very reliable indicator for the construction. This indicates that *ceteris paribus* the V–BAAR construction is easier to learn and will be acquired sooner. It is not necessarily the case that constructions which are acquired more easily will also be more productive, but it is likely that this is a tendency: speakers of a language will not always use abstract representations when they process utterances (Ferreira 2003) and for some constructions, not all speakers will actually acquire (partially) schematic

[4] The bare form of the adjective occurs when it is used attributively with an indefinite, singular, neuter noun (*een groen huis* 'a green house') and in predicative position (*het huis is groen* 'the house is green'). The form with –*e* is found in all other cases.

representations (Dąbrowska & Street, 2006). If a salient construction is acquired more easily, more speakers will develop a schematic representation, which they can then use to productively form new instantiations of the pattern.

3.1.1.3 Decompositionality[5]

Closely related to the concept of salience is that of *decompositionality*: for a speaker to be able to acquire an abstract construction, the form must be salient, but also the meaning and form need to be retraceable. Bybee and McClelland (2005) state *"in derivational morphology we have situations in which words vary by degrees in the extent to which they can be decomposed into the semantic and phonological units that originally comprised them."* (Bybee & McClelland 2005:393). They identify four continua on which each stem + affix derivation can be placed in terms of compositionality. These continua are listed in Table 3.1 below.

Table 3.1: Continua of decompositionality in derivations, adapted from Bybee & McClelland 2005

Continuum
1. Degree to which the affix is identifiable with other instances of the affix
2. Degree to which the base is identifiable
3. Extent to which the semantics differs from the semantics of etymologically related forms
4. Degree to which the phonology is predictable from the base

Whereas they look at derivational morphology and the formation of complex words, these continua are easily extended to other constructions: *affix* can be replaced by *fixed elements in the construction* and *base* by *slot fillers*. Table 3.2 contains a reformulation of the continua and indications where the V–BAAR and IS TE V constructions would fall on them.

Bybee and McClelland are careful not to suggest that high scores on their continua would imply that speakers will actually decompose the

[5] The term *compositionality* is arguably more common than its counterpart *decompositionality*. but we feel that the latter is more in line with an input-based view of language acquisition, in which it is assumed that children start out by assigning meaning to language, only analysing it into smaller meaningful units when this is necessary (cf. the *needs only analysis* in Wray 2008).

derivations in their language use. As with salience, however, we can argue that high scores on these continua facilitate the recognition of a pattern. Both constructions score relatively high on these continua.

Table 3.2: V-BAAR and IS TE V on Bybee & McClelland's decompositionality continua

Factor	V-baar	is te V
Fixed element(s) identifiable with other instances	Yes	No, fixed elements also occur in many other constructions
Slot fillers identifiable	Quite, verbal stems	Completely, separate words
Semantic difference	Probably equal for both constructions: semantic attribution of the verb to the instantiation is clear	
Predictable phonology	Yes	Yes

While decompositionality is a *sine qua non* for productive use, this does not mean that instantiations that are transparent in form and meaning will not be stored as units. In his 2003 book on resultatives, Boas points out that not all constructions are either conventional *or* compositional. According to him, there is an important intersection of cases that are both conventional *and* transparent, which he calls *motivated constructions*. Boas observes that a lot of resultatives are regular and transparent, which seems to point to an account of the resultatives in terms of general constraints, i.e. a decompositional analysis. The distribution however, actually shows type specific behavior that cannot be captured within these general constraints. He concludes that *"…in the case of resultatives, it is possible to observe great regularity (in the types of resultative phrases that go with certain types of verbs) that is conventionalized in a specific way (i.e., individual verbs determining the types of postverbal constituents they combine with in resultatives)"* (Boas 2003:141-2). An example is the verb *drive* which, in the resultative construction, pretty much exclusively occurs with negative mental states in the slot of the resultative phrase (e.g. *mad, insane, up the wall*).

The notion of motivated constructions is highly relevant to our discussion about productivity: the fact that types of a construction can be both conventional *and* transparent translates easily in an account that focuses on psycholinguistic reality and the linguistic inventory of speakers. These motivated constructions can be decoded regardless of whether they have been encountered before, because they are decompositional. The *encoding* is a different story: specific conventionalizations must be stored in order for appropriate use. Our corpus data show an abundance of specific conventionalizations, which suggests that speakers typically produce stored types and do not form instantiations productively. But that does not mean that they are not able to decode new types: the motivated constructions are largely transparent and, as such, provide information on a possible partially specific representation.

In sum, we have argued that constructions that are frequent in terms of types, salient and decompositional in form and meaning are likely to be productive. We have also emphasized that transparent forms need not be formed productively. In fact, if we find specific conventionalizations of the kind Boas observed with resultatives, this is evidence for the storage of these instantiations, even if they are transparent. We will now turn to our case study: an analysis of two Dutch constructions that express potentiality.

3.2 Corpus study of V-BAAR and IS TE V

In the remainder of this chapter, we focus on two constructions. Both convey potentiality, in the sense that they provide a comment on the possibility of a certain action to be carried out. Languages seem to have a need for expressing this notion; while we have not undertaken a typological survey, it does seem that most if not all languages have constructions for expressing it, and probably all have more than one. Surface constructions tend to vary, from morphological inflection of the verb (e.g. in Turkish) to syntactic constructions with modal auxiliaries ('can'). There are two structurally very different constructions in Dutch that express potentiality, one morphological and one syntactic in nature, which has prompted our theoretical interest in them.

While our primary research interest is in productivity and whether its manifestations in morphology and syntax are essentially identical, and the two focal constructions are mostly chosen for their illustrative potential, inevitably questions also come up about the semantic domain they derive from. Our analyses will point towards the characteristics of the verbs that occur with one or both of the constructions. Note it is not really to be expected that any verbs (except modals and other auxiliaries) are in principle excluded from potentiality constructions, given that any activity for which we have a verb will likely be something that can potentially happen or be carried out. If we find common patterns in the distribution of verbs, therefore, the constructions are likely to encode something slightly more specific than just potentiality. It is also worth pointing out here that the nameworthiness of potentiality or the lack thereof is context dependent; speakers will not go about stating the obvious (i.e. that the activity named by the verb could potentially be carried out).

For each of the two constructions, the deverbal adjective V–BAAR construction and the modal infinitive IS TE V construction, the following sections provide a corpus analysis. A corpus search can provide insight in how a particular construction is used. In this article, we use the recent Corpus Gesproken Nederlands (Corpus of spoken Dutch, hereafter CGN). This is a 10 million-word corpus consisting of spoken data from different genres, ranging from telephone conversations to official speeches. Around two thirds of the corpus is from speakers in the Netherlands, and one third is from the Dutch-speaking part of Belgium. These distributions, and the extensive metadata available, allow us to analyze differences between groups of speakers, regions, and type of discourse etc. All tokens are tagged in a number of ways, including the lemma they belong to and the word class of each token (e.g. noun, article etc.).[6]

3.2.1 The V-BAAR construction

3.2.1.1. Identifying the instantiations of the construction

In order to track all instantiations of the construction in the CGN corpus, we initially searched for all tokens of lemma's ending in –*baar* and tagged as an adjective.[7] In addition to this first search, we looked for cases in which the word in –*baar* was part of a larger word. The most frequent of these were comparatives (N = 34) such as *betrouwbaarder,* 'rely-able-er, more reliable'. There were also a few superlatives like *kwetsbaarst,* 'hurt-able-st, most vulnerable' (N = 5). Finally, the corpus contained 19 instantiations with –*baar* of the nominalizing construction IETS X_{ADJ}-S, 'something that is X', as in *iets vergelijkbaars*, 'something comparable'. In all of these cases, the –*baar* word

[6] For more information on the CGN corpus, see http://tst.inl.nl/cgndocs/doc_English/start.htm.
[7] This meant the net was cast a little bit too wide: the monomorphemic adjective *baar* (only used in the fixed expression *baar geld*, cash) was included (but N = 1).

itself also occurred separately in the corpus. These words that are derived from V–BAAR instantiations were not included in the analyses.

Since the IS TE V construction is very rarely found prenominally (see Section 3.2.2), we decided to limit the contrastive analyses to the predicative use of V–BAAR. Therefore, a more restrictive corpus search was executed, including only non-attributive tokens of -*baar*[8], followed by a manual search to exclude all tokens whose grammatical function could not be established (e.g. in one word utterances). All adverbial tokens, as in (4), were discarded from the analyses as well, because this function, like the attributive, is not available for the IS TE V construction.

(4) En toen riep hij **hoorbaar** voor iedereen kom eens hier schat...[9]
And then he yelled audible to everyone just come here honey
"And then, audible to everyone, he yelled, come over here, honey" (fv800018.24)

The following pages report the quantitative results.

3.2.1.2 Main search results V-BAAR construction

There are no fewer than 261 different types in CGN which are tagged as an adjective and end in –*baar*. In all, we find 3908 tokens of adjectives ending in –*baar*. As is common with constructions, a relatively small number of types accounts for a majority of the tokens (Zipf, 1935; Ellis 2005:320). Table 3.3 lists the ten most frequently occurring words, totaling 2189 tokens, or slightly more than half of the instantiations. Their translations will be given in Table 3.4.

Among these ten most frequent types, we encounter a lot of variation with regard to stem types. Two words have an adjectival stem (*openbaar* and *middelbaar*). These forms have been suggested to derive from a different affix (Van den Berg, 1974) and in any case their meaning has nothing to do with potentiality.

[8] These are tagged as "adjective, free" in the CGN corpus.
[9] All examples are taken from the CGN, unless a different source is specified.

Table 3.3: 10 most frequent adjectives ending in –*baar* found in CGN

Type	Nr. of occurrences
blijkbaar	1134
openbaar	306
beschikbaar	173
middelbaar	115
zichtbaar	103
onvoorstelbaar	81
schijnbaar	74
haalbaar	71
bereikbaar	69
dankbaar	63

The fact that the accent in *openbaar* falls on the last syllable (contrary to all other adjectival words in –*baar*) means that -at least synchronically- this word must be analyzed as simplex. The word *zichtbaar* has a nominal stem, and the stem in *dankbaar* could be either nominal or verbal. Both have been analyzed to be remnants from an earlier stage in which the affix –*baar* attached to nominal stems (Koelmans, 1981; Van Marle, 1990). Potentiality is present in *zichtbaar*, where the noun *zicht* 'sight' combined with –*baar* results in 'visible'. *Dankbaar*, on the other hand, can probably be analyzed as an older form: its meaning 'grateful' can be paraphrased as 'bearing or entailing thanks', with *dank* meaning 'thanks' (cf. *vruchtbaar*, 'fertile', literally 'fruit-bearing'). A further two words, *blijkbaar* and *schijnbaar*, have copular verbs as stems (*blijken* 'seem', *schijnen* 'appear'). The affix –*baar* hardly contributes to their meanings, 'seemingly' and 'apparently', respectively.

The remaining four words each have a verbal stem, with two participants: a subject and either a nominal direct object, a prepositional object or a sentential object, and match better with the meaning a compositional semantic analysis would yield. Table 3.4 summarizes the variation in both form (stem) and meaning.

Table 3.4: Form and meaning of the ten most frequent adjectives ending in *-baar* found in CGN

Type	Stem type	Meaning
blijkbaar	Verb (copula)	Apparently
openbaar	Adjective	Public
beschikbaar	Verb (subject, prepositional object)	Available
middelbaar	Adjective	Secondary, middle(aged)
zichtbaar	Noun	Visible
onvoorstelbaar	Verb (subject, nominal object/sentential complement)	Unimagin**able**
schijnbaar	Verb (copula)	Seemingly
haalbaar	Verb (subject, object)	Feasible
bereikbaar	Verb (subject, nominal object/sentential complement)	Reach**able**
dankbaar	Noun/verb (subject, nominal object)	Grateful

At the other end of the frequency scale we are more likely to encounter non-lexicalized forms: instantiations that are formed productively on the basis of a more abstract template are less likely to occur in more than one conversation, or to be produced by different speakers independently from each other. For this reason we selected all types that occur no more than four times, rather than only selecting hapaxes legomena, since simple repetition within a single discourse context may cause a productively formed word to occur more than once. From these we excluded all words that are found in the latest edition of Van Dale's Hedendaags Nederlands Woordenboek (14th edition, 2005), the most widely used Dutch dictionary, leaving us with a list of fifty words. Of these, 48 have a simple verbal stem, the exceptions being *waterverdunbaar* and *zelfopblaasbaar* (literally: 'water-thin-able' and 'self-inflate-able'), which do contain a verb but have incorporated an extra element. In terms of the selection of the stem, therefore, this list is much more homogeneous than that of the most frequent types. The meaning of 46 out of the 50 types could be paraphrased as 'can be V-ed', i.e. they are semantically transparent and have a strong potentiality related meaning. Four words have an intransitive verb as their stem: *adembaar*, *roddelbaar*, *bederfbaar* and *onoordeelbaar* ('breathe-able', 'gossip-able' 'perish-able' and 'un-judge-able' respectively). For each of these verbs, there is some sort of entity undergoing the action of the verb: air, the topic of gossip, food that goes off and the case or behavior that cannot be judged. The interpretation of the derivation is that this entity can undergo the action or event conveyed by the verb. When used as a verb, *breathe* need not have a lexical expression of the object *air*, but it is always part of its conceptual base (cf. Langacker 1991:364). In the V–BAAR construction, the entity which undergoes the action is the trajector. In prototypical transitive verbs this is the object, but it seems that this analysis is easily extended to objects often left implicit.

In sum, a first analysis of all tokens of the V–BAAR construction shows that these tend to have a transitive verb as a stem, with the exception of a small number of very frequent types, and a few non-prototypically transitive verbs.

The corpus search for all non-attributive words ending in *–baar* results in 1901 tokens (240 types). Because of our interest in forms that might be productively formed, we exclude all types that do not contain a recognizable verbal stem (14 types) and the two types with a (diachronically present but semantically opaque) copula verb stem, *blijkbaar* and *schijnbaar*, that both translate as 'apparently'. A manual check to exclude adverbial use and non-analyzable tokens (mainly one or two word utterances) leaves us with 1282 tokens (206 different types[10]). The restriction to *–baar* forms which are used predicatively therefore does not have a substantial impact on the total number of types; remember that 261 different adjectives ending in *–baar* were found in the corpus. This means that most types (206) occur predicatively at least once. The total number of tokens does go down dramatically: from 3908 to 1282 tokens. The main reason for this is the fact that some of the most frequent types completely disappear: *blijkbaar, openbaar, middelbaar, schijnbaar* and *dankbaar* do not occur predicatively. The ten most frequently occurring predicative forms are listed in Table 3.5. Although the ten most frequent types still account for close to half of all the tokens (501/1282), from the translations it is clear that we are now dealing with a much more homogeneous category in terms of the meaning of the instantiations.

[10] This number is smaller than 240 – 16 excluded types, because for 18 types the token(s) were excluded in the manual check.

Table 3.5: Most frequent V–BAAR types used predicatively

Type	Nr. of occurrences	Meaning
beschikbaar	141	Available, reachable
haalbaar	62	Feasible
bereikbaar	56	Reachable
onvoorstelbaar	47	Unimaginable
vergelijkbaar	38	Comparable
verstaanbaar	35	Intelligible
bespreekbaar	33	Discussible
herkenbaar	31	Recognizable
bruikbaar	30	Usable
kenbaar	28	Knowable

In predicative contexts, V-BAAR instantiations are most often used with a form of *zijn* ('be', see example 5), but there are also a number of tokens with a causative verb. This is mainly *maken*, to make, as in example 6, but some words combine with one or more other verbs such as *stellen* ('put', see examples 7 and 8). These combinations sound very familiar to the native speaker, which suggests that the use of such verbs seems to be restricted to some types, but since they are quite rare, our corpus is of insufficient size to determine whether the non-occurrence of such verbs with other types is more than a coincidence.

(5) Ik **ben** eigenlijk alleen overdag **bereikbaar**
I am really alone daytime available
"I can really only be reached during daytime."(fn006994.85)

(6) maar je kan bepaalde delen van je harde schijf niet **leesbaar maken** voor Linux.
But you can certain parts of your hard disk not legible make for Linux
"but certain parts of your hard disk cannot be made legible for Linux" (fn000363.218)

(7) tenzij de wetgever lidmaatschap van het Blok openlijk **strafbaar** stelt
Unless the legislator membership of the Block openly **punishable** puts

"unless the legislator openly makes membership of the Front [a political party] a punishable offense" (fv601284.19)

(8) verkeersminister Tineke Netelenbos stelt opnieuw tien beurzen **beschikbaar** voor vrouwen die hun vrachtwagenrijbewijs willen halen.
Minister_of_transport Tineke Netelenbos puts again ten stipends **available** for women who their truck_driver's_license want get
"Minister of transport T.N. makes another ten stipends available for women who want to get a license for driving big trucks" (fn004372.1)

Some utterances contain a lexical marker of the speaker's stance (e.g. when the verb *lijken*, to seem, is used, as in (9).

(9) dat lijkt me ondenkbaar.
That seems me unthinkable
"that seems unthinkable" (fn007189.63)

Finally, there are many instantiations with low token frequency in the corpus. These instantiations are all semantically transparent. This suggests the pattern is productive; the degree to which this really is the case is not so clear, however: only few of the low-frequency occurrences actually come across as 'unfamiliar', a notion that will be substantiated in subsection 3.2.1.5.

3.2.1.3 Collostructional analysis of V–BAAR

Up until this point, the corpus data have only been discussed in terms of simple type and token frequencies. In essence, we are looking at the co-occurrence pattern of a partially specific construction (Subject V_{fin} V-BAAR) and fillers of the slot for the verbal stem that precedes *–baar*. Raw frequencies are not optimally informative in this case: the fact that a certain *–baar* form occurs predicatively five times in the corpus, as is the case for *betwistbaar* 'questionable', *brandbaar* 'flammable' and *eetbaar* 'edible', may or may not be significant given the overall frequency of *betwisten, branden* and *eten*. One way to resolve this issue is to do a collostructional analysis.

This procedure has been proposed by Stefanowitsch and Gries (2003) and measures whether a given lexical item, in this case a verb, is significantly attracted to a certain construction, here: the V–BAAR construction. It does so by contrasting the expected frequency under the null hypothesis that the lexical item will occur as frequently in the construction as it does elsewhere in the corpus with the observed frequency of the lexical item in the construction. Only four numbers are required for the calculation: the total N of instantiations of the construction, the N of instantiations with the particular lexical item, the total N of the lexical item, and the N of constructions. For each verb, lemma counts in CGN were chosen as the comparison and the total number of predicates in the corpus (1 076 041) as the background number. This implies that V-BAAR tokens are not included in the token counts for each verb, while for IS TE V (see subsection 3.2.2.4) they are. While this certainly causes a certain degree of imbalance, we see no better alternative: if V-BAAR tokens are to be included in the count, other derivations should be too. It is far from clear, however, to what extent morphologically derived forms are really derived in speakers' constructicons, especially in the case of semantically or orthographically opaque derivations, diachronically motivated but synchronically non-interpretable forms, etc. In this particular case, the ranking would largely remain the same if V–BAAR tokens were included in the verb token count. The statistical test used is the Fisher-Yates exact test. Applying this test to each verb results in a list of verbs that occur significantly more often than would be expected by chance (given their frequency throughout the corpus). The strength of the association given is the log-transformed p-value of the Fisher-Yates test. In sum, the list of collexemes provides an answer to the question which words are typically used with a particular construction, given their overall frequency in the corpus. Collostructional analysis has been used in various contexts: among others in combination with experimental data (Gries, Hampe & Schönefeld, 2005), on a diachronic corpus (Hilpert, 2006), and to show differences between spoken and written language (Stefanowitsch & Gries, 2008).

This method provides a useful tool for the description of constructions. Knowing which verbs are attracted to the construction will help in establishing the characteristics of the open slot in the construction and the meaning of the construction in general. The analysis will not tell us whether the construction is productive, but it does provide input for the formulation of hypotheses about which types of verbs can easily combine with the construction to form acceptable utterances. These can subsequently be tested experimentally.

Stefanowitsch and Gries chose the Fisher-exact test for their analyses, because this test does not make any distributional assumptions or require a particular sample size (Stefanowitsch & Gries 2003:218). This is important, because often, as in this case, there is no normal distribution, and many types occur only a few times in the corpus. We used R and the script made available by Gries (2007) to compute the test.

Collostructional analysis can also be used to determine which words are repelled by a construction, i.e. which words occur significantly less often than would be expected given their frequency and the frequency of the construction. It is even possible to include words that do not occur at all in the construction, and see whether that non-occurrence is significant (cf. Stefanowitsch, 2006 for a discussion on the significance of absence). In this case, we felt that a list of repelled verbs could only contribute properly to our understanding of the construction if all possible candidates were included in the analysis, i.e. if a collostructional analysis was done for all verbs in the CGN. Since this is practically impossible, it was decided to only include verbs that occurred at least once.[11]

---

[11] In CGN, all words are tagged as tokens of a lemma. To determine the total frequency in the corpus for each verb, we counted all instances of the lemma, provided they were also tagged as verbal forms. The problem is that Dutch has a number of particle verbs. While these verbs occur as single words in some contexts, such as non-finite forms and in subordinate clauses (see examples 1 and 2), when they are the finite verb in a main clause, the two elements of the verb occupy different positions in the sentence (example 3). Unfortunately, lemma tagging of these verbs in CGN is limited to the single word forms. The separated tokens of the verbs are counted as instances of the particle-less verb, which in many cases is a frequent verb in itself, for example *afbreken* 'tear down', lit. 'off-break' and *breken* 'break'. As it would be a very time-consuming job to manually retrieve these forms, we have to make do with the automatic counts.

1. en we hebben die uhm die pergola **afgebroken** dat uh lelijke ding
   and we have have that uhm that pergola broke_off that ugly thing
   "and we've torn down that ugly pergola" (fn008022.200)

2. ze wil wel dat ge haar huis een beetje proper houdt dat ge 't niet **afbreekt** hé maar ja da 's logisch
   she wants still that you her house a little clean keep that you it not off_break right but yeah that's ogical
   "she does want you to keep the house sort of clean, that you don't tear it down, right, but that's normal" (fv901053.142)

3. maar op dit moment **breekt** mijn rotbeugel **af**.
   But at this moment breaks my stupid_braces off
   "but at this moment, my stupid braces break" (fn007672.16)

In this analysis, and henceforth, forms of V–BAAR and ON-V-BAAR with the same verbal stem (as in *betaalbaar* and *onbetaalbaar*, 'affordable' and 'unaffordable') were counted as belonging to the same type. Table 3.6 lists the twenty most strongly attracted verbs (collexemes) to the V–BAAR construction.

Table 3.6: Collexemes of *–baar* (values given for collostructional strength correspond to –log (p-$_{valueFisher-Exact}$, 10))

Collexeme (N)	Translation Verb	*-baar* form	Collostructional strength
Beschikken (141)	Have at one's disposal	Available	Inf
Feilen (2)	Err	Fallible	Inf
Bereiken (68)	Achieve, reach	Can be reached	105.0175
Aanvaarden (38)	Accept	Acceptable	70.89466
Vertrouwen (37)	Trust	reliable	70.0045
Kwetsen (28)	Hurt	Vulnerable	69.96488
Verstaan (41)	Hear,understand	Audible	60.05894
Halen (72)	Get	Feasible	58.51185
Voorstellen (52)	Suggest	Imaginable	54.41769
Vergelijken (38)	Compare	Comparable	53.40882
Bespreken (36)	Discuss	Discussable	51.32487
Vatten (25)	Get, catch	Susceptible, prone to	49.13478
Herkennen (34)	Recognize	Recognizable	47.93689
Straffen (22)	Punish	Punishable	45.00074
Voorspellen (24)	Predict	Predictable	42.44541
Betalen (41)	Pay	Affordable	32.20688
Gebruiken (37)	Use	Useable, useful	27.37977
Verkrijgen (12)	Get, receive	Available	22.98648
Afwenden (8)	Avert	Avoidable	18.30624
Besturen (9)	Drive, operate	Can be driven, operated	16.78183

We analyzed *betrouwbaar* as containing the stem *vertrouwen* and *bruikbaar* as derived from *gebruiken*. The verbs *betrouwen* and *bruiken* do not exist, and the meanings of the *–baar* words link up reliably with these two verbs.

A comparison of Tables 3.5 and 3.6 reveals that most of the frequent V–BAAR forms contain verbs that are significantly attracted to the construction as well, but the lists are certainly not identical: the infrequent verb *feilen* 'err' is one of the two most strongly attracted collexemes for this construction, even though *feilbaar* 'fallible' occurs only twice. This is because there is only a single token of the verb outside the construction in the corpus. The example of *feilen* illustrates the added value of collostructional strength over raw frequency data. However, the advantage most important to us is that it allows a more fine-grained semantic characterization of the construction.

3.2.1.4 Constructional meaning of V-BAAR

Based on the corpus data, the meaning of the V–BAAR construction can be given as the most Patient-like argument of the verb (often the second argument, with the first argument being an Agent) has the property of being able to undergo the event described in the verb.

We will discuss each aspect of this definition. Firstly, the V–BAAR instances describe a **property**; this meaning aspect is prototypically expressed in Dutch with an adjective. In its predicative use, the V–BAAR construction occurs mainly as *SUBJ BE V-BAAR*, which indicates that the property ascribed to the subject is construed as stable, as in (10). There are also instances where the verb *worden* 'become' is used, which causes the V–BAAR item to be the result or development of a process (11 and 12). Finally, in a number of sentences with V–BAAR the finite verb is a causative verb like *maken* ('make', see example 6 above). In these cases, the sentences express that some Agent causes the subject to have the property, where that previously was not the case (see also example 13).

(10) maar jouw dialect moet wel **verstaanbaar zijn** voor anderen
But your dialect must still **intelligible** be for others
"but other people still need to be able to understand your dialect" (fn008312.245)

(11) door 't noodweer van de afgelopen dagen is in Frankrijk één van de belangrijkste autosnelwegen **onbegaanbaar geworden**.
Due_to the heavy_weather of the past days is in France one of the most_important highways **unridable** become

"because of the heavy weather over the past few days, it has become impossible to use one of the most important highways in France" (fn004858.1)

(12) hij verwacht dat Den Haag beter **bestuurbaar wordt**.
He expects that The Hague better **governable** becomes
"he expects that it'll become easier to govern The Hague" (fn004858.1)

(13) ze had haar keuken **onbruikbaar gemaakt** met nutteloos vergif
She had her kitchen **unuseable** made with useless poison
"she had made her kitchen unusable because she had used some ineffective poison" (fv801341.28)

Although it is not easy to put a finger on it, for quite a few instances, the construction seems to express more that a Patient may easily -without any Agent's intention- be affected rather than that it is *possible* to affect the Patient (see examples 14 and 15). The focus is not so much on the potentiality aspect in itself, but how easy it is.

(14) dus 't is heel vreselijk dan ja je ben je bent ook heel **kwetsbaar** nog wel op dat toneel.
so it is very terrible then yes you are you are also very **hurt-able** still yet on that stage
"so it's horrible; you are also very vulnerable on that stage" (fn007371.128)

(15) stress is een groot woord misschien maar ja iedereen is toch wel heel erg **prikkelbaar** denk ik.
Stress is a big word maybe but yes everyone is anyhow yet very badly **irritable** think I
"Stress is a big word maybe but yeah, everyone is extremely irritable, I think" (fn007487.73)

The property is ascribed to an entity, which in the semantic frame of the verb fills the most Patient-like role. Very often, this is the direct object of the verb when used in an active sentence. It would not be accurate, however, to state that the property is always ascribed to the direct object, as examples (16) and (17) show.

(16) dat circuit van Zon en Schild is natuurlijk ook uh zwaar **roddelbaar**.
That circuit of Son and Shield is of_course also uh heavily **gossipable**
"'Son and Shield' as a workplace of course lends itself well to gossip" (fn007487.73)

(17) die woonmaatschappij d'r zal voor zorgen dat het dorp ook leefbaar blijft.
that housing_agency there will for take_care that the village also liveable remains
"that housing agency will make sure that the village remains ok to live in" (fv600367.14)

The occurrence of forms like *roddelbaar* 'gossip-able' and *leefbaar* 'live-able' shows that the entity in question can be a complement normally left implicit or marked obliquely. Even if the topic of gossip or the location one lives in are usually not expressed as an argument of the verb, they have to be retrievable in a discourse situation for the verbs to make sense, unless there is a generic interpretation (or if *leven* means 'be alive').

The definition goes on to say that the property expressed consists of said entity's **ability to undergo the event** described in the verb. The notion of undergoing an action entails that this construction will mainly occur with verbs that have a Patient / Undergoer argument in their frame. By stating that the slot filler describes an *event*, we capture the fact that the main slot in the construction has to be filled with a lexical item that is a verb or must be interpreted as a verb. Historically, the affix *–baar* could be used with a nominal stem, with the resulting word meaning "carrying or entailing N". While new words with this meaning can no longer be formed, a few remnants have survived, such as *dankbaar* (literally thank-able, 'grateful'). Example (18), taken from the Tilburg University weekly publication *Univers*, shows that for new formations with a nominal stem, the only interpretation available is when this noun is taken to (under)represent an event:

(18) *Univers* checkt hoe 'busbaar' Tilburg is[12]
Univers checks how 'buss-able' Tilburg is (quotation marks in the original).

[12] Univers d.d. 2 November 2006, p.5.

The word *bus* normally only functions as a noun in Dutch. The header of the article, however, is *Driekwartier bussen* ('three quarters of an hour of bussing'), and clearly uses the noun innovatively as a verb 'traveling by bus'. It is with that meaning of the word *bus* that *busbaar* can be interpreted in the context of the article, which is on the topic of (in)adequate public transport to and from the university campus. The parentheses in the original text indicate that the author was aware that this is a marked use of the construction. Still, it points to the fact that newly created instances of the construction must have an event as the slot filler.

In sum, the V–BAAR construction expresses that the most Patient-like argument of the verb has the property of being able to undergo the event described in the verb. The construction is mainly used to express stable potentiality (with the verb *zijn* 'be'). This potentiality is a characteristic of the Patient (examples 14 and 15) or it is contextually nameworthy to mention that something is (im)possible.

3.2.1.5 Evidence for stored instances of V–BAAR: specific conventionalizations

In novel combinations, the construction is used productively; such environments are the best place to study constructional meaning. Going through our corpus results, we found that many instances of the V–BAAR construction 'felt' familiar. To substantiate this intuition, we discuss four types of independent evidence for claiming that a particular instance is stored. A cautionary note is in order, however: even if a certain type is identified as likely to be stored as a unit, it may still sometimes be produced productively, i.e. the evidence is never direct evidence, and in addition we know very little about variation across individual speakers in this regard.

A first kind of evidence for storage is deviation in *form* from the majority of instantiations of the construction. An example is *vruchtbaar* (lit. fruit-able, 'fertile'). Diachronically, the affix in this word is the same as contemporary –*baar*. Both derive from the Proto-Germanic verb *beran* (to bear, cf. English *fruit bearing*). Over time, the affix has shifted from taking nominal stems and meaning 'carrying or entailing N' to the construction discussed here (Van den Berg, 1974). Synchronically, this link is not transparent to speakers, and the meaning of the form cannot be reduced to its (contemporary) parts. These types were discarded from the analyses after the initial corpus search for all adjectives ending in –*baar*.

Second, there are types in which the meaning is different from what a compositional analysis would yield: *betaalbaar* (lit. pay-able, 'affordable') does not just mean that something can be paid, but that it is reasonably cheap. This additional information may not be too surprising; potentiality or the lack thereof is usually expressed when it is unexpected and we may expect that services and goods can be paid for. This only becomes information worth lexicalizing when it is easier or more difficult than one might think. Still, the specific semantics of this type must be stored in people's memory.

Third, some instances have strong collocations, which cannot be retraced to either the construction in general or to the verb that fills the slot. These collocations are specific to that adjective: they are not collocates of the V–BAAR construction in general, and typically do not occur frequently with the verb outside the construction. An example is the notion of a *maakbare samenleving* (lit. make-able society, 'social engineering'), which is a fixed expression. The verb *maken* is not strongly attracted to *samenleving* and neither is the general V–BAAR template. As collocations are stored combinations of lexical items, the form *maakbaar* in this collocation cannot be produced productively; the combination must be stored *as is*. Note that this is different from the frequent co-occurrence with modifiers like *niet* (not) and *heel* (very), which we also find. They are collocates for the whole construction, as they are found with many different types; in fact, it is likely that the constructions HEEL V-BAAR and NIET V-BAAR are entrenched on their own.

Finally, for some types, it seems that only one of the verb's multiple senses is preferred for the reading of the –*baar* form. In our corpus data, the form *bereikbaar* 'reachable', 'achievable' always refers to its more concrete senses: to be reachable *by phone*, *at certain hours*, *by car* etc. Although the verb *bereiken* is also found to describe goals and aims, this does not occur with adjectival *bereikbaar*, at least not in CGN. Again, this specialization can only be accounted for if we assume the instance is stored.

Taken together, all this type specific behavior (cf. Boas 2003) points at the storage of a large number of instances of the construction. This means that the corpus data, invaluable as they are for arriving at a refined notion of the meaning of the construction, are insufficient to determine whether the construction is productive or not, that is whether it indeed exists as an abstract representation in the minds of speakers. We do now have a much more specific description of the meaning of the construction, which is useful for the design of an experiment testing the psycholinguistic reality of the partially specific construction, that is, of the construction's productivity (Section 3.3).

3.2.2 A syntactic potentiality construction: IS TE V

This section discusses a potential rival to the V-BAAR construction, which differs from it in form but seems to have the same meaning.[13] The IS TE V construction, also sometimes called the Modal Infinitive Potentiality Construction (Boogaart 2006) appears in many different forms, but it always contains a sequence of a subject noun phrase (often the simple inanimate pronoun *het* 'it'), an inflected copula (usually 3rd person singular present tense *is*), the infinitival particle *te* ('to'), and a main verb infinitive.[14] It

[13] There are several other ways to express potentiality in Dutch, e.g. with the auxiliary *can*. Out of these, the IS TE V construction comes closest in meaning to the V-BAAR construction. We limit our analyses here to that construction, which has, as native Dutch-speaking reviewers have pointed out, a few closely related constructions in VALT TE V and BLIJKT TE V, as in

4. hierbij valt te denken aan hulp bij juridische procedures na de asielaanvraag
Hereby falls to think of help with judicial procedures after the asylum_application
"One could think, for instance, of providing aid in the legal procedure after an asylum application" (fn000020.9)

5. in de inrichting bleek de dokter niet te bereiken
In the institution turned_out the doctor not to reach
"It turned out, that the doctor could not be reached in the institution" (fv801096.26)

A complete overview of potentiality constructions in Dutch would certainly include the VALT TE V and BLIJKT TE V constructions, but is beyond the scope of this chapter.

[14] A reviewer pointed out that the construction also occurs prenominally, without the modal verb, as in *een niet te missen kans* 'a not to miss chance', 'a chance that you cannot pass up. A corpus search for sequences of *te* + V_{inf} + N resulted in 311 hits, 73 of which are indeed of this type. Although their meaning is similar to that of the predicative IS TE V construction, there are also a few differences: the majority of the tokens (N = 48) expresses necessity rather than potentiality. Also, quite a few instantiations come from conversations in a specific genre, i.e. political discourse (N = 30). This often results in the expression of events that should take place at some point in the future, as in (6).

6. en ik heb een vraag aan de staatssecretaris die handelt over de te vernieuwen richtlijn
and I have a question for the secretary_of_state which manages about to renew guideline
"and I have a question for the secretary of state about the guideline that has to be renewed" (fn000243.52)

For V-BAAR the proportion of attributive to predicative instances is approximately 3:1, whereas for IS TE V, this is 1:16. These are clearly very different distributions. We restrict all further analyses to predicative instances of both constructions.

conveys the meaning that the action named by the verb can potentially be carried out with or to the Patient-like argument named by the subject. It has, therefore, both modal ('possibility') and passive ('can be done') semantics, while syntactically it is a copula construction. The subject noun phrase denotes what is semantically the patient of the main verb, explaining the passive meaning. An example appears in (19).

(19) de expositie is nog te bezichtigen tot zeven mei
the exposition is still to see until seven May
"the exposition can be seen until the 7th of May"(fv600185.6)

Diepeveen et al. (2006:52-55) carried out a preliminary analysis of this construction in the CGN and found that it is used in both Belgium and Holland, but with a very high frequency in Belgium of four lexicalized cases, with the verbs *zien* 'see', *zeggen* 'say', *hopen* 'hope' and *doen* (*het is te zien* 'that remains to be seen'; *het is te zeggen*, 'that is; at least'; *het is te hopen*, 'hopefully'; *het is niet te doen* 'it's impossible'), of which only the third is also in frequent use in Holland. We take this analysis further in the present article, focusing more on the instantiations with low token frequency in order to investigate how productive the construction is. Similar to the analysis of the V–BAAR construction, a corpus analysis was conducted in order to see what the range of verbs was that occurred in this construction, and to ascertain whether the construction really does the mean the same thing as V-BAAR.

3.2.2.1 Identifying instantiations of the IS TE V construction

Ideally, the dataset should be the result of a search for all forms in the Corpus of Spoken Dutch (CGN) that make use of the template [Copula + *te* + Infinitive]. However, identifying all instances of the IS TE V construction in a corpus is not as simple as it is for V-BAAR. First of all, apart from *te*, none of the word forms in the construction are constant, so that searches cannot make use of a simple lexical query. Use can be made of Part-of-Speech tags, but this creates additional problems. As is typical for syntactic constructions, certainly ones that use common words like copular *is* and the infinitival particle *te*, there are many word sequences that look superficially like the IS TE V construction, but are really instantiations of other constructions. While a search for words with the –*baar* suffix yields very few false hits, there are many in the raw data one obtains after a search for all IS TE V sequences.

We searched the CGN for all occurrences of the sequence [*is* + *te* + Infinitive], with zero, one or two words intervening between *is* and *te* (searches with more than two words intervening yielded an unwieldy number of false hits) and other possible forms (past tense, non-3rd person singular, and subordinate clauses, in which Dutch has a word order that differs from main clauses), and manually weeded out the false hits. While certainly not all instances of the construction will have been found this way, we have good reason to think we did capture most of them.

3.2.2.2 Main search results for the IS TE V construction

Table 3.7 lists the verbs most frequently found in the infinitive position of the construction in the CGN.

Table 3.7: 10 most frequent verbs in IS TE V construction found in CGN

Type	Frequency	Most common surface form	Meaning
doen ‘do’	220	is niet te doen	is hard to do
zien ‘see’	162	is te zien	can be seen (in)
hopen ‘ hope’	79	is te hopen	hopefully
zeggen ‘say’	67	is te zeggen	that is, ...
vinden ‘find’	44	is niet te vinden	cannot be found
geloven ‘believe’	26	is niet te geloven	unbelievable
vergelijken ‘compare’	25	is niet te vergelijken met	cannot be compared to
krijgen ‘get’	20	is niet meer te krijgen	cannot be found anymore
merken ‘notice’	17	is weinig van te merken	is hard to notice
spreken ‘speak’	15	is niet te spreken	does not like it

The corpus search in the CGN turned up 211 different verbs in the infinitive slot, producing a total of 1108 instances, figures that are not too different from what we found for the predicative construction with V-BAAR (171 types and 1282 tokens).

In general, the distribution shows some similarities to the one observed for V-BAAR. First, there are a few clearly favoured lexemes. Again, the ten most frequent verbs account for more than half of all tokens. As for V-BAAR, there is a clear frontrunner (*doen* 'do'), a clear number two (*zien* 'see'), and a small set of other verbs that occur frequently in the construction. Second, some semantic specialization is evident for the most frequent members, as the last column in Table 3.7 illustrates. The concept lexicalized by 'hopefully' is a few steps removed from 'can be hoped for', and the pragmatic force of 'unbelievable' is much stronger than a compositional 'cannot be believed' would convey. Third, many of the types tend to co-occur with the negation *niet*. Negation also occurs frequently with words ending in –*baar*; they often contain the negation prefix *on*–. Some verbs form collocations with other adverbial modifiers, as illustrated in the third column of Table 3.7. Finally, many verbs are used only once or twice in the construction in the corpus, suggesting the template is used productively.

The most commonly found verb, *doen* 'do', is, of course, also a very frequent verb overall. Presumably, the verbs that tell us most about the construction are those verbs that are most attracted to this construction. In order to find out for which verbs this holds, we performed a collostructional analysis.

3.2.2.3 Collostructional analysis of IS TE V

The collostructional analysis was carried out in exactly the same way as the V–BAAR construction described in subsection 3.2.1.3. Table 3.8 lists the twenty most strongly attracted verbs to the IS TE V construction. Recall that for the V-BAAR construction the list of most frequently found verbs in the construction and the list of verbs most attracted to it as revealed by collostructional analysis partially overlapped but also showed some interesting differences. The same conclusion applies to the data for IS TE V. Of the ten verbs most frequently found in IS TE V, four do not appear among the twenty verbs attracted to the construction the most: *zeggen* 'say', *vinden* 'find', *spreken* 'talk to' and *krijgen* 'get'. All of these are verbs with very basic meaning, which ensures a very high token frequency overall in any corpus.

For high-frequency verbs, it is difficult to be highly attracted to *any* construction, simply because it occurs a lot in many different constructions (cf. Gries, Hampe & Schönefeld 2005:650). Be that as it may, the list of verbs most attracted to the construction reveals some interesting characteristics.

Table 3.8: Collexemes of IS TE V (values given for collostructional strength correspond to –log (p-$_{\text{valueFisher-Exact}}$, 10))

Collexeme (N)	Translation Verb	in construction	Collostructional strength
stukkrijgen (1)	Demolish	indestructible	Infinite
toerekenen (1)	Ascribe	can be ascribed to	Infinite
hopen (79)	Hope	hopefully	100.463
zien (162)	See	can be seen, found, etc.	84.739
doen (192)	Do	can be done	74.859
vergelijken (25)	Compare	should be compared	32.367
wijten (11)	Blame	should be blamed on	21.822
achterhalen (9)	Get a hold of	can be retrieved from	15.083
geloven (26)	Believe	unbelievable	13.869
verklaren (11)	Explain	can be explained	13.401
merken (17)	Notice	be obvious	12.595
terugvinden (8)	Retrieve	can be retrieved in	12.435
combineren (8)	Combine	can be combined with	10.261
verstaan (10)	Understand	be loud/clear enough	9.714
verwachten (13)	Expect	is to be expected	8.992
herkennen (9)	Recognize	can be recognized by	8.620
bereiken (8)	Reach	can be reached	6.118
bekennen (4)	See	(nowhere) to be found	5.674
aanbevelen (3)	Recommend	is to be recommended	5.364
aanmerken (2)	Complain	can(not) be complained about	5.197

First, native speakers of Dutch will recognize the familiar mulit-word chunks that underlie the high position of some of these verbs, e.g. *is niet stuk te krijgen*, literally 'cannot be broken', i.e. 'is still going strong'; *is te wijten aan X* 'X is to be blamed for'; or *is niet meer te achterhalen*, literally 'cannot be overtaken anymore', i.e. 'cannot be verified anymore' (said in a figurative sense about a particular piece of information). In general, many combinations containing a particular verb and a particular degree adverb are lexicalized, in the sense that they recur in the corpus and may thus be considered conventional chunks.

However, what the verb ranking allows us to do most of all is get a good grasp of the constructional meaning (cf. Gries et al. 2005:650-654). In order to do that, we must examine the ranking and try to understand why these verbs are so attracted to the potentiality construction. What is it about *stukkrijgen* 'demolish', *toerekenen* 'ascribe', *vergelijken* 'compare', etc. that makes expression of potentiality (or lack of it) so nameworthy? The next subsection is an attempt to answer this question by narrowing down the construction's meaning as precisely as possible, based on the verbs most saliently attracted to it.

3.2.2.4 Constructional meaning of IS TE V

The list of twenty most attracted verbs does not clearly form a single natural category. This is in contrast with, for example, Gries et al.'s (2005) analysis of the English *as*-predicative (e.g. *Does he regard that as a serious problem?*) and probably reflects the wide semantic scope of potentiality constructions. However, some generalizations can be made on the basis of the list.

Perhaps the most striking feature of the construction is that it often encodes the speaker's *assessment* of potentiality. It is not potentiality as such that is conveyed, but whether the speaker thinks a certain action or event *is likely* to happen. In the list of twenty, this holds for *stukkrijgen* 'demolish', *doen* 'do', *vergelijken* 'compare', *wijten* 'blame', *geloven* 'believe', *merken* 'notice', *combineren* 'combine', *verwachten* 'expect', *bekennen* 'see', *aanbevelen* 'recommend', and *aanmerken* 'complain'. That is not to say that every instantiation of these verbs in this construction will have this semantic nuance, but often it is the case, as becomes clear once one starts investigating the actual instantiations in the corpus. While it is trivially true that most verbs refer to an action that can happen, the IS TE V construction makes this normally unimportant aspect salient. Given the vagueness of potentiality, saying that something could possibly happen implies making the assessment that it is worthwhile to even mention it. In other words, the existence of the construction reflects the fact that it makes sense to make a comment about something potentially happening if it is not obvious that it will. Presumably this is also the reason why the construction is often used in its negative form: saying that something is not possible will often be more noteworthy than saying that it is. The construction is, therefore, often used as a stance marker: the speaker thinks that an activity is likely or unlikely to be carried out. In (20), the speaker does not just state that comparing Holland and Germany is something that cannot be done: the message is that as far as the speaker is concerned, they are incomparable. From there, it is a small step to necessity rather than mere potentiality: it is not just that the action *can* be carried out: it *should* be carried out, and asserting this is the reason for constructing the utterance in the first place. The pragmatic impact of (21) is not that the illegal repairs can be blamed, but that they should be blamed for the fire. Boogaart (2006:43) also notes that the construction is sometimes used to encode necessity; we suggest here that the two semantic categories actually shade off into each other.

(20) maar Duitsland is niet te vergelijken met Nederland hoor, absoluut niet.
but Germany is not to compare with Netherlands right, absolutely not
"but you cannot compare Germany with Holland, you know, absolutely not" (fn007569.162)

(21) de brand in de skitreintunnel in Kaprun is vermoedelijk te wijten aan illegale reparaties.
The fire in the ski_train_tunnel in Kaprun is probably to blame to illegal repairs
"the fire in the ski train tunnel at Kaprun should probably be blamed on illegal repairs" (fn002178.1)

It is, therefore, a personal judgment of the degree of potentiality rather than the assertion of absolute potentiality that is conveyed by the construction. This is often reinforced by the adverbs of degree that tend to form an essential part of the conventional chunk that frequent recurrence of the instantiation has forged. Many examples seem to be motivated by a desire on the part of the speaker to counter any expectations his audience may have: the point is, again, not to say that something is possible, but that it is more (or less) possible than the hearer might think (see examples 22 and 23).

(22) dat is niet te geloven
that is not to believe
"That's hard to believe" (fn000820.311)

(23) Ja 't is moeilijk te controleren natuurlijk hè
yes it is difficult to check of.course isn't.it
"Yes, that's of course hard to check, isn't it" (fn007057.39)

A second feature, often overlapping with the previous one, is that the event denoted by the verb is carried out mentally rather than physically. Combined with the stance character of the construction, and its tendency to be used to mark necessity rather than potentiality, it means that many examples fall within in the domain of human interaction, particularly interaction involving persuasion. Among the twenty most attracted verbs, many name a type of mental activity, e.g. *toerekenen* 'ascribe', *hopen* 'hope', and *aanbevelen* 'recommend'. Again, it is certainly not the case that all these verbs are always used for mental activities, but most of the time they are. For instance, 'expect' and 'wish' in (24) reflect mental activity.

(24) dat bij deze koerswijziging zijn humor hem verlaten zal is noch te verwachten noch te wensen.
that with this change.of.course his humor him leave shall is not to expect not to wish
"that with this change of course his sense of humor will leave him should neither be expected nor to be wished for" (fn001288.71)

A third group of verbs considerably attracted to this construction denote sensory activity. This includes *zien* 'see', *verstaan* 'understand', *bekennen* 'see' and *aanmerken* 'complain' in the list above. Obviously, these denote some type of mental activity as well, and when used in the IS TE V construction they often convey that, perhaps with a little effort, the addressee or some third party can see or hear the thing referred to by the subject. Typical examples appear in (25) and (26).

(25). er is geen spoor van make-up op haar gezicht te bekennen.
There is no trace of make-up on her face to spot
"You cannot see a trace of make-up on her face" (fn001202.40)

(26). het is moeilijk te verstaan hè
It is hard to hear/understand, right? (fv400121.294)

Finally, it should be noted that virtually all verbs found in the construction are transitive, and those that are not make salient reference to some Patient-like element, such as a person spoken about (not marked as the direct object of the verb *spreken* 'speak' in an active sentence, but contained in a prepositional phrase with *over* 'about'). This suggests that the construction tends to be used to comment not so much on the potentiality of the event, but on the potentiality of the entity named by the (sometimes unexpressed) Patient argument of that verb to undergo the event. The IS TE V construction contains a copular predicate, used to comment on an entity that it 'can be V-ed', and, like in a passive, that entity surfaces as the subject of the construction.

To conclude, the meaning of the construction can be given as

In the eyes of the speaker, the most Patient-like argument of the verb has the property of being able to undergo the event described by the verb. In particular, the construction is often used as a stance marker.

3.2.2.5 Evidence for stored instances for IS TE V: lexicalization

Some allusions have already been made to the fact that many instantiations of the construction have a very familiar ring, and often these chunks include adverbial modifiers. Recall that for the V-BAAR construction, the same was found. For many of these chunks, recurrence in the corpus is one fairly reliable source of evidence, but even many combinations that occur only once in the CGN seem conventional, at least according to our intuitions.

Given the tendency of the construction to encode relative rather than absolute potentiality, it is no surprise that the adverbial modifiers tend to be degree adverbs, such as *goed* 'good', *best* 'well', *makkelijk* 'easy', and *moeilijk* 'difficult', with the most frequently used modifier of all, *niet* 'not' as the limiting case of degree indication: combined with the IS TE V construction it conveys that something is impossible.

Lexicalization is also obvious in cases where the meaning of the instantiation is not what you would get if the meaning of the verb and that of the construction are compositionally combined. Not surprisingly, this holds especially for the most frequent cases, such as the Flemish idioms *het is te zien* 'it depends', literally 'it is to see' and *het is te zeggen* 'that is to say' (Boogaart 2006: 40). Other examples of such idiomaticity include *is niet te filmen* ('it's unbelievable'; literally 'cannot be filmed') and *is ver te zoeken* ('cannot be found', 'does not exist'; literally 'can be searched for far away', cf. English *you can search high and low*).

In conclusion, as was the case for the V-BAAR construction, the corpus data for the IS TE V construction do not provide conclusive evidence to answer the question whether the construction is really productive. Some instantiations may certainly have been coined by the speaker on the spot, and speakers of Dutch are certainly able to create new forms, but both the quantitative evidence and our intuitions about many of these instantiations suggest that the great majority of attested instantiations were taken whole from memory. This raises the question what the psychologically most realistic description of the construction is: do people mainly 'have' a long list of fixed, or 'specific', expressions, such as *is niet te filmen* (is not to film; i.e. 'unbelievable'), *is te bezichtigen* (is to view; i.e. 'can be viewed'), and *is voor te stellen* (is for to set; i.e. 'can be imagined'), or do they also have the general schematic schema [COP + *te* + INF], as well as perhaps a variety of partially schematic subschemas (cf. Langacker 1991:410)? This is a question that requires further research.

Now that both constructions have been described in some detail, we are ready to compare them. Are the V-BAAR construction and the IS TE V construction synonyms? This question is dealt with in the final subsection.

3.2.3 Comparison of corpus data V–BAAR and IS TE V

The distinctive collexeme analysis (Gries & Stefanowitsch, 2004) was designed to compare distributional patterns for two constructions. It is similar to simple collostructional analysis in that it contrasts expected frequencies with observed frequencies, on a null hypothesis that no differences are found. Whereas in the previous analyses the occurrence of verbs in the construction vis-à-vis their frequency in the whole corpus was tested, the distinctive collexeme analysis provides a contrast between the occurrences of lexical items in two constructions. We use this analytical procedure to identify which verbs have a strong preference for either the IS TE V or the V-BAAR construction. Since only the token counts for each verb in both constructions and the total number of instantiations of both constructions are incorporated in the analysis, and overall frequency of the verbs is not considered, the list of preferred verbs for each construction is likely to differ from the results of the collostructional analysis of each construction. This is because the distinctive analysis serves a different purpose: while the collostructional analysis aids in developing a fine-grained semantic analysis for each construction, the distinctive analysis looks more directly for differences between them. The ten verbs that have the strongest preference for the IS TE V or the V-BAAR construction are listed in Table 3.9.

Table 3.9: Results distinctive collexeme analysis (values given for distinctiveness correspond to –log (p-$_{valueFisher-Exact}$, 10))

Verb	Translation	V-baar	is te V	Preference	Distinctiveness
Beschikken	Have at one's disposal	141	1	baar	37.937
bereiken	Reach	68	8	baar	11.005
aanvaarden	Accept	38	0	baar	10.391
halen	Get	72	11	baar	10.143
voorstellen	Suggest	52	4	baar	9.975
bespreken	Discuss	36	0	baar	9.839
denken	Think	45	4	baar	8.266
vertrouwen	Trust	37	2	baar	7.874
kennen	Know	28	0	baar	7.634
kwetsen	Hurt	28	0	baar	7.634
zien	See	0	162	Is te	57.040
doen	Do	18	192	Is te	46.855
hopen	Hope	0	79	Is te	27.046
zeggen	Say	0	67	Is te	22.847
vinden	Find, locate	3	44	Is te	11.446
geloven	Believe	0	26	Is te	8.749
krijgen	Get, receive	0	20	Is te	6.717
spreken	Speak	0	15	Is te	5.030
verwachten	Expect	0	13	Is te	4.357
wijten	Blame	0	11	Is te	3.684

Looking at the characteristics of the verbs that significantly prefer one of the constructions over the other (44 for the V–BAAR construction, 25 for IS TE V), some interesting differences appear, but also a lot of similarities. All verbs listed have two (or three) lexical arguments, i.e. can be used in the transitive construction. Semantically, the Patient argument can be a physical object (*zien* 'see', *bereiken* 'reach') or a mental entity (*hopen* 'hope', *voorstellen* 'suggest') for both constructions. These observations illustrate that both IS TE V and V-BAAR occur with a broad range of verbs.

There are also a few interesting differences, one of which had escaped our analyses so far: a majority of the verbs that have a significant preference for the IS TE V construction can take a sentential complement as its object argument in a main declarative sentence (from the top 10, no fewer than 7: *zien, doen, hopen, zeggen, vinden, geloven, verwachten*).[15] This is not a semantic criterion, although in fact a lot of mental activity verbs occur with such complements, and it may be related to the syntactic form of the construction. A sentential complement X is apparently easier to incorporate in the frame *het is te V dat X* 'it is to V that X' than in the frame *het is V-baar dat X* 'it is V-able that X'. We have, at present, no clear idea about why this would be the case.

There are also some numerical differences: more verbs are significantly preferred by V–BAAR (44) than by IS TE V (25), where the overall number of verbs was slightly higher for IS TE V than V–BAAR (209 and 171 respectively). Verbs that prefer IS TE V tend not to occur at all with V–BAAR, but the reverse is not true. In total, 55 verbs occur in both constructions in our corpus, and only eleven of those at least five times in each construction. One possible interpretation of this difference is that IS TE V has a slightly broader range than V–BAAR. Note, however, that all attributively used forms of *–baar* were excluded prior to the analyses. Had we not done that, the picture might have been different: CGN contains 3908 instantiations of *–baar*, with some of the most frequent forms never occurring in predicative contexts.

In comparison with the lists given earlier for each construction separately, the results of the distinctive analysis resemble those of the raw frequency data more than those of the collostructional analysis. This is due to the fact that the overall frequency of the verb in the corpus does not play a role in the distinctive collexeme analysis. As a result, the very general and frequent verbs that fell out of the collostructional analysis for IS TE V (e.g. *zeggen* 'say', *vinden* 'find') are included in the list of preferred verbs. It is interesting to see that these basic verbs are very rarely found with the V–BAAR construction. This again may point to a smaller range of productivity for V–BAAR compared with IS TE V.

Finally, a closer look at some of the verbs that are found with both constructions shows that these can be divided into two groups: for some verbs, there is a clear difference in meaning between the IS TE V form and the V-BAAR form, whereas others are seemingly synonymous. *Eten* 'eat' is an

[15] Strictly speaking, this does not hold for *doen* 'do': it does not occur with subordinate clauses introduced by *dat* 'that'; however, it does co-occur with subordinate clauses introduced by a different complementizer, *om te* '(in order) to'.

example of the former category: *eetbaar* means 'edible, not poisonous' and *is te eten* 'not bad, tasty'. This differentiation ties in well with the overall inclination of *assessed* potentiality for the IS TE V construction: it is the speaker's opinion about the flavour of the food that is transmitted. With *eetbaar*, we are dealing with factual potentiality: it is possible to eat X without getting sick. Not all verbs show such specialization: we see no difference in meaning between *vergelijkbaar* 'comparable' and *is te vergelijken*. The semantics of the verb already contains some degree of assessment: it is always the speaker who decides that two things can or cannot be compared. Possibly, for these verbs, it is the sentential context and other discourse-related factors that determine which construction is actually used.

In sum, the verbs that show strong preferences cover a broad range of transitive verbs for both constructions, with some aspects of the distribution pointing at a slightly smaller range for the V–BAAR construction. Verbs that take a sentential object tend to prefer the IS TE V construction. When verbs do occur in both constructions, in some cases there is semantic specialization, i.e. the two instantiations do not mean the same thing.

3.3 Experiment: the Magnitude Estimation task

The corpus data discussed in the previous sections of this article provide ample evidence that the V–BAAR construction and the IS TE V construction occur frequently in present day Dutch. The instantiations also turn out to fit semantically with the generalized template. In other words: there are many forms that look like they are instantiations of the constructions. The problem is, however, that for natural production data it is impossible to determine whether these sequences were productively 'built' using the template. It is more than likely that people have many forms available as units in their mental lexicons, as was argued in subsections 3.2.1.5 and 3.2.2.5. With natural language data, therefore, one cannot answer the question whether speakers of Dutch have a partially specific representation or template for the constructions at hand, although the corpus data make this a likely proposition.

In order to find out more about the knowledge Dutch speakers have concerning the types of verbs preferred in the constructions, an experiment was designed. In this experiment, the acceptability of various instantiations of both constructions is tested. The experiment focuses on two aspects of the construction: the participant structure of the verbs and the notion of *assessment of potentiality*. With regard to the former, the CGN-data show that verbs with certain participant structures (i.e. transitive verbs, with Agent and Patient participant roles) are much more likely to occur in the constructions than others. The *assessment* of potentiality especially came to the fore in the data analysis of the IS TE V construction.

If people make a consistent distinction in the acceptability of instantiations that can be traced back to the participant structure of the verb they contain, this indicates that they have a representation for these constructions that includes information about the types of verbs with which they can be combined. Earlier, a construction's productivity was defined as the existence of (partially) abstract representations in speakers' constructicons, that they can use to make and interpret novel instantiations of that construction. In our operationalization of this mental representation, we now argue that a consistent distinction in acceptability of forms that corresponds to certain characteristics of these forms is evidence that the mental representation actually exists and is used by speakers. The next section introduces the methodology and items used in the test, as well as the participants and the procedure.

3.3.1 Method

3.3.1.1 Task

The experiment conducted is a *magnitude estimation task* (Ryan, 1973; Sorace, 1996; Bard, Robertson & Sorace, 1996; Sorace & Keller, 2005; Wulff, 2009). This is a specific kind of judgment task, in which participants make up their own acceptability scale. In contrast to other grammaticality judgment tasks, participants are told to assign a random number to a first stimulus, and then grade subsequent stimuli relative to the previous one. This offers various advantages: participants are free to distinguish as many different categories or grades of acceptability as they like, whereas grammaticality judgment either limits these to two (yes-no) or a fixed number (when a Likert-scale is offered). When a Likert-scale is used, usually with a range of 5 or 7 options, responses in the middle of the scale are notoriously difficult to interpret: participants may be unsure about their response and therefore opt for the middle of the range, or they may be very certain that this stimulus belongs in the middle. Both problems are avoided with the magnitude estimation (ME) task.

ME was first applied to linguistic stimuli by Ryan (1973), who tested attitudes towards accented speech with this measure after it had already been used successfully to gauge reactions to stimuli in physical domains, such as people's perceptions of the loudness of sound or the intensity of pain (Stevens, 1957). Such stimuli can be measured along an objective scale intensity (decibel etc.), which does not exist for the acceptability of linguistic stimuli. The method can be used, however, to determine whether people consistently distinguish between types of stimuli theoretically postulated to be different, in our case between verbs with different thematic and grammatical role assignments, and instantiations that either do or do not reflect an *assessment* of potentiality (cf. Sprouse *in press* for an overview of linguistic experiments with ME and a discussion of its merits).

In the current experiment, two groups of participants are distinguished: adults (mean age 43) and children (mean age 12;3). To our knowledge, magnitude estimation tasks have not been done before with children. There is evidence that children in primary school are able to perform grammaticality (cf. McDonald, 2008 for an overview) and acceptability judgment tasks (Sorace et al., 2009). The experiments discussed in Chapter 2 only involved children. Children's performance was evaluated against a more or less implicit norm. For the word formation and word definition task (Sections 2.2 and 2.3 respectively), this meant correct responses, and for the lexical decision task (Section 2.4), a comparison was made with an earlier experiment, performed by adults (De Jong et al., 2000). For the present experiment, a group of adult participants was added to the research design, in order to obtain a measure of adult native-speaker knowledge.

3.3.1.2 Test items

Two versions of the experiment were constructed. These differed in the number of items: adults rated 84 items (48 test items and 36 fillers) and children rated 64 items (36 test items and 28 filler items). The children's set of items was a subset of the adult version. The number of test items differed, because we feared the full set of 84 items would be too taxing for children's concentration span.

Test items contained either an instance of the IS TE V construction or V–BAAR (always used predicatively). Filler items were made up from other ways to express potentiality, such as 'it is possible that', or 'X can be V-ed'. In the test items, five categories of verbs were used, with four verbs in each category (three in the children's version) and all verbs occurring once with V–BAAR and once with IS TE V. Almost none of these verbs appeared with either of the constructions in the corpus data; the only exceptions are *combineren* 'combine' and *breken* 'break'. *Breken* only occurs with *–baar* (5 tokens in CGN); *combineren* is found twice with *–baar* and 8 times with *is te.* This, combined with the fact that we constructed all of the test sentences, makes it likely that the great majority of test items constituted novel combinations for the participants. In addition, for one category of verbs, test items were construed with either a clear *assessment* element or factual potentiality. Figure 3.1 reflects the distribution of test items.

Construction type / Verb category	V-baar (N)		is te V (N)		Filler items (N)	
	adults	children	adults	children	adults	children
1 Assessment	4	3	4	3	8	4
Factual	4	3	4	3		
2	4	3	4	3	8	6
3	4	3	4	3	8	6
4	4	3	4	3	8	6
5	4	3	4	3	4	6
Total	24	18	24	18	36 / 84	28 / 64

Figure 3.1: Distribution of test items Magnitude Estimation

Verb categories

The verb categories are described below, listing the verbs used and one example test item for each category. For a complete list of all test items, see Appendix 4.

Prototypical transitive verbs, with Agent-Subject and Patient-Object

These verbs are typically found with a human Agent in the subject position and a Patient affected by the verb (*schrijven* 'write', *schilderen* 'paint', *combineren* 'combine', *maaien* 'mow').

(27) Deze gladde muur is eenvoudig schilderbaar met een roller
This smooth wall is easily paint-able with a roller

Optionally transitive verbs with Agent-Subject and Patient-Object or Undergoer-Subject roles

These verbs are found both in highly transitive utterances with an Agent and a strongly affected Patient and in sentences with only one argument expressed: the Undergoer[16] (*smelten* 'melt', *scheuren* 'tear', *breken* 'break', *drogen* 'dry').

(28) Chocolade is makkelijk te smelten in de magnetron
Chocolate is easy to melt in the microwave

Optionally transitive verbs with a (generally) implicit second argument

These verbs usually occur with only one explicit argument: an Agent-Subject (*schreeuwen* 'yell', *wandelen* 'stroll', *zingen* 'sing', *roken* 'smoke'). They can only be understood, however, with reference to a semantic base which includes a second argument: a motion verb like *wandelen* 'stroll' always presupposes a path and a sound emission verb such as *schreeuwen* 'yell' makes reference to a sound/message. The inclusion of this category allows us to differentiate between verbs which require one expressed argument in de active declarative (categories II and III) or two expressed arguments (in categories I, IV and V).

(29) Sigaren moeten eerst rijpen voordat ze rookbaar zijn
Sigars must first ripe before they smoke-able are
"Sigars must ripen before they can be smoked"

Non-prototypical transitive verbs, with Stimulus-Experiencer roles and BE-perfect

These verbs require two explicit participants, filling Subject and Object position, but thematically they differ from the verbs in I: the subject is usually filled by a Stimulus, and the object position by an Experiencer (*lukken* 'succeed', *meevallen* 'turn out better than expected', *ontgaan* 'escape', 'fail to grasp', *ontschieten* 'elude').[17] The perfect tense is formed with the

[16] Compare transitive use *ik smelt chocolade altijd 'au bain marie'* 'I always melt chocolate in a water bath' and intransitive use *chocolade smelt op 36 graden* 'chocolate melts at 36 degrees'.

[17] Although Dutch and English are related languages, the Dutch verbs in this category do not have a near-equivalent in English with a similar argument structure. For example, the English equivalent of *lukken* is *to succeed*. While the Dutch verb prototypically occurs in the Stimulus-Experiencer construction '*het lukte me*' (literally: 'it succeeded me'), English construes it as an active clause with an Agent: '*I succeeded*'.

auxiliary *zijn* 'be' (e.g. *het is me gelukt* 'it is me succeeded', 'I have managed to do it').

(30) Het is een ambitieus plan, maar als iedereen helpt is het zeker lukbaar.
It is an ambitious plan, but if everyone helps is it surely succeed-able

Non-prototypical transitive verbs, with Stimulus-Experiencer roles and HAVE-perfect

Similar to category IV verbs, these occur with a Stimulus-Subject and an Experiencer-Object (*fascineren* 'fascinate', *verbazen* 'amaze', *afschrikken* 'deter', *verrassen* 'surprise'). The distinction between verbs in IV and V is that for category V different construals are available: in addition to perfect tense with the auxiliary *hebben* 'have', and its transitive construal, forms with the auxiliary *zijn* 'be' are also marginally acceptable, and these have a passive interpretation (*het heeft me gefascineerd* 'it has me fascinated', 'It has fascinated me' versus *ik ben gefascineerd* 'I am fascinated').

(31) Ongewenst bezoek is af te schrikken met een alarmsignaal
unwanted visit is off to fright with an alarm_signal
"Unwanted visitors can be deterred with an alarm signal"

Assessment and factual potentiality

Based on our corpus data, our intuition is that the V- BAAR construction is used more to express 'factual' potentiality, whereas the IS TE V construction often reflects the speaker's stance or assessment of potentiality. In order to investigate whether these semantic nuances are part of the participants' mental representations of the constructions, sentences were included with either an explicit stance-marking element (e.g. *denk ik*, 'I think', example 32) or where the potentiality was clearly determined externally (e.g. based on rules or laws, example 33). For practical reasons (the test could not become too long) this was only done for each of the four verbs in the first category.

(31) Pannenkoeken met stroop zijn goed te combineren met een glas koude melk, vind ik. (Assessment)
Pancakes with syrup are good to combine with a glass cold milk, find I
"Pancakes with syrup go well with a glass of cold milk, I think"

(32) Deze korting is niet te combineren met andere aanbiedingen (Factual)
This reduction is not to combine with other offers
"You cannot combine this discount with other special offers"

3.3.1.3 Participants and procedure

The adult participants were 72 native speakers of Dutch (aged between 18 and 82, mean age 43). All were born in the Netherlands and presently living there. They participated on a voluntary basis. The 148 participating children were pupils in sixth grade ('groep acht' in the Dutch school system), mean age 12;3 years.[18] They came from three primary schools in Tilburg, a large city in the south of the Netherlands. All children participated on a voluntary basis, with consent given by their parents. About one quarter of the participating children (N =38) reported speaking another language than Dutch on a regular basis. Table 3.10 illustrates the distribution of the pupils over the three schools.

Table 3.10: Child participants magnitude estimation task per school

School	Nr of children	Monolingual	Bilingual	Mean age
1	106	98	8	12;3
2	25	11	14	12;4
3	17	1	16	12;5
Total	148	110	38	12;3

The adult participants received an email with a link to a website. Here they could log in and, after providing some personal information, participate in the experiment at a time convenient for them. A short introduction informed them that this experiment was set up to find out what kinds of sentences people rate as 'better' or 'worse', where 'bad' sentences were defined as unintelligible or ungrammatical –thus pointing at both form and meaning. The participants were told that they were to rate each sentence on how 'good' it seemed to them, with higher numbers reflecting better sentences.

[18] The sample size for the children is substantially larger than for adults. This is thanks to the cooperation of three different schools, one of which is one of the country's largest primary schools.

They were then introduced to the experimental technique with a number of examples, first with non-linguistic stimuli (oddly-shaped objects whose size they had to grade relative to one another), then with sentences in which one or more word order restrictions were violated. Before starting they were told that the sentences in the experiment all contained information about something that was *possible*. They were instructed to read each sentence carefully, but not think too long about the number they filled in. From the logged data, it was clear that participants spent between 12 and 22 minutes online for the whole task. Test items appeared in a random order, different for each participant.

The children did the experiment at school. The participating schools had a dedicated classroom equipped with enough computers to let all pupils in one class do the experiment simultaneously. The children were told that they would participate in an experiment designed to find out what kinds of sentences are 'better' or 'worse' by asking this question to children. For that reason, it was explained, they would have to rate a number of sentences, with higher numbers for better sentences, and lower numbers for worse sentences. They too were introduced to the technique by practicing assigning numbers to differently sized objects first, and then given a practice trial with sentences containing word order violations. The instructions for the task were slightly modified in comparison with the adult version, in order to make sure that the children understood what they were asked to do. The modifications only included simplifications in sentence structure and word choice. An example is the first sentence of the introduction. For children, this reads "*Als je leest, merk je dat sommige zinnen beter zijn dan andere*", 'when you read, you notice that some sentences are better than others'. Adults read "*Bij het lezen van een tekst valt soms op dat sommige zinnen 'beter' zijn dan andere",* 'When reading a text, it some sentences seem noticeably 'better' than others'. Each child worked at an individual computer, taking approximately 20 minutes to complete the task. The researcher was present in the classroom.[19]

[19] She observed that not all children were focused on the task: one of them simply put down '999' for most sentences, for example. In the analyses, these responses did not influence the findings to disturb the overall picture described below. Because it would be a subjective procedure to remove some participants' data from the analyses, they were left intact. One reviewer notices that it would be interesting to do further analyses on the data to see if certain groups of participants behave similarly. This falls outside of the scope of the present work, but would certainly be valuable. The MultiRep model can accommodate the existence of different schematic representations for different language users.

3.3.2 Results

For each participant, scores were recoded into Z-scores to make a direct comparison of individual scores possible. Although participants reported feeling unsure about being consistent in their grading, the test was found to be reliable (Cronbach's α = .85 for adults, α = .93 for children). Each participant responded to all test items, making this a within-subject design.

Effects of construction type and verb category

We applied a GLM repeated measures analysis for each of the two participants groups, with verb category (5 levels) and construction type (2 levels) as within-participant factors. The results of this test are represented visually in Figure 3.2 (adults) and 3.3 (children), with mean Z-scores and standard deviation in Tables 3.11 and 3.12.

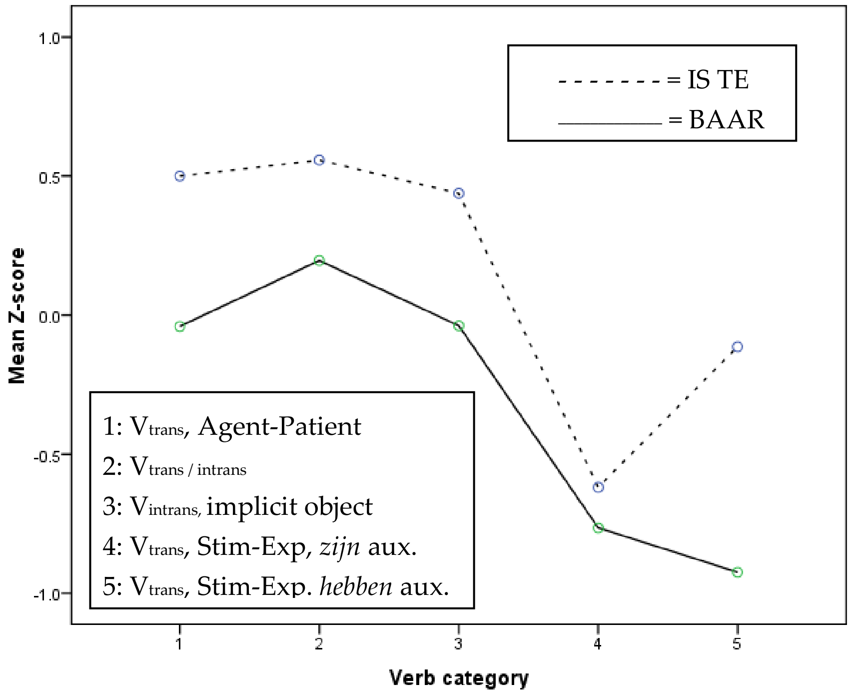


Figure 3.2 Mean Z-scores per item category for adults

Table 3.11: Mean Z-score per verb category and construction type for adults

	is te V (sd)	V- baar
I. V_{trans}, Agent-Patient	.500 (.343)	-.041 (.315)
II. $V_{trans\,/\,intrans}$	.557 (.416)	.196 (.392)
III. $V_{intrans,}$ implicit object	.438 (.406)	-.038 (.521)
IV. V_{trans}, Stim-Exp, *zijn* aux.	-.620 (.385)	-.765 (.477)
V. V_{trans}, Stim-Exp. *hebben* aux.	-.114 (.253)	-.925 (.489)

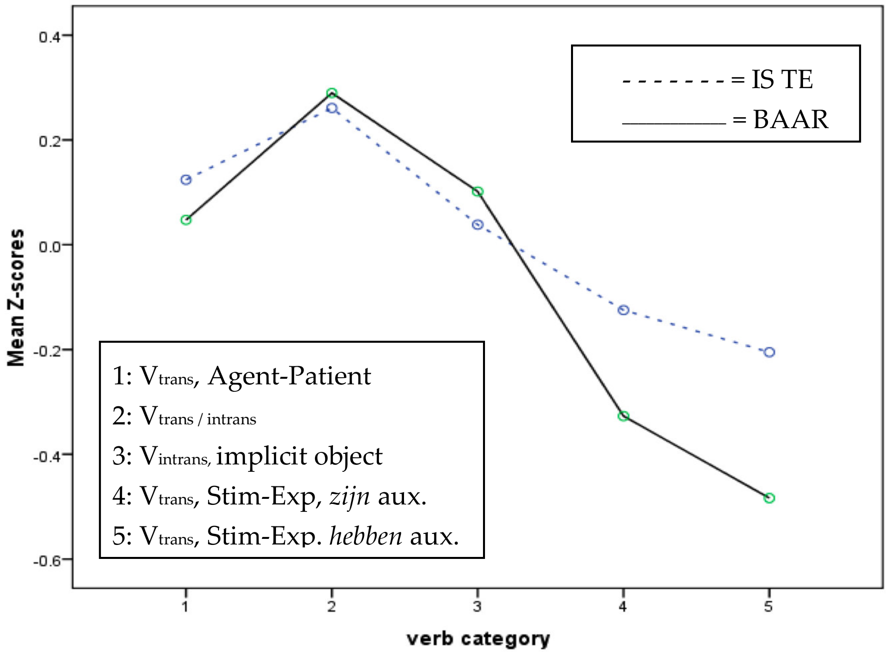


Figure 3.3 : Mean Z-scores per item category for children

Table 3.12: Mean Z-score per verb category and construction type for children

	is te V(sd)	V-baar
I. V_{trans}, Agent-Patient	.124 (.361)	.047 (.469)
II. $V_{trans\,/\,intrans}$	.261 (544)	.290 (.549)
III. $V_{intrans,}$ implicit object	.038 (.523)	.102 (639)
IV. V_{trans}, Stim-Exp, *zijn* aux.	-.125 (.476)	-.327 (.510)
V. V_{trans}, Stim-Exp. *hebben* aux.	-.205 (.475)	-.484 (.596)

The acceptability data of children and adults were analyzed separately because of the differences between the two test versions, which included a smaller set of test items for children, slight differences in the formulation of the instruction and the setting in which participants took place. Since the sphericity assumption is not met (the Mauchly test is significant), Huynh-Feldt corrected F-values will be reported.

For adults, the main effect of construction type is significant ($F_{(1.00,\ 71.00)}$ = 103.99, $p < .001$) as is the effect of verb category ($F_{(2.63,\ 186.59)}$ = 189.07, $p < .001$) and the interaction ($F_{(3.47,\ 246.52)}$ = 18.28, $p < .001$). Pairwise comparisons show that for each pair of verb categories both main effects and the interaction effect are significant. The same applies for the children's data: the main effect of construction type is significant $F_{(1.00,\ 147.00)}$ = 10,58, $p < .001$. The effect of verb category is also significant $F_{(3.58,\ 526.71)}$ = 59.42, $p < .001$ as is the interaction $F_{(3.87,\ 568.49)}$ = 6.43, $p < .001$. Again, pairwise comparisons show significant differences between all pairs.

The effect of construction type shows that adult participants rated novel instantiations of the IS TE V construction in general as more acceptable than novel V-BAAR forms. This may well be caused by the fact that the occurrence of novel words is a much rarer phenomenon than the creative or new combination of words. A novel instantiation of the V–BAAR construction leads to a new word, a word that the participants were unlikely to have encountered before. For the IS TE V items, the novel combination of a specific verb and the construction did not lead to new words. This means that the V–BAAR forms were more salient as novel instantiations, which might explain the overall lower degree of acceptability.[20] For the children's data, the overall difference between the two constructions is also significant, with V–BAAR forms rated lower. The two figures show, however, that they disliked V–BAAR test items especially for verb categories IV and V, rather than the more general preference for IS TE V items that is visible in adults' responses.

With regard to the verb categories, two elements in the distribution of scores are particularly striking. The first of these is the clear distinction in acceptability between verb categories I, II and III on the one hand and IV and V on the other. This distinction is present in adults' and children's responses alike. The first three verb categories have in common that they are regularly expressed with an Agent-Subject. In addition they are expressed with an obligatory Patient (category I), an optional Patient (II) or an implicit

[20] Incidentally, a number of people made a remark about the occurrence of *–baar* in the test items after participating. No-one mentioned *is te*.

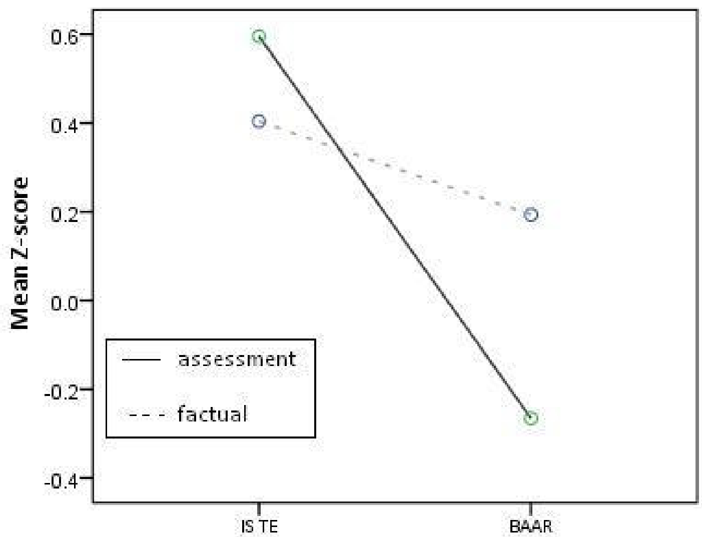
Patient (III). Clearly, the acceptability of novel instantiations does not so much require an expressed Patient, but depends on this argument's interpretability.

The strongest interaction effect between verb category and construction type is visible in the pairwise comparison of verb categories IV and V for adults (verb effect: $F_{(1,71)} = 21.77$, p < 001, construction effect: $F_{(1,71)} = 112.75$, p < .001, interaction effect: $F_{(1, 71)} = 63.43$, p < .001). For V–BAAR items, it does not make much of a difference whether a Stimulus-Experiencer verb takes *hebben* 'have' or *zijn* 'be' as its auxiliary in perfect tense. For IS TE V items, on the other hand, those verbs that occur with *zijn* 'be' are a lot better than those that do not.

Assessment of potentiality and factual potentiality

For the analysis of factual potentiality versus assessment of potentiality, we only have a limited number of items: 16 for adults and 12 for the children's data (8 and 6 respectively for each of the two constructions, divided over 3 or 4 verbs: each verb occurs in all contexts). Again, we applied a GLM repeated measures analysis.

Figure 3.4 below shows that, in the adults' data, the general preference for IS TE V is also present with these test items. More interestingly, IS TE V items with an assessment of potentiality, i.e. explicit stance markers like *denk ik* 'I think', score higher than factual potentiality. The reverse is true for V-BAAR, with an especially strong rejection of assessed potentiality (mean Z-score = -.266). The main effects and the interaction are significant (construction effect: $F_{(1,71)} = 74.92$, p <.001, potentiality type: $F_{(1,71)} = 7.81$, p < .01, interaction: $F_{(1, 71)} = 51.83$, p <.001).



In

Figure 3.4: mean Z-scores for baar and is te items with factual and assessed potentiality for adults

children's responses, the difference between the expression of factual or assessed potentiality is not significant (p = .12). There are two possible explanations for this difference between the analyses of adults' and children's data. Firstly, the analyses are based on few observations: 8 for adults and only 6 for children. Secondly, the semantic distinction between factual and assessed potentiality is subtle, and it may be the case that this fine-grained difference has not been acquired (yet) by the children that were tested in this experiment. Further experiments with older children and a larger number of test items are necessary to substantiate this hypothesis.

3.3.3 Summary and analysis

In sum, the acceptability of novel V-BAAR and IS TE V forms is much higher for verbs for which a Patient argument is interpretable than for other verbs with two arguments, at least for Stimulus-Experiencer verbs. Participants give higher acceptability ratings to novel IS TE V instantiations than V-BAAR forms, possibly due to the salience of new words compared to multi-word combinations. In general, verbs with a Stimulus-Experiencer argument structure are dispreferred compared to other test items, but this tendency is weaker for IS TE V with those verbs that take *hebben* as their auxiliary in perfect tense. Finally, the expression of either factual potentiality or the assessment of it leads to rather different acceptability ratings from adults for both constructions: IS TE V is rated higher in sentences marked for stance, and V-BAAR for factual potentiality.

Most of the results of the ME task correspond remarkably well with the corpus data. The acceptability ratings for different verb types reflect frequency of occurrence in the corpus. For IS TE V, we found a somewhat larger number of different verb types in the corpus than for V-BAAR, which ties in well with the overall higher acceptability ratings for IS TE V. The corpus analysis for IS TE V revealed that this construction is often used to mark a speaker's assessment of potentiality rather than an objective fact. This too is replicated in adults participants' responses to test items, albeit in a slightly different way: here it seemed more like assessment was rejected for V-BAAR items than valued for IS TE V items.

3.4 Discussion

3.4.1 The V-BAAR and IS TE V constructions in the MultiRep model

In this chapter, we have investigated whether two seemingly synonymous potentiality constructions in Dutch, referred to here as V-BAAR and IS TE V, can be said to be productive. If a template can be shown to be used productively, this provides evidence for the cognitively real existence of a partially specific pattern, i.e. a syntactic construction; otherwise we can only show that there are various instantiations, each stored as a (lexical) unit, that happen to instantiate the same underlying pattern.

The MultiRep model that was introduced in Chapter 1 assumes that knowledge about CLIs and the patterns they instantiate develops in a bottom-up fashion: based on the input children receive, they will first acquire specific form-meaning pairings. As they hear and read more instantiations of a pattern, they might develop (partially) abstract representations of the construction. It was argued in the first section of the current chapter that the simple occurrence of forms is insufficient evidence for the productivity of a pattern. In this chapter, by focusing on the issue of productivity we investigated what kinds of the abstract representations speakers have for the V-BAAR and IS TE V constructions (cf. Figure 1.8 in Chapter 1).

To the extent that the experimental task we used contained novel instantiations of the constructions, it forced participants to make use of abstract representations they may have had for these patterns. The acceptability ratings that the participants provided showed a pattern related to the verb types that the test items contained. We interpret this as evidence for the existence of an abstract representation for each of the two constructions: the consistency in rating differences can only come about if participants had a general representation available to them to which the test items could be compared.[21]

In subsection 3.1 we argued that the burden of proof is on the abstract pattern: the default assumption is that forms are stored. The data we subsequently presented show, we think, that for the V-BAAR and IS TE V constructions we have now proven that the abstract representations are also cognitively real.

3.4.2 Discussion and conclusions

A comparison of two constructions with virtually the same meaning allowed us first of all to determine whether the two constructions are really synonymous. They turned out to differ in subtle ways, which suggests that the constructions do exist as independent templates, ready for use in novel ways, i.e. with new verbs in the open slot reserved for the verb stem (in the case of V-BAAR) or the infinitive (in IS TE V). These differences were revealed through collostructional analyses of the two constructions and a distinctive collexeme analysis, directly comparing them. While a few verbs occur frequently in both constructions, many were attracted to only one of them. This suggests that each construction has its own meaning, slightly different

[21] One reviewer asked whether unproductive constructions could also have an abstract representation, provided that the set of instantiations is large enough. In our view, this would be unlikely, albeit not impossible under the model's assumptions. It would require a set of instantiations with sufficient overlap in form, meaning and distribution for an abstract representation to develop. At the same time, all items that match with the (underspecified) characteristics of the slot in the abstract representation would already be part of the known set of instances that are stored at the lexically specific level. Note that the experiment in this chapter does not have anything to say about this issue. Here we only argue that the consistency in the ratings of novel instantiations of these two constructions proves that participants have abstract representations of these constructions.

from the other one, that matches better with some verbs than with others. The main difference appears to be that IS TE V functions as a stance marker, enabling speakers to convey their assessment of whether something is likely or not to happen, rather than factual possibility. In order to verify the cognitive reality of the schemas, we conducted a Magnitude Estimation experiment, in which subjects judged the acceptability of various instantiations using the two constructions. The evidence this yielded converges with the corpus evidence, in that the acceptability of novel instantiations of the two constructions is better for some classes of verbs than for others.

In addition, acceptability ratings for the constructions differ from each other, with IS TE V rated significantly higher. We tentatively interpret this to mean that IS TE V is more productive than V-BAAR, and that this is related to their character as multi-word phrase ('syntax') versus derived word ('morphology'), respectively. This may be linked to differences in salience between the two constructions (see subsection 3.1.1.2), with novel instantiations of the more salient pattern V-BAAR rated as less acceptable.

In sum, the main conclusions are that the two constructions investigated are productive and that they are not synonymous. The two sources of evidence used in this chapter, corpus data and acceptability judgments, support each other: they demonstrate different aspects of the same phenomenon.

Chapter 4:

Processing CLIs: A case study of Fixed Adjective-Preposition constructions (FAPs)[1]

4.0 Introduction

A central question in this book is what kinds of units people access when they use language. In the first chapter, a model was outlined which stipulates that language users have information about linguistic elements stored at various levels of specificity. The experiments that were discussed in the previous chapters investigated the existence and use of abstract representations alongside lexically specific representations. In this chapter, we take one further step and look at the (temporary) storage of a multi-word CLI with two lexically specific elements as a unit. As such, this study expands the range of CLIs we look at. In the first chapter (Section 1.2) it was argued that CLIs may consist of more than one word. The IS TE V construction discussed in Chapter 3 contains only one lexically fixed element: the word *te*, while the CLIs investigated in this chapter are a type of multi-word conventional pairing. They are instantiations of the Fixed Adjective-Preposition (FAP) construction, which is introduced in the next section.

It is usually impossible to observe what level(s) of representation someone is using when producing or understanding a sequence of morphemes or words. When a speaker utters the word *unable*, this could be the result of combining the UN+ADJ construction with *able* or by retrieving the unit *unable*: there is no difference in the resulting utterance. In the current chapter, we report on an experiment in which we attempt to discover the boundaries between units by asking participants to briefly remember sentences. The expectation is that they cannot remember the whole sentence at once; the the sequences they do remember inform us about the chunks that they divide sentences into, thus telling us whether a CLI is stored as a whole or in parts.

---

[1] This chapter is a revised and extended version of a chapter to appear in the forthcoming ICLC/CorpLing 2009 Volumes, edited by Dagmar Divjak and Stefan Gries and is co-authored by Antal van den Bosch and Peter Berck. We are grateful to the audiences at the 2009 Corpus Linguistics Conference in Liverpool and the Sixth Semantics in the Netherlands Day at Leiden University in 2008 for their comments.

This chapter also explores the role of contextual distribution. It was suggested in Section 1.5.1 that knowledge about collocates is part of the representation that speakers have of any given lexical item. In this chapter, we look more closely at one collocate that these FAPs have: the finite verb. In addition, a computational memory-based language model is used to assign a word predictability level to each word in the sentences. This measure, referred to as *word perplexity*, is based on the preceding words, and tells us how (un)likely the next word is, given its context. Although it is a relatively limited measure, only taking into account the previous three words, this is a first approximation at quantifying distribution and context as an element that influences language processing.

Words are not distributed randomly; they tend to co-occur in more or less predictable patterns. Some sequences are highly fixed: given the sequence *they lived happily ever...* most speakers of English would assume that the next word is *after*, based on their previous exposure to this expression. Because such word combinations are conventional, they must be part of speakers' linguistic repertoires. Not all collocations are as fixed as *happily ever after*. We do not know to what extent speakers process such combinations as one unit. On the one hand, they consist of more than one meaning-carrying element, while on the other hand they are unit-like in that they have a clear joint meaning.

In spontaneous language, it is difficult to determine if a multi-word sequence is really one unit, although sometimes there are indications, e.g. when there is phonological reduction in speech. Bybee and Scheibman (1999), for instance, observe that *do not* is most likely to be reduced to *don't* in its most frequent context, i.e. when preceded by *I* and followed by *know, think, have to* or *want*. Advances in technology in the past decade have made it possible to follow people's gaze in reading, offering a very interesting tool for investigating 'units' in language processing. An example of such research is Schilperoord and Cozijn (2010) who used an eye-tracking experiment to investigate anaphor resolution. They found reading times were longer for antecedents and anaphoric information if the antecedent was part of a longer fixed expression (e.g. ...*by the skin of his teeth$_i$. They$_i$ have to be brushed twice every day*), than if the antecedent was not part of a fixed expression. This indicates that elements inside the fixed expression are less available for anaphor resolution, which in turn may be interpreted as evidence for the unit-status of such expressions in reading. Besides measuring eye movements in reading, however, it is hard to tap into processing directly without interfering with it. We use an experimental technique attempting to do just that: investigating the units people divide an utterance into when they memorize it for reproduction.

In this chapter, we focus on a specific complex lexical item, the Fixed Adjective-Preposition construction (FAP). First, we examine human sentence processing data for this construction and its sentential context. Second, we look at the probability of words given the previously occurring words in FAP sequences using a measure of word-level perplexity. By then comparing these data, we are able to analyze to what extent the likelihood of a FAP sequence influences sentence processing.

In the experimental task, we focus on six specific adjective-preposition sequences in Dutch, and contrast two contexts (varying the finite verb in the utterance) and two interpretations (i.e. as a unit or as a coincidental sequence). The task and test items are outlined in Sections 4.2.1.4 and 4.2.1.5. One likely indicator of what is used as a unit is co-occurrence patterns in attested speech. For that reason, the processing data from the experiment are compared to a likelihood measure for observing the next word, based on a corpus. This corpus does not contain any tagging other than letter sequences and word boundaries; no Part of Speech tags or constituent structure is externally added to the input the model receives, on which to base its measures. Thus, the model can be classified as a knowledge-poor stochastic language model as typically used in speech recognition systems (Jelinek, 1998).

The final section of this chapter discusses the relation between such measures and human language processing and relates the results to the MultiRep model introduced in Chapter 1. On the basis of a correlation analysis of task results (reflecting processing) and likelihood measures (reflecting co-occurrence patterns) we attempt to point out what aspects coincide and where the two differ; finding the latter would indicate that humans do something else (or more) than what a simple stochastic model does.

4.1. The Fixed Adjective-Preposition construction (FAP)

4.1.1 The fixed adjective-preposition construction: form

In this chapter we look at a specific type of conventional word pairs: the fixed adjective-preposition construction in Dutch. An example of such a combination is *trots op* 'proud of'. At first glance, this seems to be a two-word fixed expression, but a closer look reveals that the adjective is always preceded by a subject and a verb (see below) and that the preposition is followed by a nominal constituent, with which it forms a prepositional phrase. Example (1) is a prototypical instance of the pattern.

(1) de boer is **trots op** zijn auto
the farmer is **proud of** his car (fn001204.26)[2]
structure: NP V_{fin} [Adj [Prep $[N]_{NP}]_{PP}]_{AP}$

The fixed elements in this construction are the adjective and the preposition. This combination is conventional: the selection of the preposition is not semantically transparent (in fact, the literal translation of *trots op* is 'proud on', instead of 'proud of'). There are quite a number of these conventional pairings in Dutch (see Section 4.1.1.2 below). Because they are purely conventional and fixed, it must be assumed that speakers of Dutch have stored these patterns in their constructicon. In addition to these two lexically specific elements, the construction also contains a number of underspecified elements. An underspecified element of a construction is an element that must be expressed in each instantiation, but not always with the same exact words (see Section 1.4). The FAP construction has three such elements, which are described below.

4.1.1.1 Underspecified element 1: the nominal constituent

The lexical content and internal structure of the noun phrase are not specified for the construction; it may take different forms. These range from anaphoric pronominal expressions (example 2) to referential lexical NPs with a noun (+ optional modifiers, examples 3 and 4) and even full clauses.

(2) Jonas wou dat niet dus die was **boos op ons**
Jonas wanted that not so that was **angry at us**
"Jonas didn't want that, so he was angry at us" (fv901185.238)

(3.) Titia is voortdurend boos op haar vader
Titia is continually angry at her father (fn001240.20)

[2] All Dutch examples are taken from the Corpus of Spoken Dutch (CGN).

(4) die is heel erg boos op z'n uh Deense opponent omdat ie zomaar gaat liggen
he is very much angry at his uhm Danish opponent because he simply lies down. (fn007444.108)

When there is a reference to a full clause, the construction includes the placeholder *er* 'there' followed, possibly at a distance, by the adjective + preposition unit, followed by a clause starting with *dat* 'that' or *om* 'to' (examples 5 and 6). If the reference is mentioned in the previous utterance, it is possible to refer back with *hier* 'here' or *daar* 'there' (7).

(5) we zijn **er trots op dat** dit gebouw er is zoals het er staat
We are **there proud of that** this building there is so_as it there stands
"We're proud that the building is there the way it is" (fn000070.10)

(6) nee ik uh ben **er** maar matig **enthousiast over om** nog even wat erbij te pingelen
No I uhm am **there** but meagre **enthusiastic about to** still quickly something there_with to barter
"No, I am not too enthusiastic about the option to barter for something extra" (fn000931.181)

(7) maar zoals wij uh dat ingericht hebben nou **daar** was iedereen **jaloers op**
But like we uhm that arranged have well **there** was everyone **jealous of**.
"but the way we arranged that, well everyone was jealous of it." (fn008213.286)

4.1.1.2 Underspecified elements 2 and 3: the verbal constituent and the subject

All example sentences so far contain a copula. The adjective in the FAP construction is used predicatively, and the utterance expresses that a property is ascribed to the subject. Most instantiations of the FAP construction take this form. In Dutch, adjectives used attributively occur before the noun. The pre-nominal adjectival phrase is usually short, as it is in English. Adjectives can be stacked (*I saw a tall, dark, handsome man*) or modified (*he was an extremely good-looking guy*), but, in English and Dutch alike, they cannot contain a modifier in the form of a prepositional phrase (**I saw the jealous of his colleague professor*).

The absence of FAP instantiations in attributive positions means that this construction typically co-occurs with one of a very short list of verbs, namely those that are found in copula constructions. Of these, *zijn* 'be' is by far the most frequent; it occurs in all examples given so far (ex. 1-7). Other copula verbs with the FAP construction are rare, but they can be found (8).

(8) je **zou** toch **jaloers op** die beesten **worden**
You **would** still **jealous of** those animals **become**
"It would make you jealous of those animals" (fn001227.46)

In terms of likelihood, the occurrence of a FAP instantiation is a much more reliable cue for the co-occurrence of a copula verb than vice versa: copula constructions are a lot more frequent than FAP constructions and the complement constituent may take many different forms, with both adjectival (*she is very intelligent*) and nominal constituents (*she is a doctor*). The verb links the property expressed in the adjective to a subject; the copula construction means that this link consists of ascribing the property to the subject. Since the verb is a strong collocate for the FAP sequence, this will be part of the experimental design (see Section 4.1.1.4 below). As there do not seem to be any clear distributional patterns for specific subjects, other than that subjects are typically references to humans, that underspecified element of the construction will not be varied systematically in the test items.

4.1.1.3 Summary and comparison to the prepositional verb construction

In sum, the fixed adjective-preposition construction consists of two fixed elements: an adjective and a preposition, which are conventionally paired. The underspecified elements are the subject, the verb and the nominal complement to the adjective. Although there are a large variety of subjects, the verb that occurs with this construction is usually *zijn* 'be'. The nominal element can take different forms.

In this respect, the FAP construction is similar to a much more commonly discussed construction in Dutch: prepositional verbs. These too allow for a variety of nominal constituents to fill the slot in the prepositional phrase. For prepositional verbs as well as the FAP construction, the non-prepositional element (the verb and the adjective, respectively) does not always occur in this construction: we may find *trots* 'proud' without *op* 'of' (see example 13) and *geloven* 'believe' without *in*.[3] This is a relevant fact with regard to the likelihood of a word's occurrence given another word: if *trots* is not always followed by *op*, this means that the likelihood of *op* after *trots* is less than 100%. In this respect these combinations differ from completely fixed expressions like *happily ever after*. In the case of *geloven*, two different argument structures are possible: either the object of faith is not expressed at all, and the verb is used intransitively, as in (9), or it is a direct object, as in (10). When *geloven* is part of the prepositional verb *geloven in*, the direct object is also expressed (11). For adjective-preposition combinations, the object can only be expressed in the same clause if the preposition is used.

(9) 't belangrijkste teken voor iemand die **gelooft**
The most_important sign for someone who **believes**.
(fv400022.232)

(10) niemand **gelooft** dat
Nobody **believes** that (fv600368.151)

(11) **gelooft** u **in** buitenaardse beschavingen?
Believe you **in** extraterrestrial civilizations?
"do you believe in extraterrestrial civilizations?" (fn001374.5)

We now turn to a description of the meaning of the FAP construction.

4.1.2 The fixed adjective-preposition construction: meaning

The adjectives that occur as part of a FAP construction also occur outside of these, carrying a very similar meaning. A direct comparison of a FAP

[3] It is here that FAPs differ from many one-word CLIs, i.e. derivations: by definition, bound morphemes such as *–baar* (cf. Chapter 3) do not occur separately; it is only used as part of the V–BAAR construction. This is especially relevant from the point of acquisition: the occurrence of *–baar* is an extremely reliable cue to identify an instance of the V–BAAR construction. Hearing or seeing *geloven* or *trots*, on the other hand, is not sufficient to determine if the prepositional verb *geloven in* or the FAP construction TROTS OP are used.

instance and an utterance with the same adjective but no prepositional phrase allows for a first approximation of the construction's meaning:

(12) hij is **jaloers op** jullie mooie huis
He is **jealous of** your beautiful house (fn001175.116)

(13) en als je met de postbode stond te praten was ik echt heel erg **jaloers**
and when you with the postman stood to talk was I really very much **jealous**
"and when you were talking to the postman I was really very jealous" (fn001041.21)

In both (12) and (13) there is a person to whom the property of an emotion is ascribed - jealousy. The prepositional phrase *op jullie mooie huis* 'of your beautiful house' in (12) is a lexically specific reference to the object the emotion is aimed at, the cause of this emotion. In (13) there is no prepositional phrase. The cause of the jealousy, however, is very clear from the context. A corpus search of *jaloers* in the Spoken Dutch Corpus (118 occurrences) reveals that the object of the jealousy is not always so explicitly expressed. In those cases where *jaloers* co-occurs with *op* (39 tokens, allowing for intervening sequences of up to 5 words), this prepositional phrase always refers to the object or person that causes this emotion, i.e. who or what the jealousy is aimed at.

The case of *jaloers op* is not unique. In many FAPs the prepositional phrase contains a reference to the cause and/or recipient of the emotion that the adjective describes (e.g. *blij met* 'happy with', *bang voor* 'afraid of', *verbaasd over* 'surprised about' etc.). While not all adjectives that occur in FAPs refer to emotions, very often the prepositional phrase refers to a cause, as in *allergisch voor* 'allergic to', *kwijt aan* 'lose to', *ziek van* 'sick of' etc. (see examples 14 and 15).

(14) dus hij is een week van zijn vakantie **kwijt aan** stage?
So he is a week of his holiday **lost on** internship?
"So he will lose a week of his holidays to doing an internship?" (fv400194.22)

(15)ze ging wel vroeg naar bed want ze was **ziek van** 't hete eten of zo
She went sure early to bed because she was **sick of** the spicy food or something

"She did go to bed early, because the spicy food or something had made her sick." (fn000384.169)

4.1.3 The FAP construction: form and meaning

The general pattern that seems to underlie all the examples reviewed so far is summarized in Figure 4.1. For reasons of space, the subject and the verb are left out. The adjective expresses a property, and the noun phrase is the cause of that property. They are linked in a complement relation. The lexical expression of this relation is the preposition.[4] Note that the representation in Figure 4.1 is an abstraction over specific FAPs.

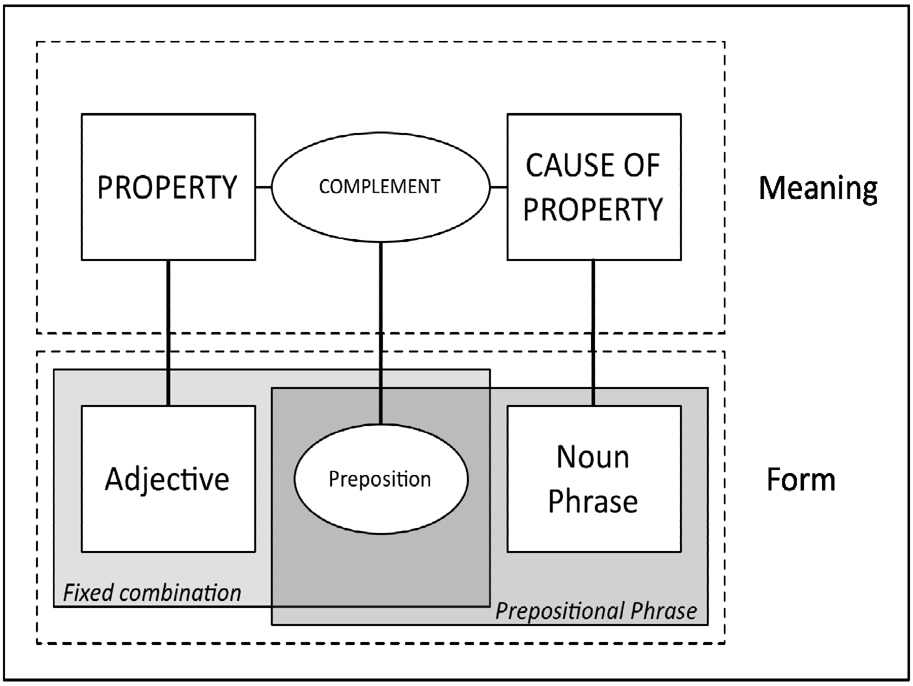


Figure 4.1: Visual representation of the FAP construction

[4] It is possible to express both the property and the cause in other ways, e.g. in two separate clauses (*He scored a goal. His mother was very proud*) or with a nominalization (*His goal made his mother very proud*).

The fact that it is possible to formulate a form-meaning pattern that the different FAPs all seem to follow, does not mean that this general schema is a level of representation that is psycholinguistically real, i.e. that speakers of the language have this in their constructicons. In this chapter, we will not attempt to find out whether the latter is the case, as was the goal for the two constructions in Chapter 3. Instead we focus on the degree to which specific FAPs are used as a unit in human sentence processing. This distinction also serves to show that constructions can be studied with different degrees of detail: each specific FAP is an instantiation of the construction as depicted in Figure 4.1, and utterances with different lexical items in the underspecified elements are instantiations of a FAP.

Even when both fixed elements of the FAP sequence occur in the same utterance, they may still not trigger the FAP interpretation, as is the case in (16) -although these examples are very rare.

(16) ja dan kan ik wel begrijpen dat men hier **op** Urk een beetje **trots** is en spreekt van het wonder van Urk
Yes then can I sure understand that one here **on** Urk a little **proud** is and speaks of the miracle of Urk
"Yes, in that case I can understand that people here on Urk (a former island) are kind of proud and speak of the miracle of Urk" (fn007548.24)

The prepositional phrase *op Urk* 'on Urk' in this sentence could in principle refer to the cause of pride. The discourse context makes it clear that the reference is to pride possessed by inhabitants of Urk *men hier op Urk* 'people here on Urk': the prepositional phrase refers to a location. This example shows that not all FAP sequences necessarily have a FAP interpretation; the sequence may also be accidental.

The cause of a property is not something that a speaker always needs to express. For that reason it is not surprising that adjectives that sometimes occur in FAPs are also found without the prepositional phrase expressing that cause. The fact that adjectives referring to emotions seem to take up a large part of the distribution of FAPs may be explained by the construction's semantics: emotions are always caused by something or someone, and this is a relatively salient aspect of this type of adjectives, compared to many other semantic groups of adjectives (e.g. colors, adjectives related to size etc.). For adjectives referring to emotions, when this property is ascribed to someone, the cause of that emotion must be retrievable in the discourse context. It can be expressed lexically with a FAP, but this is not necessary: sometimes an earlier mention suffices, or the cause could be left unexpressed on purpose. The relative salience of the cause of an emotion entails that in a significant number of instances it will be useful information to express lexically. An in-depth analysis of the distribution of FAPs, for instance with the help of a behavioural profile (Divjak & Gries, 2008), could shed more light on the specifics of this construction, but is outside of the scope of the present chapter.

In the preceding section, where the form of the construction was described, we noted that there are some parallels with prepositional verb constructions. With regard to meaning, however, we see a clear difference between the two patterns here: the function of the prepositional phrase in the FAP construction is that of an extension. It allows for the lexical expression of the cause of the property described in the adjective. As it is a non-compulsory addition, it may be appropriate to call this a modifier. With prepositional verbs, the prepositional phrase usually contains an object-like complement, as is for instance the case with *geloven in* 'believe in'. The distinction between complements and modifiers, however, is a gradual one (Taylor 2002:234-5). Whereas the lexical expression of the cause of the emotion is optional in an utterance, this cause must be retrievable in the discourse context; its optionality is relative to the context.

4.2 Experiment: The copy task

Determining what units people use when they produce language is nigh on impossible, because the unit boundaries are not always visible: speech is continuous. Earlier research has shown that pauses in speech often occur at boundaries between constituents, although they are also found after function words, i.e. inside nominal and prepositional phrases (e.g. Hawkins, 1971; Goldman-Eisler, 1972; Grosjean & Deschamps, 1975). Recent research into the effects of global and local text structure on pauses in speech shows that most speakers use pauses to indicate text structure to listeners (cf. Den Ouden, Noordman & Terken, 2009). This means that pauses in speech serve a communicative purpose, but does not provide conclusive evidence about the processing units in speech production.

An alternative to the analysis of speech is to ask participants to divide sentences into units or ask them how strongly consecutive words are related, as Levelt (1969) did. The strength-of-relationship measure thus obtained strongly resembles the constituent structure. The problem or shortcoming of this technique, however, is that participants have to rely on explicit knowledge (see Section 2.1.2). It is an offline task and therefore does not measure language processing. Participants, who have been taught in school to analyze sentences in terms of constituents, may use this to perform the task, regardless of whether this actually conforms to the units they use themselves. Unfortunately, it is impossible to determine with certainty whether this strategy is employed (participants may even do this subconsciously, such that asking them about their strategies may not solve this problem).

Griffiths (1986) introduced an inventive design that sought to overcome the vagueness of pauses in speech as a measure, while maintaining the online character of the task. He asked participants to copy sentences in writing. They saw a sentence, which they then had to copy on a sheet of paper. The sentence was not in sight of the copy sheet. Participants were told that they could look back at the original sentence whenever they were unsure about how to continue. Each time they did this, it was registered at what point they were in the copying process (i.e. after which word they had to look at the sentence again). Much like the pauses, the look-back points are an indication of a unit boundary, although not looking back does not mean that there is no unit boundary: participants can remember more than one unit at a time. Since Griffiths' introduction of the method, technological advances have made it much easier to execute such an experiment. Straightforward software programs can create log files in which switches are registered (cf. Ehrismann (2009) for a similar experiment).

This experimental design is not a speech production task, but it is a production task nonetheless. In Gilquin and Gries' classification of kinds of linguistic data in terms of naturalness (Gilquin & Gries 2009:5) this task rates rather low, though. All experimentally elicited data rank in the lower half of the naturalness scale, and experiments involving participants to do something they would not normally do with units they do not usually interact with come at the bottom of the range. Arguably, memorizing long sentences verbatim with the aim of reconstructing them is not a 'natural' task. We return to this issue in the closing section.

The task requires participants to store (a part of) a sentence in working memory for the duration of the copying process. Because participants have to reconstruct the exact sentence —paraphrasing is not allowed— they will have to store the specific sequences the sentence is made up from. This provides us with the advantage that we can see into what parts the participants break up the test sentences, including the FAPs, while they perform the task, thus giving us an insight in the units they process. The switch behaviour between the original sentence and the copy screen indicates unit boundaries.

We use this experimental design to answer the following research questions:

1. Are FAPs a unit in human processing?
   a. Is the switching behaviour different for sentences in which the adjective-preposition sequence is coincidental (no semantic link between the adjective and the prepositional phrase) or if the prepositional phrase expresses the cause of the property (i.e. has a 'FAP interpretation')?
   b. Is the switching behaviour influenced by the identity of the verb that precedes the coincidental sequence or the FAP?

2. Does a stochastic word perplexity measure treat FAPs as one unit?
   a. Does the metric distinguish between coincidental sequences and those with a FAP interpretation?
   b. Is the metric influenced by the identity of the verb that precedes the coincidental sequence or the FAP?

3. Is the word perplexity measure relevant in a predictive sense to aspects of human sentence processing?

In order to answer these questions, we designed an experiment in which participants were asked to reconstruct sentences that they had just seen. These sentences each contained an adjective-preposition sequence. By varying the context, we were able to determine the influence on processing units of the presumed unit status of the sequence (research question 1a and 2a) and of the verb hypothesized to be associated with it (question 1b and 2b). The participants' data are compared to a word probability measure to answer the third research question. The relevant details of the experimental set-up are explained in the following paragraphs.

4.2.1 Experimental design

4.2.1.1 Participants

The participants were 35 children in sixth grade ('groep acht'), mean age 12;5 years. They were a subset of the participants in the experiment described in Chapter 3, and came from two primary schools in Tilburg, a city in the south of the Netherlands. All children participated on a voluntary basis, with consent given by their parents.

4.2.1.2 Item selection

In order to select frequent adjective-preposition pairs, we first made an inventory of all combinations that were listed in the *Prisma woordenboek voorzetsels* (Reinsma & Hus, 1999). This dictionary lists combinations of a preposition and another word for over 5000 different lexical items, and contains 472 different FAPs. Some adjectives occur with more than one preposition. In total 365 different adjectives are listed. Since our primary interest is in frequent combinations, we restricted this list to items where the adjective occurs at least 100 times in the Spoken Dutch Corpus, and the combination is found at least 10 times as a continuous sequence. 75 combinations met this requirement. For the experimental task we selected six combinations that allowed for the construal of test sentences in which the prepositional phrase could express either the 'cause' of the adjective ('FAP interpretation') or a separate location / prepositional phrase connected to the verb (see below for a detailed description of test items).

When the *Prisma woordenboek voorzetsels* contains two entries for the same adjective, in many cases there is a clear semantic difference between the two types of extensions, e.g. with *blij met* and *blij voor*, 'happy with' and 'happy for', respectively (see examples 17 and 18). While both types of referents can be construed as 'causes' for the emotion, the prepositional phrase introduced with *met* refers to the thing (physical object or achievement/result) that causes happiness, and the *voor* phrase contains a reference to a person whose presumed happiness causes vicarious happiness in the speaker.

(17) laten we dat vooropstellen we zijn natuurlijk wel **blij met** onze vrouwen
let us that first_put we are of_course indeed **happy with** our wives
"we must stress that we are of course happy with our wives" (fn000377.289)

(18) ik ben **blij voor** de prins dat hij eindelijk een mooie jonge vrouw heeft kunnen vinden
I am **happy for** the prince that he finally a beautiful young woman has can find
"I am happy for the prince that he has finally been able to find a beautiful young woman." (fv600215.13)

With other fixed combinations, this is less the case, e.g. *aardig tegen* and *aardig voor*, 'nice towards' and 'nice to'. Both prepositional phrases express the person kindness is directed towards (see 19 and 20).

(19) ja één keer en toen was ie best **aardig tegen** mij
yes one time and then was he kind_of **nice towards** me
"Yes, once, and he was fairly nice to me that time" (fn006990.20)

(20) maar iedereen is altijd heel erg **aardig voor** mannen in Japan hè
but everyone is always very much **nice to** men in Japan, right
"but in Japan people are always very nice to men, aren't they?" (fn007565.1380)

The distribution of these two combinations may be influenced by various factors, including region (i.e. one form is prevalent in a certain part of the country), genre and others. None of these combinations were selected as test items for the experimental task.

4.2.1.3 Corpus for the memory-based language model

The corpus on which the memory-based language model is trained is a combination of two newspaper corpora: the Twente news corpus[5] and the ILK newspaper corpus[6]. The former contains newspaper articles, teletext subtitles and internet news articles. The latter contains data from several regional Dutch newspapers. The first 10 million lines of the corpus, containing 48.207.625 tokens, were taken to train the language model. Tokens which occurred five times or less were replaced by a special token representing low frequency words.

4.2.1.4 Test items

The task consisted of 24 sentences, each containing one of the six selected adjective-preposition sequences. For each pair, we determined the most frequently co-occurring verb: *zijn* 'be' in five cases and *doen* 'do' for one pair. A sentence was created with this verb, a subject, the fixed combination, and an appropriate nominal phrase as the complement of the preposition, referring to the object the emotion named by the adjective is aimed at (see example 21). These sentences thus contain the frequent verb + adjective + preposition combination, with the prepositional phrase related to the adjective ('FAP interpretation').

(21) Al in april **was** Esra **enthousiast over** de vakantie naar haar familie in het buitenland (TYPE A sentence)
Already in April **was** Esra **enthousiastic about** the vacation to her family in the foreign_country
"In April already, Esra had been enthusiastic about the holiday trip to her family abroad"

In order to allow us to find out how the participants divide this sequence into smaller units, we placed the target sequence near the middle of the test sentence. Had the sequence been placed at the beginning, results would be clouded as the first words of a sentence are easily remembered. Positioning the target sequence too close to the end of the sentence would also create

[5] http://inter-actief.cs.utwente.nl/~druid/TwNC/TwNC-main.html
[6] http://ilk.uvt.nl/ilkcorpus

problems, because by that time it is easy to reconstruct the remainder of the sentence (see below for the task procedure).

Subsequently we created sentences with exactly the same subject, adjective and prepositional phrase sequence, but in which the prepositional phrase does *not* relate to the adjective. These sentences were created both with the frequent verb (example 22, TYPE B sentences) and with a verb that makes a different interpretation for the prepositional phrase, e.g. a location or topic, more likely (example 23, TYPE C sentences).

(22) Voor de pauze **was** Esra **enthousiast over** de vakantie aan het kletsen met haar vriendin (TYPE B sentence) [7]
Before the break **was** Esra **enthousiastically about** the vacation on the chatting with her friend
"Before the break, Esra was chatting enthusiastically to her friend about the vacation"

(23) Na het weekend **begon** Esra **enthousiast over** de vakantie te vertellen aan haar hele klas (TYPE C sentence)
After the weekend **began** Esra **enthousiastically about** the vacation to tell to her whole class
"After the weekend, Esra began to tell the whole class enthusiastically about the vacation"

The 2x2 design is then completed by generating sentences with a non-frequently co-occurring verb and an FAP interpretation (example 24, TYPE D).[8] Table 4.1 provides an overview of the sentence types.

(24) Lang voor vertrek **begon** Esra **enthousiast over** de vakantie alvast haar tas in te pakken (TYPE D sentence)
Long before departure **began** Esra **enthusiastic about** the vacation already her bag in to pack
"Long before departure, Esra, enthusiastic about the vacation, began packing her bag."

[7] Note that the English gloss for *enthousiast* in this sentence contains an adverbial suffix. In Dutch the same form can be used both adjectivally and adverbially.

[8] Various people commented that the type D sentences sounded 'unnatural'. Given that they combine a verb that usually does not co-occur with a 'FAP- interpretation', this sentiment is understandable.

Table 4.1: Types of test sentences *copy task*

		Frequent verb	
		+	-
FAP interpretation	+	Type A (example 21)	Type D (example 24)
	-	Type B (example 22)	Type C (example 23)

In addition to these requirements, for each adjective-preposition pair the sentences were constructed in such a way that the four types contained an equal number of words and differed by no more than two letters in total length. This was done in order to minimize the influence of these factors, so that any difference in test behaviour could be reliably attributed to sentence type. Table 4.2 summarizes these data for the six word pairs tested.

Table 4.2: Characteristics of test sentences

Pair	Translation	Frequent verb	Non-frequent verb	Word count	Letter count
Boos op	Angry at	Was	Stond stood	13	59-61
Enthousiast over	Enthusiastic about	Was	Begon began	15	70-71
Geïnteresseerd in	Interested in	Was	Stond stood	16	76-78
Jaloers op	Jealous of	Was	Begon began	13	63-64
Voorzichtig met	Careful with	Deed *did*	Liep walked	15	67-68
Trots op	Proud of	Was	Stond stood	15	78-79

Each test sentence had a similar grammatical structure, which is reproduced in Figure 4.2 below.

Test sentence (example)						
Na het weekend	begon	Esra	enthousiast	over	de vakantie	te…
Constituent 1	V_{finite}	Subj.	adjective	Prep	NP	rest

Figure 4.2 Grammatical structure of test sentences *copy task*

In sum, the test items varied on three points: six fixed pairs of adjectives and prepositions, two finite verbs per pair, and two types of relation between the adjective and the prepositional phrase. The underspecified element 'verb' is thus a variable that is manipulated. The other two underspecified elements, the subject and the noun phrase, remain the same for the different variations of each FAP sequence. A complete list of test sentences can be found in Appendix 5.

4.2.1.5 Procedure

The experiment took place in the children's schools, in their computer rooms. The participating schools had a dedicated classroom equipped with enough computers to let all pupils in one class do the experiment simultaneously. The children were told that they would participate in an experiment designed to find out what kinds of sentences are difficult or easy to remember. Each child worked at an individual computer.

After starting up the program, the children saw a brief introduction, outlining what they had to do. A short text explained that they would see a sentence, which they were to read. They then had to press the space bar, which replaced the sentence with a new screen. On this screen, they saw a number of words in the top half, and an empty bar at the bottom (see screen shot in Figure 4.3). The task was then to drag the words down to the bar in the right order to form the sentence they had just read. Only if the correct word was dragged down, it would stay there. Other words would pop back up, i.e. they could only reconstruct the sentence from left to right and had to start at the beginning. If at any point in the sentence they forgot how it continued, they could return to the original sentence by pressing the space bar again.

Figure 4.3: Screen shot *copy task* (practice sentence). The sentence at the top of the screen reads 'drag the words in the right order down (space bar = return to example)'. At the moment this screen shot was taken, a participant had already dragged the first four words of the sentence *dit is een zin* 'this is a sentence' down in the bar at the bottom of the screen.

After completing a sentence, a message in Dutch appeared telling the participant to press Enter to continue with the next sentence. All children completed the whole task, copying 24 sentences, with the order randomized for each participant. Before starting with the first test sentence, they had to do a practice sentence first. The researcher was present in the computer room and answered questions about the procedure when necessary.

On average children spent nearly 20 minutes on the task, but some took up to half an hour. Staying focused on the task turned out to be difficult, in spite of verbal admonitions by the researcher to try and be as quick as they could. Some children complained that the task seemed 'endless'. Because the order of the sentences was different for each participant, we assume that this has not compromised our data. All switch data were logged online and later retrieved for analysis.

4.2.1.6 Variables

Item-based variables

The test sentences vary with regard to the adjective-preposition sequence (six different pairs), the finite verb (two per pair) and the type of relation between adjective and preposition (two per pair).

The switches that the participants made between the sentence and the reconstruction screen were all coded for their position: at which point in the reconstruction process did a participant go back to the original sentence. This position was defined with regard to the last word reconstructed. If a participant correctly copied *lang voor vertrek*, the first three words of the test sentence given as example (24), and then switched before continuing with *begon*, this was registered as a switch after *vertrek*.

Processing variables

The software program made specifically for this experiment logged for each word how it was handled. The log files therefore show at what points in the sentences each participant switched. Since people will only store a limited number of units in working memory at one time, they will have to switch when that storage has run out. Each switch is therefore a sign of a boundary between two units, but the absence of a switch does not indicate that the child is dealing with only one unit. The sum of switches by all participants was determined for each word boundary (`*n* switches adjective', for example, for the word boundary following the target adjective).

Likelihood variables

We trained a computational memory-based language model (Van den Bosch & Berck, 2009) on the aforementioned newspaper text corpora. The memory-based language model predicts, based on a context of *n* consecutive words to the left, predicts a distribution of possible following words. The computational model can be likened to standard stochastic models that employ backoff smoothing, but without additional smoothing (Zavrel & Daelemans, 1997). Hence, if the model finds a matching context in memory that points to a single possible following word, the model predicts this word with a probability of 1.0. If there is a mismatch between the current local context of the *n* preceding words and the contexts in memory, the model backs up iteratively to find a match in the preceding *n-1* words, producing estimates that do include more than a single possible word, with their probabilities adding up to 1.0. We set $n = 3$, yielding a 4-gram memory-based model that was subsequently applied to the 24 sentences to establish word-level perplexities. This means that the model assigns a value of word perplexity to all of the words in the 24 test sentences, based on its predictions given the three preceding words.

For each word, we take the negative base-2 logarithm of the probability assigned by the model to the word that actually occurs as the next word. This measure is typically referred to as the word-level *logprob* (Jelinek, 1998). The metric is strongly related to word-level perplexity, another often-used metric in statistical language modeling to express the degree of surprise of a language model to observe a word given an earlier sequence of words: word-level perplexity is $2^{logprob}$. In the remainder of the text, we perform tests on the logprob measure, but occasionally refer to this metric as the "perplexity" measure.

4.2.2 Results

4.2.2.1 Descriptives

The 35 children who participated in the experiment and each copied 24 sentences switched a total of 1794 times: 2.14 switches on average per sentence. Each switch (back and forth between the sentence and the reconstruction screen) was coded for position. In other words, for each switch we know after which word it was made. In order to be able to sum switches over different test sentences, the positions were defined in terms of the word's function in the sentence (e.g. 'subject'). Figure 4.4 visualizes the aggregated switch behaviour for one sentence in the form of a dendrogram. The switch behaviour shown here is representative for the other sentences. The dendrogram illustrates which word sequences are more 'unit-like'. Sequences were iteratively combined starting with the word boundary that caused the fewest switches (in this particular example the last three words *een volle trein*). At each iteration, the sequences are linked that required the least amount of switches.

The dendrogram arising from the aggregated switches is remarkably consistent with the constituent structure of the sentences: the three prepositional phrases in the sentence *door de veranderingen*, *op haar school* and *in een volle trein* each are linked together before they are integrated in the rest of the sentence. The noun phrase inside the prepositional constituent is also visible in the switch data. The sequence finite verb – subject is a relatively strong unit as well. All of these are of course semantic as well as structural units. The first constituent is linked to the rest of the utterance at the last iteration. This is a recurrent pattern for all test sentences: at the end of the three- or four-word sequence that constitutes the first phrase, many of the participants switched. These patterns are significant: the number of switches between the last word of the first constituent and the finite verb, and between the finite verb and the subject are significantly different (mean number of switches at boundary first constituent = 12.2 (sd 2.8) and after the finite verb = 3.1 (1.6), $t_{(23)} = 12.75$, $p < .001$).

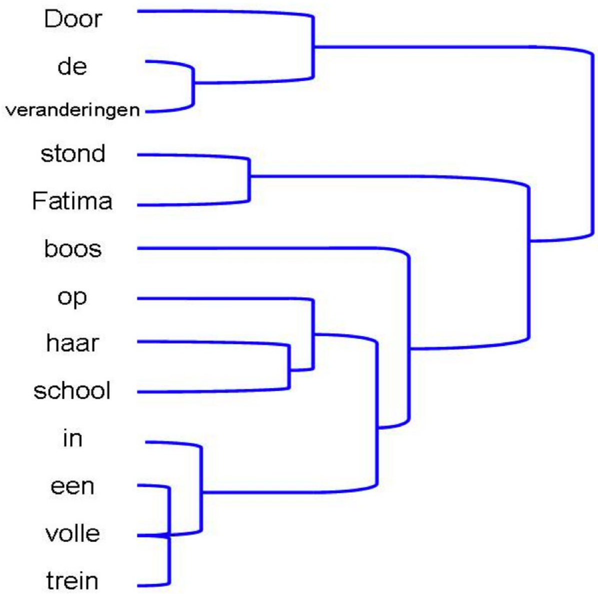


Figure 4.4: Visual representation of switch behaviour (sentence contains *boos op* 'angry at', a FAP interpretation and a non-frequent verb).
Door de veranderingen stond Fatima boos op haar school in een volle trein
Because_of the changes stood Fatima angry at her school in a full train.

Finally, the number of switches tapers off near the end. This is most likely due to the research design: the participants had to select the next word from the remaining words in the sentence, all visible on the screen. For the last couple of words, there were only a few left to choose from.

The same kind of dendrogram can be construed on the basis of the probability measure. This time, sequences were iteratively combined starting with the word with the lowest perplexity and the one it precedes. In the sentence given in Figure 4.5, the lowest perplexity value was assigned to the word *op*, which was therefore linked first with preceding word *boos*, after which sequences were linked that are progressively less likely to follow each other. Figure 4.5 contains a dendrogram for the same sentence as Figure 4.4, this time using the probability measure to construct it.

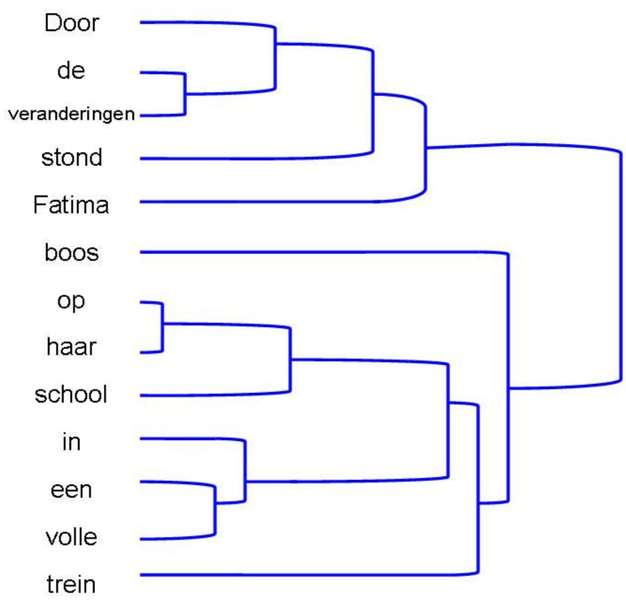


Figure 4.5: Visual representation of logprob (sentence contains *boos op* 'angry at', a FAP interpretation and a non-frequent verb).

The two figures are rather similar: the probability measure too results in a constituent-like structure for prepositional phrases. Note that within the prepositional phrase, the structure is slightly different for *op haar school*. In terms of probability, the preposition – determiner sequence is more of a unit than the determiner – noun sequence. This is a recurrent pattern: whereas the switch data tend to cluster a noun phrase within a prepositional phrase, the probability data often produce a preposition – determiner clusterA second difference is the absence of the sequence finite verb – subject as a unit in the probability-based dendrogram. Again, this is a recurrent finding for many of the sentences. We will return to these differences and discuss them in terms of human processing and the role of local co-occurrence probabilities therein in the last section of this chapter. Finally, unlike the human data, the probability measure does not profit from a reduction in possible candidates towards the end of the sentence, as the stochastic model has no access to the diminishing list of possible continuations that the human subjects have. We already suggested that this effect in the switch behaviour is due to the experimental design. This possible explanation is corroborated by the absence of the effect in the probability data.

4.2.2.2 Statistical analyses

Research question 1: human switch data

For each adjective-preposition sequence, there were test sentences with and without a frequent verb and with and without a semantic link between the prepositional phrase and the adjective, a FAP interpretation.

To determine whether the verb made any difference to the switch behaviour, sentences with a frequent verb (frequent in co-occurrence with the FAP) and those with an infrequent verb were contrasted. The only word boundary where switch behaviour was significantly different was at the 'subject' position. Fewer switches were made after the subject when the verb was a collocate of the adjective-preposition sequence: the mean number of switches was 10.33 (sd 2.64) for sentences with a collocate verb, and 14.00 (2.95) for a non-collocate verb ($t_{(22)} = 3.21$, $p < .01$). This is an effect of the verb on the amount of switches made at a *later* point in the sentence: in each sentence, the verb came directly before the subject. Participants switched fewer times *after* the subject if the *preceding* verb was a collocate of the sequence.

The effect of the interpretation (FAP interpretation or coincidental sequence) was not significant: In a direct comparison of switch behaviour for all four types of sentences (recall the 2x2 design), an analysis of variance was carried out with verb and interpretation as factors. Whereas the main effect of verb was significant ($F_{(1,20)} = 11.08$, $p<.01$), the effect of the interpretation was not ($F_{(1,20)} = 1.85$, n.s.) and there is no significant interaction.

The 2x2 design for the test sentences entails that some of the sentences contained a verb likely to co-occur with the adjective-preposition sequence in a FAP interpretation, but where the prepositional phrase turned out to serve a different purpose (type B sentences). Likewise, other test sentences include a verb that does not co-occur often with the FAP, but was still followed by it (type D sentences). Such sentences may lead to garden path effects: the first part seems to lead to one interpretation, but as the sentence continues, it becomes clear that a different interpretation is the correct one. To check whether this caused a difference in switch behaviour, we contrasted type A and C sentences (no garden path effect likely) with type B and D sentences (garden path likely). There was indeed a significant effect: the likely garden path sentences required more switches at the word boundary after the first word following the prepositional phrase, that is once the whole FAP construction, including the full prepositional phrase, had been copied (mean nr. of switches for non-garden path sentences is 1.75 (1.54) and for potential garden path sentences = 4.75 (3.11), $t_{(22)} = 2.99$, $p < .05$). This is a clear indication of a garden path effect: the sentences are processed with equal effort until *after* the entire target sequence. It is at this point that participants had to reanalyze the structure of the sentences.

In sum, we find a small significant effect of the verb in the test senctences: a frequent verb reduces the amount of switches needed after the subject in comparison with a verb that did not occur frequently with the adjective-preposition sequence. The distinction between FAP interpretation and coincidental sequences did not directly lead to differences in switch behavior. This suggests that co-occurrence patterns are influential for the temporary storage of the sentence by the participants, an observation that is confirmed by the effects of the likely garden path sentences.

Research question 2: probability measure

For each word in the 24 test sentences, the model determined the probability of that word occurring. The measure used to reflect this probability is the aforementioned logprob measure. Analogous to the sum of switches for the copy task data, this results in a numerical value for each word boundary. Values for the logprob measure are higher when the likelihood is lower: a completely expected word with no competitors scores closest to zero, while strongly unexpected words receive a high value. The statistical tests that were used to identify any significant differences between groups of sentences and/or position, were also applied to the logprob measure.

Unlike the switch data, the logprob data were influenced by one variable only: the choice of verb. Sentences with a frequent verb have a perplexity score that is closer to zero for the subject than sentences with an infrequent verb (note that the subject is the first word after the verb, ($t_{(22)} = 3.21$, $p < .01$). This indicates that the subject is more expected after the frequent verb than after the less frequent verb. Note again, that the infrequent verb is especially infrequent in combination with the FAP construction, and not infrequent in and of itself.

Similarly to the switch data, there are no significant differences in word perplexity for sentences with or without FAP interpretation. Unlike the switch data, any visible effects of garden path sentences are absent.

Research question 3: human switch data and probability measure

The stochastic model provides us with a probability measure for each word, and the switch data summed over all participants present a numerical indication of processing units. These two measures turn out to correlate quite strongly: Pearson's correlations are significant ($p < .05$) for 19 out of 24 sentences (r ranging from .518 to .841), with correlations for a further 4 sentences higher than .400. The remaining sentence has a correlation of .249. We may remind ourselves that the probability measure is based on the preceding three words only. This means that for the first word, particularly, overall frequency is the only guideline. Near the end of each sentence, the switch data go down a lot, due to the experimental design (see above), whereas there is no such help in the selection of the next word for the model. Given these restrictions, the correlations are reasonably strong, and indicate that the participants were more likely to remember the next word when this word was a probable word. We will further discuss the similarities and differences in the final section.

4.3 Conclusion and discussion

The experimental task that the participants performed provided evidence about the psycholinguistic reality of the FAP sequences and the FAP constructions that were tested.[9] Sentences that contained a verb that frequently co-occurs with the FAP sequence, were apparently easier to

[9] Note that this is not the same as saying that the representation in Figure 1 is psycholinguistically real: that figure abstracts over all FAP sequences in one highly general knowledge representation, whereas we tested five *specific* FAPs.

remember (i.e. fewer switches after the subject were observed). Moreover, if the sentence contained a collocate verb, but no FAP interpretation, or vice versa, this led to more switches later on in the sentence, a finding which suggest the participants experienced a garden path effect.

Clearly, the FAP sequence is not word-like for all children and all sentences: there are sometimes switches between the adjective and the preposition. The results do suggest, however, that the sequence is more likely to be processed as a unit when it is preceded by a verbal collocate. In our opinion, this result is most compatible with a view of 'units' as a gradient phenomenon: sequences are more or less unit-like, and this is influenced by distributional patterns, i.e. co-occurrence frequency, and by constituency.

Our stochastic measure, reflecting the probability of the next word given the three preceding ones, was also significantly influenced by the main verb. In contrast to the switch data, however, no garden path effect was found. This lack of effects can be attributed to the fact that the stochastic model bases its estimates on a local window of three words, meaning that it is oblivious to dependencies spanning beyond this width. Standard stochastic language models are known to be empirically limited by a local window of three or four neighboring words (Jelinek, 1998), beyond which observations become too sparse, and estimates too unreliable. In addition to its limitations in locality, the perplexity measure is also based on overall frequencies and co-occurrence patterns alone: effects of (recent) context such as activation and decay are not part of the model. In human processing, this is akin to only having a long-term memory. Recency and priming effects are not captured by the present model. There is ongoing work on more flexible language models (e.g. Guthrie et al., 2006) that could prove useful here – this is a departure point for future research.

Yet, the human switch data and the probability measure are not unrelated. For most of the sentences, the two measures correlated quite strongly in spite of limitations in the comparability (switches taper off near the end of the sentence, probability is quite low for any first word of a sentence). A more detailed comparison of the two measures reveals, in addition to the similarities (see also the two dendrograms in Figures 4.4 and 4.5), two points of divergence. These differences concern prepositional phrases and the finite verb-subject sequence.

The 24 test sentences contain a total of 71 prepositional phrases (PPs), each consisting of a preposition, a determiner (sometimes an adjective) and a noun. In the switch data, for only 14 PPs the number of switches between the preposition and the determiner was lower than that between the determiner and the noun. The stochastic measure showed a lower value for the preposition-determiner sequence than for the determiner-noun sequence 28 times. This seems to indicate that the noun phrase within the PP was sometimes more unit-like for the participants than would be expected on the basis of the stochastic measure (see Figure 4.6 below for the comparison).

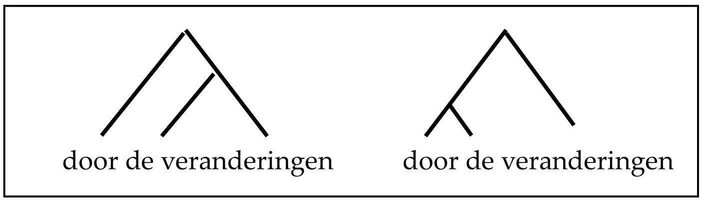


Figure 4.6: Clustering an NP within a PP (left) vs. a PREP+DET sequence within a PP (right)

The sequence of two relatively frequent closed-class items (preposition + determiner) is often associated with a high probability: the combinatorial possibilities are quite limited. This can be taken as a weakness of the stochastic measure, but at the same time reflects the intrinsic ambiguity between regarding the preposition-determiner pair or the determiner-noun pair as the primary unit; it could be argued that these two readings may exist in parallel. This would be another reason to expect gradient aggregated switching behaviour rather than all-or-nothing.

The second point of divergence concerns the sequence of two open class words: the verb and the subject. In the test sentences, the subject was always a proper noun. Although this is a theoretically infinite set, the word boundary between the verb and the subject was never a more frequent switch point than either the word boundary before or after this sequence in the switch data. In other words, for all 24 test sentences, more participants switched before the verb and after the subject than between these two words. The perplexity measure does not reflect this completely: in one third of the sentences (8 times) the logprob value for either the preceding or the following word is closer to zero. It is not unexpected that there are no clear predictions at this point of the sentence from the memory-based model. For the participants in our experiment, however, the sequence was very much part of one unit in memory.

The copy task proved to be an informative task; the switch data show clear effects of co-occurrence patterns. An attempt to replicate the experiment with adult participants, however, failed: the task proved so simple that adult participants hardly ever switched (less than 1.5 switches per sentence). Apparently, having to remember sentences of the type that was used in this experiment does not tax their memory enough to cause switches. Increasing the task demands is one possible way to make this experiment more difficult for adults. Ehrismann (2009) did this by adding a secondary task: participants had to add a second sentence to each utterance they had copied. Other options include adding distracting sound or limiting view time to the original sentence.[10]

4.3.1 Processing FAPs and the MultiRep model

The broader question behind the comparison between experimental data and a measure that is based in frequency and co-occurrence patterns is to what extent human sentence processing reflects these distributional patterns. Given the caveats that were discussed earlier, we feel that the correlations are encouraging. Although correlations cannot simply be equated to causal relations, in this case we feel justified to suggest that the distribution-based *logprob* measure reflects one of the contributing factors to the switch data. Thus, the simple stochastic model explains more than 25% explained of the variance in switches, which is quite an achievement.

Earlier in this chapter, we indicated that our experimental task ranks rather low on the 'naturalness scale' that Gilquin and Gries (2009) introduced. At the 2009 Corpus Linguistics conference in Liverpool a workshop with the theme of 'converging and diverging evidence' was organized. The participants at this workshop each compared corpus-based frequency measures to human language use. Interestingly, tasks that scored relatively high on the naturalness scale, such as defining different word senses or recognizing neologisms (Littlemore & MacArthur, 2009; Svanlund, 2009), correlated less strongly with frequency measures than the more unnatural task of a word class identification task (Teddiman, 2009), which is also not normal language use.

In this case, 'naturalness' may not be the most appropriate scale to place these tasks on. The word class identification task and the copy task we report on here have in common that they are online tasks: they measure language processing as it happens. In word definition tasks, on the other hand, people have time to reflect on their answers, or at least do not have to instantly decide or act. We suggest that it is this difference that is reflected in

[10] We are grateful to the audience at the Fifth Corpus Linguistics Conference in Liverpool (July 2009) for this suggestion.

a higher (online tasks) or lower (offline tasks) correlation with frequency measures. Time allows people to not only rely on frequency, but to take other factors (possibly pragmatics, larger textual and co-textual contexts, and metalinguistic knowledge etc.) into account.

Seen on the gradient scale of time allotted to test subjects, from immediate to indefinite, our task is not among the most immediate online tasks: participants read the sentence first before they started the copying process. We believe that this is the reason we find no significant correlation of our simple stochastic metric and the presence or absence of the FAP construction – while at the same time we do observe a quite strong overall correlation. Arguably, the room that our task grants to participants to think about the sentence causes divergence of their behaviour with respect to the simple stochastic model.

This chapter reported on an online task in which the participants sometimes did and sometimes did not process instantiations of the FAP construction as a whole. Whether the sequence was broken up or not was significantly influenced by its context: in contexts likely to co-occur with an instantiation of the construction, it was most often not divided. In our opinion, the fact that context plays a significant role underlines the need to incorporate contextual knowledge as part of our representation of lexical items – complex or not. Much like the definition task and the word-formation task described in Chapter 2, performance varied: not all instantiations were processed as wholes by all participants. This pattern is similar to what we found in the experiments discussed earlier: not all definitions in the word definition task contained references to the morphemic elements, and in the word formation task, correct word formations with a particular affix in a particular word did not reliably predict correct responses to other items with the same affix in the word formation task. This variation in performance is interesting. The experimental data show that it is not completely random: it is at least partially determined by context and by frequency.

Describing the participants' performance in terms of the MultiRep model that was introduced in the first chapter, it is highly plausible that participants employed different levels of representations. Note that with the current experimental design we can never be certain that a sequence is in fact analyzed as a whole: when there is a switch between the adjective and the preposition, this is proof that the sequence was broken up into two parts. When there is no switch, however, the sequence may or may not be stored as a whole. It can still be the case that it is remembered as two units that are stored together in short-term memory (just like any other sequence of words). The fact that switch behaviour is not random, though, makes this a less likely interpretation.

Both unit-like processing of the FAP sequence and word-by-word storage are possible in the MultiRep model. The results of the copy task show that the likelihood of unit-like processing is influenced by both the context of the sequence and its interpretation. The MultiRep model suggests that collocates are part of the representation of units in language, which seems to be corroborated by the data presented here. The observation that interpretation is influential too (cf. the garden path effect) is also predictable from the model, as this is based on form-meaning pairings, making the semantics an inseparable part of the representations.

In sum, we feel that the results of our experimental task provide converging evidence for the relevance of frequency in human processing. At the same time, however, the data show that people take into account more aspects of structure and meaning than can be captured by the current simple stochastic model. The MultiRep model that was introduced at the end of the first chapter of this book can accommodate these findings.

Acknowledgements: The software program for this experiment was written by Michiel van der Ros. For more information about this program, please contact the author.

Chapter 5:

Discussion

5.0 Introduction

The first chapter of this book introduced the MultiRep model for CLIs, short for model of multiple representations for complex lexical items. In this introduction, I presented the view that children start acquiring CLIs and the patterns that instantiate them by learning specific form-meaning pairings, as language learning is usage-based and proceeds in a bottom-up fashion. In assuming a bottom-up generation of incrementally more abstract representations, I follow cognitive linguists such as Tomasello (2003) and Diessel (2004). Following the ideas of Langacker (1987, 1991, 2008), Goldberg (1995, 2006) and Croft (2001), I reject the strict distinction between lexicon and grammar and instead assume that people have a *constructicon*, consisting of networks of specific lexical items and more abstract patterns and generalizations.

As children encounter different instantiations of the same pattern, they may come to recognize similarities in form and meaning. This stimulates the formation of (partially) abstract representations. Lexically specific and abstract representations are assumed to co-exist, with links between stored elements on the basis of co-occurrence patterns and overlap in form and/or meaning. The MultiRep model is designed to reflect both the specific and the abstract knowledge that speakers have and does not aim for the elegant abstractions a linguist may detect (cf. Boas, 2008). This bottom-up perspective entails that it is abstract representations that must be proven to be cognitively real; in other words: if we assume that speakers tend to store and use lexically specific sequences whenever possible (cf. Wray, 2002; Dąbrowska, 2008), this is the default option, and the burden of proof is on detecting the use of abstract representations.

Five different experimental techniques were employed in addition to corpus analyses to investigate speakers' knowledge and use of specific and abstract representations. In what follows below, I summarize the results of these experiments and analyses and discuss the extent to which they are compatible with the MultiRep model. In doing so, I will also point out promising directions for further research, and identify issues that, in my view, deserve more attention.

5.1 The MultiRep model

In the MultiRep model, the emphasis is on the co-existence of representations at different levels of specificity. At the initial, basic level, these representations only contain a single lexically specific element. This can be either a monomorphemic word or a larger unit that is stored without any internal structure in its representation. Those larger units may contain more than one form-meaning element, but the fact that a linguist can recognize different morphemes does not mean a speaker necessarily has to store these elements separately (see also the discussion sections in earlier chapters). More complex representations will contain underspecified obligatory elements or likely but not obligatory elements. In this view, a construction is a form-meaning pairing that contains one or more lexically specific or underspecified elements, which may or may not be obligatory elements of the construction. At first sight, this definition might seem very general and this is done intentionally: it covers *all* stored form-meaning pairings which together constitute a speaker's constructicon or linguistic repertoire. The MultiRep model assumes that speakers have representations available to them at varying degrees of specificity. Recognizing that elements of a representation can be either lexically specific or underspecified and obligatory or likely is useful for the interpretation of the experimental results that were presented in earlier chapters.

Lexically specific vs. underspecified elements, obligatory vs. likely elements

All five experiments investigated knowledge about constructions with at least one lexically explicit element. This is what sets complex lexical items and the patterns they instantiate apart from fully underspecified argument structure constructions such as the ditransitive construction: CLIs contain a lexically explicit element (hereafter Type A element) also at abstract levels of representation. The derivational affix constructions that featured in the word formation task (Section 2.2) each contain a lexically fixed element: the affix. The results showed that children mainly use their knowledge of specific instantiations to solve this task. The relative frequency of the test items and vocabulary size of participating children (mean age 9;4) were strongly correlated with performance on the task. The test items in the word definition task (Section 2.3) were of very low frequency and therefore probably new to the participants. The results proved that the children are aware of morphological structure, although they are more likely to refer to elements of a complex word in general (66% of responses, including references to the affix) than specifically to the affix (33%). In this task, references to the affix can only be explained if knowledge about a partially abstract representation is used.

Affix constructions by definition have an obligatory slot, because the affix cannot occur without a stem. This is the second type of elements that may occur in a representation: underspecified slots that need to be filled in actual instantiations. It is possible to define the types of lexical elements that can fill these obligatory slots (hereafter Type B elements). The characteristics of potential slot fillers may be more or less specific, ranging from basic word class (e.g. the plural suffix in English is preceded by a count noun) to semantically more narrowly delimited descriptions, e.g. the argument expressing the acting participant in instantiations of *lachen* 'laugh' has a human or human-like referent. Sometimes the characteristics are very specific; in the IS TE V construction (Chapter 3) a form of the paradigm of *zijn* 'be' has to be part of any instantiation. An element in a representation is clearly of Type A when the string of letters is always the same, and clearly of Type B when a potentially endless set of lexical items can fill the slot, as for the English plural construction. In cases where the open slot is filled with an element from a more specifically defined set, e.g. forms of a verbal paradigm, the distinction between Type A and Type B becomes gradual. The categories A and B meet when there is a closed subset of possible slot fillers. Constructions are productive if speakers have a representation of the pattern with (at least) one Type B element in it. This means that they can form new instantiations of the pattern.

In the corpus analyses for the V–BAAR and IS TE V constructions, distributional patterns and collostructional analyses showed some clear characteristics of slot fillers in instantiations of both constructions. Results of a magnitude estimation task provided evidence, moreover, that adults and children (mean age 12;3) are sensitive to these characteristics: they rated instantiations with some types of verbs consistently higher than others. These differences in acceptability can only be accounted for by assuming that the participants have representations of the constructions that contain information about the characteristics of the slot fillers.

The third type of elements a representation may contain is the non-obligatory element (Type C element). An example of these is the adverbial modifier that is very frequent in instantiations of the IS TE V construction. Many of the occurrences contain a lexical expression of the degree of possibility, such as *niet* 'not', *best wel* 'kind of' and *prima* 'fine'. Looking at the IS TE V construction in general, this Type C element is underspecified: it is a degree-marking adverbial modifier. For a number of specific IS TE - verb combinations, specific adverbs are found to collocate, e.g. *is niet te doen* 'is hard to do', literally 'is not to do'. Within the category of Type C elements, both lexically specific and underspecified elements may be found. In contrast to elements of type B, a Type C element does not occur in all instantiations of the pattern. Much like the Type A - Type B distinction, the boundary between Type B and Type C is graded. Type B can be seen as a special case of type C: the likelihood of the element to occur in some shape or form is one hundred percent. Some types of the IS TE V construction have a lexically fixed marker of degree of possibility (e.g. *niet te harden* 'unbearable' only occurs in the negated form). For this instantiation of the general IS TE V construction the adverb *niet* 'not' is a Type A element. The results of the copy task discussed in Chapter 4 provide evidence that language users are sensitive to these non-obligatory elements. In the copy task, participants were asked to read a sentence on a computer screen, press a button and recreate that same sentence. Whenever they were unsure about the following word, they could look back at the original sentence. This look-back behavior showed that a FAP, a fixed adjective-preposition sequence (e.g. *trots op* 'proud of'), required less memory capacity if it was preceded by a verb it frequently co-occurs with than if the verb is less commonly found with this sequence. The presence of a non-obligatory collocate aids in the ease with which the sequence is (briefly) stored in memory.

The co-existence of representations at various levels of specificity

The MultiRep model implies that speakers have representations at different levels: of completely fixed instantiations, i.e. only consisting of Type A elements, of partially specific schemas (Type A and Types B/C), and of abstract patterns (only Types B and C). The distinction I just sketched between these three types is not commonly made. At least two independent tendencies are likely causes of a lack of attention for this three-way distinction.

First, part of the distinction resides in the recognition of fixed instantiations of a construction (see Sections 3.2.1.5 and 3.2.2.5 for examples with the V-BAAR and IS TE V constructions) and in analyses of sentence constructions that contain a fixed word (e.g. the V-*ing* *TIME away* construction). Analyses within the Construction Grammar (CxG) framework are usually centered either around a construction with very little if any lexically fixed elements (e.g. the ditransitive construction, Goldberg, 2006; Barðdal, 2007; and the resultative construction, Boas, 2003; Goldberg & Jackendoff, 2005) or around a verb and its senses with concomitant argument structure(s) (e.g. *eat* and *drink* in Newman & Rice, 2006). Hitherto, smaller constructions such as derivational word formation have not received much interest within this framework. This is a surprising omission, because CxG's basic assumptions (see Section 1.4.1) seem to lend themselves especially to the study of these types of constructions in which both fixed and underspecified elements play an important role.

Second, elements that are likely but not obligatory in instantiations of a construction have not been studied much at all. From a psycholinguistic perspective, it is not surprising that such elements exist, and that their presence speeds up recognition of a pattern: they are collocations at the construction level rather than at the word level. Corpus analyses of specific constructions indicate that certain extensions are more likely than others, e.g. a marker of the degree of potentiality in the V–BAAR and IS TE V constructions and the collocate verb in the Fixed Adjective Preposition constructions. The actual co-occurrences are semantically motivated, although specific combinations need not be completely transparent to a language user.

Formalizing representations

The descriptions of characteristics of underspecified elements in representations may come across as rather informal. The aim of the MultiRep model is to describe speakers' knowledge of constructions as it is reflected in what occurs in a language and how they respond to linguistic stimuli. Whereas highly schematic representations are 'elegant' and 'economical' in theoretical terms (Chomsky, 1995) and computationally (Rissanen, 1983) these schemas do not necessarily correspond to the knowledge that speakers of a language actually possess, or to the specific distributions of instantiations of constructions. In defining characteristics of underspecified elements, I have deliberately stayed away from the notion of restrictions on slot fillers, because speakers' knowledge is based on what *does* occur. Also, radically excluding certain types of elements as possible slot fillers would often be too strict for a realistic description of a construction: people stretch the characteristics of slots in their creative use. An example is the creative formation of *busbaar* 'bus-able' for 'reachable by bus' (see Section 3.2.1.4). As an aside, it is probably useful to note that the notion of 'slot fillers' seems to imply that a construction is chosen first, and that slots are filled in a second step. This terminology suggests that abstract constructions often take precedence: they are selected first after which slots are filled. In real language use, elements of different size and different degrees of specificity will be activated alongside each other.

5.2 Identifying CLIs and the patterns they instantiate: The researchers' perspective

For research purposes, it is useful to maintain the distinction between instantiations and underlying patterns (descriptive linguistic notions). In terms of speakers' linguistic knowledge these notions translate as lexically specific sequences and (partially) abstract representations. In spontaneous language use it is nigh on impossible to determine whether an uttered sequence has been produced as a stored chunk or by combining an abstract representation and (a) slot filler(s). The only exceptions are novel creations that are explicitly marked (e.g. *...customisable if that's a word*, see Chapter 2.1.1). There are, however, elements that make it unlikely that something is a productively formed instantiation: if the sequence is high in frequency and/or semantically not transparent, it will probably be stored as a unit in a speaker's constructicon. The corpus analysis of the V-BAAR and IS TE V constructions in Chapter 3 revealed that many instantiations showed so-called type specific behavior, which I take to be evidence that these specific instantiations are stored.

In all of the experiments discussed in this book, the focus was on constructions with at least one lexically specific or type A element. Elements of type A are not necessarily a component of all constructions. Argument structure constructions such as the ditransitive do not have any word or morpheme that is present in all instantiations. Even if a construction does have one or more fixed elements, these need not uniquely identify the construction. This is one of the differences in form between the V-BAAR and IS TE V constructions. Whereas the morpheme *–baar* is a very reliable indicator of the V-BAAR construction, infinitival *te* is not for the IS TE V construction. The notion of *cue validity* (Goldberg, 2006) thus quantifies the likelihood of dealing with construction Y, given lexical element X. In this respect, the affix *–baar* has a very high cue validity for the V–BAAR construction, whereas *te* does not for the IS TE V construction. While both constructions have a Type A element, one of these is unique to the construction (*-baar*) where the other (*te*) is not. This means that within CLI constructions, for the researcher it is not always equally easy to identify what constitutes an instantiation of the construction.

To what extent constructions with at least one type A element are really essentially different from fully underspecified patterns (e.g. DET. ADJ. N), remains an open question. Traditionally, the acquisition of these patterns has been studied separately from the learning of linguistic units with a fixed element: the former is syntax and the latter (more) lexicon. Under the assumption that all linguistic pattern learning starts with simple form-meaning pairs, there is no reason to maintain this distinction. The Construction Grammar framework offers ways to investigate the extent to which constructions with and without type A elements differ from each other.

The results of the five experiments discussed in this book suggest that speakers use different levels of representation for different tasks. This is not always a conscious and explicit choice; speakers may not control which levels they can use. In the first two experiments, the word definition and word formation tasks, results showed that the children who participated could not always make use of abstract levels of representation. The (slightly older) children who took part in the magnitude estimation tasks did employ abstract representations of the V-BAAR and IS TE V constructions in their responses on the magnitude estimation task. While all three tasks allow for conscious reflection and are offline, this indicates that giving people time means they can apply abstract levels of representation. This is important from the point of view of experimental research. It means that in testing and researching proficiency, if task demands influence what levels of representation are activated, task performance may over- or underestimate speakers' abilities. In research on second and foreign language acquisition, this issue is currently heavily scrutinized, both in attempts to operationalize implicit and explicit knowledge and in studies on the effects of task characteristics on task performance (Ellis, 2005; Hulstijn, 2005; Kuiken, Vedder & Mos, 2005; Van den Branden, Bygate & Norris, 2009). For any model of linguistic repertoire, this variation entails that we must assume that different levels co-exist, and that speakers do not have full control over which they use to execute a task. How these levels are connected and what exactly influences what levels are activated, are questions requiring further investigation.

5.3 Identifying CLIs and the patterns they instantiate: The language learners' perspective

For learners, input is essential. In the acquisition of their first language(s), children build up their linguistic repertoire from the language they encounter. This repertoire starts with the storage of unanalyzed form-meaning pairings, but as children come across multiple instantiations of a pattern, they may begin to recognize similarities between those instantiations. These then allow for the development of (partially) abstract patterns, which contain one or more open slots. This path of development has been described extensively for a number of aspects of language, such as the role of verbs in acquiring argument structure constructions (e.g. Tomasello's notion of *verb islands*, Tomasello, 1992; Casenhiser & Goldberg, 2005 for experimental research involving nonce verbs; Diessel, 2004 for the development of complex sentences).

The role of skewed distributions and prototypicality in acquiring constructions

Much emphasis has been placed on the beneficial effects that a skewed distribution of instantiations of a pattern can have (Goldberg, 2006; Ellis & Collins, 2009). Goldberg (2006, chapter 4) shows for argument structure constructions that the input children receive contains a strongly skewed distribution: one verb accounts for at least 20% in each of the different patterns she investigated, e.g. *give* in the ditransitive construction. In addition, the dominant verb is also prototypical in meaning for the pattern it instantiates: its semantics match very closely with the constructional meaning. Ellis and Ferreira-Junior (2009), on the other hand, show that for at least some argument structure constructions this is not the verb speakers identify as most representative for each construction. Both Ellis and Ferreira-Junior (2009) and Boyd and Goldberg (2009) discuss what may cause the differences here.

Ellis and Ferreira-Junior suggest that *"it will be important in future research to investigate a wide range of constructions in native language corpora (...) to assay the generality of Zipfian distributions within constructions"* (Ellis & Ferreira-Junior 2009:383). The detailed analyses in this book for the V–BAAR and IS TE V constructions in Chapter 3 are an example of such research. They reveal that, whereas the distribution of types over tokens is Zipfian, the most frequent types are not prototypical in meaning for the pattern they instantiate. Examples of frequent forms with the –baar affix include kostbaar, 'valuable' or blijkbaar 'apparently'. Although they look like instantiations of the general pattern, they deviate from the general pattern both in meaning and in form (neither of the two forms has a potentiality connotation and blijkbaar does not occur predicatively). What does this mean for the acquisition of the pattern? For the development of a partially abstract representation, types with low token frequency are more useful, as they do conform to the meaning and form characteristics. It stands to reason that constructions are difficult to acquire when the most frequent apparent instantiations diverge from the general pattern. For the V-BAAR and IS TE V constructions it is possible that high type frequency compensates for the lack of information value the high token types provide. Bybee (1995) suggests that for regular morphology high frequency types do not contribute to the productivity of the pattern they instantiate. Hay and Baayen (2002, 2005) argue that morphologically complex words that are relatively frequent, i.e. have a high frequency compared to that of the stem they contain, are most likely to be stored separately. Taken together, this implies that high frequency instantiations of a pattern are not particularly useful to acquire a representation of the underlying pattern. As they are likely to be stored separately, they will show signs of semantic drift, i.e. have their own particular meaning or connotation. This is pretty much the opposite of the claim made by Goldberg (2006) about the role of the most frequent verb in the argument structure constructions she investigated. Since Hay and Baayen and Bybee deal with morphological patterns, and Goldberg with argument structure constructions, the question becomes whether the existence of a highly frequent and formally and semantically prototypical instantiation is relevant for argument structure constructions' distributions but not for morphological constructions' distributions. The corpus analysis of the IS TE V construction, though, suggests that at least for some larger constructions the most frequent instantiations are not prototypical. Clearly, more research is necessary here, in which the distribution of types and tokens in a construction and the prototypicality of high frequency instantiations is investigated.

The role of schooling in acquiring constructions

Performance on the experiments described in this book points out that for the tasks requiring explicit knowledge and meta-linguistic awareness, the nine-year-olds have not reached adult competence yet. Experiments such as the word definition task and the word formation task call for segmentation of sequences in meaning-carrying elements. Wray (2002, 2008) suggests that in first language acquisition input is only broken up into meaningful segments when children have a need to do so: *"nothing [is] broken down unless there [is] a specific reason"* (Wray 2002:130). She calls this the Needs Only Analysis, and says that this is where L1 and L2 acquisition differ: speakers who learn a second language tend to break sequences up into smaller parts and to produce utterances by putting separate elements together. This is why fixed combinations such as FAPs (e.g. *trots op* 'proud of') are not problematic in mother tongue acquisition, but very difficult for L2 learners. For linguistic constructions that are acquired relatively late, typically after children have learned to read, formal education may play a role. In school, explicit attention is paid to forms and structures. Reading allows for metalinguistic reflection, which is stimulated by school through exercises, structuring of the input etc. Perhaps this makes late L1 acquisition more L2-like. The question of the extent to which formal attention to language in school influences development of (abstract) representations needs further investigation.

The role of input in acquiring constructions

Taking input as the basis for acquisition, which is the starting point for the MultiRep model, the expectation is that not all speakers will develop exactly the same general pattern: the input differs from person to person, both quantitatively and qualitatively. This is not necessarily a problem: earlier (Section 3.1.1) it was discussed that for the occurrence of an instantiation it cannot be determined whether this was produced from storage (the whole instantiation is part of the speaker's linguistic repertoire) or formed productively (the speaker invokes the general pattern and (a) lexical item(s) for the underspecified element(s)). In language use, communication can be successful regardless of whether both speakers have the same general patterns or not, as long as the uttered instantiations can be recognized as meaningful.

The analyses in Chapter 2 showed that the size of children's linguistic repertoire was an influential factor for performance on the tasks. There are significant differences in vocabulary size, with bilingual children on average scoring at the same level as monolingual children who are one year younger. If we assume that input is the essential determinant of linguistic development, this is not surprising. It is highly likely that the bilingual children have had less input in Dutch than their monolingual peers. Bilingual children's performance on the experimental tasks was structurally similar; they did as well as the monolingual children when the differences in vocabulary size were corrected for. Although the observed differences in vocabulary size do not come as a surprise, they are not unproblematic: it is an undesirable state of affairs in a school system where performance in Dutch is expected and tested, and in which all text books assume certain linguistic proficiency levels. While the research questions in this book were not aimed at providing a solution for this problem, the data suggest that offering substantial quantities of contextually rich input with a variety of different instantiations of a pattern is necessary for acquisition.

Although the data suggest that frequency is an important factor in language acquisition, there are reasons to assume that language learning is not just stochastic: salience is likely to play a part as well (see Section 2.1.1.2

and also Wulff et al., 2009, who observed that acquisition orders in adult learners of English were best predicted by combining frequency, form and function in the input). The comparison between the stochastic word perplexity measure and switch behavior in the copy task (Chapter 4) showed that a significant part of the 'chunking' that participants did can be explained by the likelihood measure, but not all. The stochastic measure did not incorporate the garden path effects the participants in the experiment produced, and some word sequences were much more unit-like with one measure than the other: the perplexity level is particularly low for sequences of two frequent closed-class words, such as preposition-determiner, and relatively high for finite verb-subject sequences, which was a strong unit for the participants in the experiment.

Without a doubt, input is an absolute prerequisite for the acquisition of constructions. Because different children will receive different input, the representations will not be the same for each child. The performance on the experiments described throughout this book showed that there is much individual variation. What is interesting is that this variation is especially prominent if knowledge is tested of individual items (e.g. in the word formation and word definition tasks, both Chapter 2). The tasks that measured more abstract representations (magnitude estimation, Chapter 3) and the structure of the constructicon (family size effect in the lexical decision task, Chapter 2) mainly showed similarities. In the lexical decision task, there were large individual differences in reaction times that correlated strongly with children's vocabulary size. The presence of a family size effect was not related to vocabulary size. This raises the question whether in a usage-based account it would be reasonable to expect large individual differences at the level of specific representations, but more similarities at abstract levels. At present, there are not enough data to go beyond this speculative hypothesis, but this is clearly a point that should be pursued in further research.

5.4 Conclusion

In this book, I have taken a usage-based approach to language. Construction Grammar and, at a more general level, cognitive linguistics, was chosen as a theoretical framework, because it comfortably accommodates the linguistic phenomenon that is the topic of this book. By combining corpus analyses and experimental research, I have attempted to shed more light on the question what people know about complex lexical items and the patterns they instantiate.

Reviewing the choices in the experimental designs, several potential improvements could be identified. First, some of the experimental tasks did not contain a sufficient number of items to make general claims possible about specific affix constructions. Second, it would have been interesting to test the same children at different points of their linguistic development. Denser and longitudinal data collection is necessary to answer some of the questions that the present work has raised. Other voids cannot be filled easily: the Corpus of Spoken Dutch (CGN) is a useful tool, but too small for any research question pertaining to relatively infrequent phenomena, and may not be similar enough to the linguistic input that children get, for which no reasonably large corpus is available in Dutch. A theoretical problem lies in the identification of multi-word linguistic units: both the analyses of the constructions in CGN and the experimental data seem to point out that these larger units are cognitively real, but it remains difficult to formulate criteria for identification.

In spite of these limitations, the data and analyses have shown their merit, particularly in bringing to light a developmental pattern in older children's constructicons. Some of the experimental techniques have rarely (lexical decision tasks) or never (magnitude estimation) been used with children as participants. The performance on these tasks shows that children are in fact able to execute them and that these experiments can provide useful information about their language skills. Combining corpus analyses and experimental research has been a fruitful exercise; showing that people have detailed knowledge about the possible slot fillers and semantics of the V-BAAR and IS TE V constructions, for example, would not have been possible without these two approaches.

In sum, bringing together aspects from different subfields within linguistics −the theoretical stance from cognitive linguistics, data-driven analyses from corpus linguistics, the research traditions from applied linguistics and psycholinguistics− leads to further insights in language as people use it. Within each of these disciplines, I believe that complex lexical items and the patterns that they instantiate deserve more attention in further research.

References

Achard, M. & Niemeier, S. (2004). *Cognitive linguistics, second language acquisition, and foreign language teaching*. Berlin: Mouton de Gruyter.

Albright, A. (2002). Islands of reliability for regular morphology: Evidence from Italian. *Language, 78,* 684-709.

Albright, A. & Hayes, B. (2003). Rules vs. analogy in English past tenses. A computational/experimental study. *Cognition, 90,* 119-161.

Allegre, M. & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language, 40,* 41-61.

Anderson, R. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.

Anglin, J. M. (1993). *Vocabulary development: A morphological analysis*. Chicago, IL, IL: University of Chicago Press.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Aronoff, M. (1994). Morphology by itself; Stems and inflectional classes. Cambridge, MA: MIT Press.

Aronoff, M. & Anshen, F. (1998). Morphology and the lexicon: Lexicalization and productivity. In A.Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 237-247). Oxford: Blackwell.

Baayen, R. H. & Lieber, R. (1991). Productivity and English derivation: A corpus based study. *Linguistics*, 29, 801-843.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database [Computer software]. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. Journal of Memory and Language, 37, 94-117.

Baayen, R. H. & Schreuder, R. (2003). *Morphological structure in language processing*. Berlin: Mouton de Gruyter.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412.

Baayen, R. H. (2008). Corpus linguistics in morphology: Morphological productivity. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 899-920). Berlin: Mouton de Gruyter.

Backus, A. (2003). Units in code switching: Evidence for multimorphemic elements in the lexicon. *Linguistics, 41,* 83-132.

Balota, D. A. (1990). The role of meaning in word recognition. In D.A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 9-32). Hillsdale, NJ: Lawrence Erlbaum.

Bard, E. G., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language, 72,* 32-68.

Barðdal, J. (2007). The semantic and lexical range of the ditransitive construction in the history of (North) Germanic. *Functions of Language, 14,* 9-30.

Barðdal, J. (2008). *Productivity; Evidence from case and argument structure in Icelandic*. Amsterdam: John Benjamins.

Bauer, L. (2001). *Morphological productivity*. Cambridge: Cambridge University Press.

Benelli, B., Belacchi, C., Gini, G., & Lucangeli, D. (2006). 'To define means to say what you know about things': The development of definitional skills as metalinguistic acquisition. *Journal of Child Language, 33,* 71-97.

Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150-177.

Bertram, R., Schreuder, R., & Baayen, R. H. (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 489-511.

Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language, 42,* 390-405.

Biber, D. (1999). Longman grammar of spoken and written English. Harlow: Longman.

Blom, E. (2003). From root infinitive to finite sentence : The acquisition of verbal inflections and auxiliaries. Utrecht: LOT.

Boas, H. C. (2003). *A constructional approach to resultatives*. Chicago, IL: University of Chicago Press.

Boas, H. C. (2005). Determining the productivity of resultatives: A reply to Goldberg and Jackendoff. *Language, 81,* 448-464.

Boas, H. C. (2008). Determining the structure of lexical entries and grammatical constructions in Construction Grammar. *Annual Review of Cognitive Linguistics, 6,* 113-144.

Bod, R. (2009). Constructions at work or at rest? *Cognitive Linguistics, 20,* 129-134.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10,* 245-261.

Boogaart, R. (2006). "Het is te zien hoe dat je het ziet" De modale infinitief in Nederland en België ["It is to see how that you see it" The modal infinitive in the Netherlands and Belgium]. *Voortgang, jaarboek voor de neerlandistiek, 24,* 37-50.

Booij, G. (2002). *The morphology of Dutch*. Oxford: Oxford University Press.

Booij, G. (2005). Compounding and derivation: Evidence for Construction Morphology. In W.U. Dressler, D. Kastovsky, O. E. Pfeiffer, & F. Rainer (Eds.), *Morphology and its demarcations* (pp. 109-132). Amsterdam: John Benjamins.

Booij, G. (2009). Construction morphology and compounding. In R. Lieber & P. Štekauer (Eds.), *The Oxford Handbook of Compounding* (pp. 201-216). Oxford: Oxford University Press.

Booij, G. (n.d.). *Recycling morphology: Case endings as markers of Dutch constructions.* Retrieved from http://www.hum2.leidenuniv.nl/booijge/publicaties.htm.

Boon, T. D. & Geeraerts, D. (2005). Groot woordenboek van de Nederlandse taal, 14e edition [Dictionary of the Dutch language, 14th edition]. Utrecht: Van Dale Lexicografie.

Borsley, R. D. & Newmeyer, F. J. (2009). On Subject-Auxiliary Inversion and the notion purely formal generalization. *Cognitive Linguistics, 20,* 135-143.

Bosman, A. M. & De Groot, A. M. B. (1996). Phonologic mediation is fundamental to reading: Evidence from beginning readers. *Quarterly Journal of Experimental Psychology, 49A,* 715-744.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Automatic semantic activation of embedded words: Is there a "hat" in "that"? *Journal of Memory and Language, 52,* 131-143.

Boyd, J.K., & Goldberg, A. E. (2009). Input effects within a constructionist framework. *Modern Language Journal, 93*, 418-429.

Brooks, P. J. & Tomasello, M. (1999). How children constrain their argument structure constructions. *Language, 75*, 720-738.

Brown, R. W. (1973). *A first language: The early stages.* London: Allen & Unwin.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language production. Volume 2: Development, writing and other language processes* (pp. 257-294). London: Academic Press.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425-455.

Bybee, J. & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics, 37,* 575-596.

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65-85). Chicago, IL: University of Chicago Press.

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition, 24,* 215-221.

Bybee, J. & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22*, 381-410.

Carlisle, J. F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics, 9,* 247-266.

Carlisle, J. F. & Nomanbhoy, D. (1993). Phonological and morphological awareness in first graders. *Applied Psycholinguistics, 14,* 177-195.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading & Writing: An interdisciplinary journal, 12,* 169-190.

Casenhiser, D. & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science, 8,* 500-508.

Castles, A., Davis, C., & Letcher, T. (1999). Neighbourhood effects on masked form priming in developing readers. *Language and Cognitive Processes, 14,* 201-224.

Chipere, N. (2003). Understanding complex sentences: Native speaker variations in syntactic competence. Basingstoke: Palgrave.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Clahsen, H. & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27,* 3-42.

Conklin, K. & Schmitt, N. (2008). Formulaic Sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29,* 72-89.

Croft, W. (2001). Radical construction grammar; Syntactic theory in typological perspective. Oxford: Oxford University Press.

Croft, W. (2005). Logical and typological arguments for Radical Construction Grammar. In J.O. Östman & M. Fried (Eds.), *Construction Grammars* (pp. 273-314). Amsterdam: John Benjamins.

Croft, W. (2009). Constructions and generalizations. *Cognitive Linguistics, 20,* 157-165.

Dąbrowska, E. & Lieven, E. (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics, 16*, 437-474.

Dąbrowska, E. & Street, J. (2006). Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences, 28,* 604-615.

Dąbrowska, E. (2008). The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language, 58,* 931-951.

Dąbrowska, E. (2009). Words as constructions. In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 201-223). Amsterdam: John Benjamins.

Davis, C., Castles, A., & Iakovidis, E. (1998). Masked homophone and pseudohomophone priming in children and adults. *Language and Cognitive Processes, 13,* 625-651.

De Jong, N. H., Schreuder, R., & Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes, 15,* 329-365.

De Jong, N. H., Schreuder, R., & Baayen, R. H. (2003). Morphological resonance in the mental lexicon. In R.H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 65-88). Berlin: Mouton de Gruyter.

Deacon, S. H. & Kirby, J. R. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics, 25,* 223-238.

Deacon, S. H. & Bryant, P. (2005). The strength of children's knowledge of the role of root morphemes in the spelling of derived words. *Journal of Child Language, 32,* 375-389.

Den Ouden, H., Noordman, L., & Terken, J. (2009). Prosodic realizations of global and local structure and rhetorical relations in read aloud news reports. *Speech Communication, 51,* 116-129.

Derwing, B. L. (1976). Morpheme recognition and the learning of rules for derivational morphology. *Canadian Journal of Linguistics, 21,* 38-66.

Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming: Unraveling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes, 20,* 75-114.

Diepeveen, J., Boogaart, R., Byloo, P., Janssen, T., & Nuyts, J. (2006). *Modale uitdrukkingen in Belgisch-Nederlands en Nederlands-Nederlands: corpusonderzoek en enquête [Modal expressions in Belgian-Dutch and Dutch-Dutch: corpus research and questionnaire*]. Münster: Nodus Publikationen.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge: Cambridge University Press.

Divjak, D. & Gries, S. Th. (2008). Clusters in the mind? Converging evidence from near synonymy in Russian. *The Mental Lexicon, 3*, 188-213.

Doğruöz, A. & Backus, A. (2009). Innovative constructions in Dutch Turkish: An assessment of ongoing contact-induced change. *Bilingualism: Language and Cognition, 12,* 41-63.

Drews, E. & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 21,* 1098-1116.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2008). Does darkness lead to happiness? Masked suffix priming effects. *Language and Cognitive Processes, 23,* 1002-1020.

Eddington, D. (2000). Analogy and the dual-route model of morphology. *Lingua, 110,* 281-298.

Ehrismann, M. (2009). Eén enkele eenheid? De wijze waarop gefixeerde expressies in het lexicon zijn opgeslagen [One single unit? The way in which fixed expressions are stored in the lexicon]. Unpublished MA-thesis, Tilburg University, The Netherlands.

Elbers, L. & van Loon-Vervoorn, A. (2000). Lexicon en semantiek. In S. Gillis & A. Schaerlaekens (Eds.), *Kindertaalverwerving: Een handboek voor het Nederlands* (pp. 185-224). Groningen: Martinus Nijhoff.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*, 305-352.

Ellis, N. C. (2006). Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review, 19,* 100-121.

Ellis, N. C. & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function. Introduction to the special issue. *Modern Language Journal, 93,* 329-335.

Ellis, N. C. & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal, 93,* 370-385.

Ellis, R. (2006). Modeling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics, 27,* 431-463.

Erman, B. & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20,* 29-62.

Feldman, L. B. (2000). Are morphological effects distinguishable from the effects of shared meaning and shared form? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1431-1444.

Feldman, L. B. & Pastizzo, M. (2003). Morphological facilitation: The role of semantic transparency and family size. In R.H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 233-258). Berlin: Mouton de Gruyter.

Feldman, L. B., Soltano, E. G., Pastizzo, M., & Francis, S. E. (2004). What do graded effects of semantic transparency reveal about morphological processing? *Brain and Language, 90,* 17-30.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47,* 164-203.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language, 64,* 510-538.

Freyd, P. & Baron, J. (1982). Individual differences in acquisition of derivational morphology. *Journal of Verbal Learning and Verbal Behavior, 21,* 282-295.

Gibbs, R. W. & O'Brien, J. E. (1990). Idioms and mental imagery: The metaphorical motivation for idiomatic meaning. *Cognition, 36,* 35-68.

Gilquin, G. & Gries, S. Th. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory, 5,* 1-25.

Giora, R. (2003). *On our Mind. Salience, context, and figurative language.* Oxford: Oxford University Press.

Giraudo, H. & Grainger, J. (2001) Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review, 8,* 127-131.

Goldberg, A. E. (1995). Constructions: A construction grammar approach to argument structure. Chicago, IL: University of Chicago Press.

Goldberg, A. E. & Jackendoff, R. (2005). The end result(ative). *Language, 81,* 474-477.

Goldberg, A. E. (2006). Constructions at work: The nature of generalization in language. Oxford: Oxford University Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and Speech, 15,* 103-113.

Goldschneider, J. M. & DeKeyser, R. M. (2001). Explaining the "Natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning, 51,* 1-50.

Gómez, R. L. & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4,* 178-184.

Green, L., McCutchen, D., Schwiebert, C., Quinlan, T., Eva-Wood, A., & Juelis, J. (2003). Morphological development in children's writing. *Journal of Educational Psychology, 95,* 752-761.

Gries, S. Th. & Stefanowitsch, A. (2004). Extending collostructional analysis; A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics, 9,* 97-129.

Gries, S. Th., Hampe, B., & Schönefeld, D. (2005). Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics, 16,* 635-676.

Gries, S. Th. (2007). Coll.analysis 3.2. A program for R for Windows 2.x [Computer software].

Griffiths, P. (1986). Constituent structure in text-copying. *York Papers in Linguistics, 12,* 75-116.

Grosjean, F. & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français. *Phonetica, 31,* 144-184.

Guthrie, D., Allison, B., Liu, W., Guthrie, L., & Wilks, Y. (2006). *A closer look at skip-gram modeling. Paper* presented at the Fifth International Conference on Language Resources and Evaluation (LREC-2006) in Genoa, Italy.

Halle, M. & Marantz, A. (1994). Some key features of distributed morphology. In A. Carnie, H. Harley, & T. Bures (Eds.), *MITWPL 21: Papers on phonology and morphology* (pp. 275-288). Cambridge, MA: MIT Working Papers in Linguistics.

Hawkins, P. R. (1971). The syntactic location of hesitation pauses. *Language and Speech, 14,* 277-288.

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics, 39,* 1041-1070.

Hay, J. & Baayen, H. (2002). Parsing and productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 2001* (pp. 203-235). Dordrecht: Kluwer.

Hay, J. B. & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences, 9,* 342-348.

Hilpert, M. (2006). Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory, 2,* 243-257.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning. *Studies in Second Language Acquisition, 27,* 129-140.

Jackendoff, R. (1997). Twistin' the night away. *Language, 73,* 534-559.

Jelinek, F. (1998). *Statistical methods for speech recognition*. Cambridge, MA: MIT Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: John Benjamins.

Kapatsinski, V. M. (2005). To schema or to rule: Contra the dual mechanism model. In R.T. Cover & Y. Kim (Eds.), *Proceedings of the 31st annual meeting of the Berkeley Linguistics Society* (pp. 193-204). Berkeley, CA: Berkeley Linguistics Society.

Kaschak, M. P., Loney, R. A., & Borreggine, K. L. (2006). Recent experience affects the strength of structural priming. *Cognition, 99,* B73-B82.

Kemp, N. (2006). Children's spelling of base, inflected, and derived words: Links with morphological awareness. *Reading & Writing, 19,* 737-765.

Keuleers, E. & Sandra, D. (n.d.) *Similarity and productivity in the English past tense*. Retrieved from http://www.cpl.ua.ac.be/Members/emmanuel/documents/KeuleersSandra-submitted.pdf/file_view.

Koelmans, L. (1981). Uit de geschiedenis van de Nederlandse achtervoegsels III: Het suffix -baar [From the history of Dutch affixes III: the suffix *-baar*]. *De nieuwe taalgids, 74,* 330-339.

Krashen, S. (1982). The input hypothesis: Issues and implications. New York, NY: Longman.

Krott, A., Gagné, C. L., & Nicoladis, E. (2009). How the parts relate to the whole: Frequency effects on children's interpretations of novel compounds. *Journal of Child Language, 36,* 85-112.

Kuiken, F., Vedder, S. C., & Mos, M. (2005). Cognitive task complexity and second language writing performance. In S. Foster-Cohen & P. García-Mayo (Eds.), *EuroSLA Yearbook 2005* (pp. 195-222). Amsterdam: John Benjamins.

Kuperman, V. (2008). *Lexical processing of morphologically complex words: An information-theoretical perspective*. Wageningen: Ponsen & Looijen.

Kurland, B. F. & Snow, C. (1997). Longitudinal measurement of growth in definitional skill. *Journal of Child Language, 24,* 603-625.

Langacker, R. W. (1987). *Foundations of cognitive grammar, Volume 1: Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, R. W. (1991). *Foundations of Cognitive Grammar, Volume 2: Descriptive application*. Stanford, CA: Stanford University Press.

Langacker, R. W. (2008). *Cognitive Grammar; A basic introduction*. Oxford: Oxford University Press.

Langacker, R. W. (2009). Cognitive (Construction) Grammar. *Cognitive Linguistics, 20,* 167-176.

Larsen, J. A. & Nippold, M. A. (2007). Morphological analysis in school-age children: Dynamic assessment of a word learning strategy. *Language, Speech, and Hearing Services in Schools, 38,* 201-212.

Lehtonen, M. & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition, 6,* 213-225.

Lehtonen, M., Niska, H., Wande, E., Niemi, J., & Laine, M. (2006). Recognition of inflected words in a morphologically limited language: Frequency effects in monolinguals and bilinguals. *Journal of Psycholinguistic Research, 35,* 121-146.

Leong, C. K. (1989). The effects of morphological structure on reading proficiency - A developmental study. *Reading & Writing: An interdisciplinary journal, 1,* 357-379.

Levelt, W. J. M. (1969). The scaling of syntactic relatedness: A new method in psycholinguistic research. *Psychonomic Science, 17,* 351-352.

Lewis, D. J. & Windsor, J. (1996). Children's analysis of derivational suffix meanings. *Journal of Speech and Hearing Research, 39,* 209-216.

Libben, G. (2006). Why study compound processing? In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 1-23). Oxford: Oxford University Press.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30,* 333-370.

Lieven, E. (2009). Developing constructions. *Cognitive Linguistics, 20,* 191-199.

Littlemore, J., & MacArthur, F. (2009). *Figurative extensions of word meaning: How do corpus data and intuition match up?* Paper presented at the Fifth Corpus Linguistics Conference in Liverpool, UK.

Mahony, D., Singson, M., & Mann, V. (2000). Reading ability and sensitivity to morphological relations. *Reading & Writing: An interdisciplinary journal, 12,* 191-218.

Martens, V. E. G. & De Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language, 98,* 140-149.

McBride-Chang, C., Wagner, R. K., Muse, A., Chow, B. W. Y., & Shu, H. (2005). The role of morphological awareness in children's vocabulary acquisition in English. *Applied Psycholinguistics, 26,* 415-435.

McDonald, J. L. (2008). Grammaticality judgments in children: The role of age, working memory and phonological ability. *Journal of Child Language, 35,* 247-268.

McDonough, K. (2006). Interaction and syntactic priming: English L2 speakers' production of dative constructions. *Studies in Second Language Acquisition, 28,* 179-207.

McMurray, B. & Hollich, G. J. (2009). Core computational principles of language acquisition: Can statistical learning do the job? Introduction to special section. *Developmental Science, 12,* 365-368.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York, NY: Taylor & Francis.

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

Mintz, T. H. (2005). Finding the verbs: Distributional cues to categories available to young learners. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 31-63). Oxford: Oxford University Press.

Mos, M. (2006). Complexe woorden in het mentale lexicon van kinderen. [Complex words in children's mental lexicon]. *Toegepaste Taalwetenschap in Artikelen, 75,* 53-66.

Mos, M. & Vermeer, A. (2009). Wat weten kinderen nou eigenlijk helemaal van derivationele morfologie? [What DO children know about derivational morphology?]. In A. Backus, M. Keijzer, I. Vedder, & B. Weltens (Eds.), (pp. 253-263). *6e Anéla Conferentie*, Kerkrade, The Netherlands.

Mos, M., Van den Bosch, A., & Berck, P. (in press). The predictive value of word-level perplexity in human sentence processing: A case study on fixed adjective-preposition constructions in Dutch. In: D. Divjak & S. Th. Gries (Eds.) *ICLC/CorpLing 2009 Volumes*. Berlin: Mouton de Gruyter.

Nagy, W., Diakidoy, I., & Anderson, R. (1993). The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives. *Journal of Reading Behavior, 25,* 155-170.

Nagy, W., Berninger, V., Abbott, R., Vaughan, K., & Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. *Journal of Educational Psychology, 95,* 730-742.

Nap-Kolhoff, E. (2010). Second language acquisition in early childhood: a longitudinal multiple case study of Turkish-Dutch children. Utrecht: LOT.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York, NY: Newbury House.

Newman, J. & Rice, S. (2006). Transitivity schemas of English EAT and DRINK in the BNC. In S. Th. Gries & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (pp. 225-260). Berlin: Mouton de Gruyter.

Nicoladis, E. (2005). The acquisition of complex deverbal words by a French-English bilingual child. *Language Learning, 55,* 415-443.

Nunes, T., Bryant, P., & Bindman, M. (1997). Morphological spelling strategies: Developmental stages and processes. *Developmental Psychology, 33,* 637-649.

Nunes, T, & Bryant, P. (2006). Why morphemes are useful in primary school literacy. *Teaching and Learning Research Programme, 14*, Retrieved from http://www.tlrp.org/pub/documents/no14_nunes.pdf.

Östman, J. O. & Fried, M. (2005). Construction grammars: Cognitive grounding and theoretical extensions. Amsterdam: John Benjamins.

Pinker, S. & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S.D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 321-351). Amsterdam: John Benjamins.

Plag, I. (2003). *Word-formation in English*. Cambridge: Cambridge University Press.

Qualls, C. D., Treaster, B., Blood, G. W., & Hammer, C. S. (2003). Lexicalization of idioms in urban fifth graders: A reaction time study. *Journal of Communication Disorders, 36,* 245-261.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review, 11,* 1090-1098.

Rastle, K. (2007). Visual word recognition. In M.G. Gaskell & G. Altmann (Eds.), *The Oxford handbook of psycholinguistics* (pp. 71-88). Oxford: Oxford University Press.

Raveh, M. (2002). The contribution of frequency and semantic similarity to morphological processing. *Brain and Language, 81,* 312-325.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10,* 355-371.

Reinsma, R. & Hus, W. J. B. (2005). *Prisma Woordenboek Voorzetsels [Prisma dictionary of prepositions].* Utrecht: Het Spectrum.

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *Annals of Statistics, 11,* 416-431.

Rubin, H. (1988). Morphological knowledge and early writing ability. *Language and Speech, 31,* 337-355.

Ryan, E. B. (1973). Subjective reactions towards accented speech. In R.W. Shuy & R. W. Fasold (Eds.), *Language attitudes: Current trends and prospects* (pp. 60-73). Washington, DC: Georgetown University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition, 81,* 149-169.

Salamoura, A. & Williams, J. N. (2006). Lexical activation of cross-language syntactic priming. *Bilingualism: Language and Cognition, 9,* 299-307.

Schilperoord, J. & Cozijn, R. (2010). The representation and processing of fixed and compositional expressions. In M. Everaert, T. Lentz, H. de Mulder, Ø. Nilsen, & A. Zondervan (Eds.), *The linguistics enterprise: From knowledge of language to knowledge in linguistics* (pp. 291-314). Amsterdam: John Benjamins.

Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.

Schoonen, R. & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. *Language Testing, 25*, 211-236.

Schönefeld, D. (2006). From conceptualization to linguistic expression: Where languages diversify. In S. Th. Gries & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (pp. 297-344). Berlin: Mouton de Gruyter.

Schreuder, R. & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language, 37*, 118-139.

Schrooten, W. & Vermeer, A. (1994). *Woorden in het basisonderwijs: 15 000 woorden aangeboden aan leerlingen [Words in primary school education: 15.000 words offered to children]*. Tilburg: Tilburg University Press.

Schwanenflugel, P. J. & Gaviskà, D. C. (2005). Psycholinguistic aspects of word meaning. In D. A. Cruse, F. Hundsnurscher, M. Job, & P. R. Lutzeier (Eds.), *Lexikologie/Lexicology: Ein internationales Handbuch zur Natur und Stuktur von Wörtern und Wordschätzen/ An international handbook on the nature and structure of words and vocabularies* (pp. 1735-1748). Berlin: Walter de Gruyter.

Seidenberg, M. S. & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences, 4*, 353-361.

Singson, M., Mahony, D., & Mann, V. (2000). The relation between reading ability and morphological skills: Evidence from derivational suffixes. *Reading & Writing: An interdisciplinary journal, 12*, 219-252.

Smith, M. & Nicoladis, E. (2000). The developing lexicon: Morphosyntactic knowledge in fifth grade children. In A. K. Melby & A. R. Lommel (Eds.), *LACUS Forum XXVI: The lexicon* (pp. 189-200). Fullerton, CA: Linguistic Association of Canada and the United States.

Snow, C. E., Cancino, H., De Temple, J., & Schley, S. (1991). Giving formal definitions: A linguistic or metalinguistic skill? In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 90-112). Cambridge: Cambridge University Press.

Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 375-409). San Diego, CA: Academic Press.

Sorace, A. & Keller, F. (2005). Gradience in linguistic data. *Lingua, 115,* 1497-1524.

Sorace, A., Serratrice, L., Filiaci, F., & Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua, 119,* 460-477.

Sprouse, J. (n.d.). *Evaluating the assumptions of magnitude estimation*. Retrieved from http://www.ling.cogsci.uci.edu/~jsprouse/papers.html.

Stefanowitsch, A. & Gries, S. Th. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics, 8,* 209-243.

Stefanowitsch, A. (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory, 2,* 61-77.

Stefanowitsch, A. & Gries, S. Th. (2008). Channel and constructional meaning: A collostructional case study. In G. Kristiansen & R. Dirven (Eds.), *Cognitive Sociolinguistics* (pp. 129-152). Berlin: Mouton de Gruyter.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153-181.

Strating-Keurentjes, H. (2000). *Lexicale ontwikkeling in het Nederlands van autochtone en allochtone kleuters [Lexical development in migrant and Dutch children's Dutch]*. Veghel: Bek.

Svanlund, J. (2009). *Frequency, familiarity, and conventionalization*. Paper presented at the Fifth Corpus Linguistics Conference in Liverpool, UK.

Taft, M. & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14,* 638-647.

Talmy, L. (1996). The windowing of attention in language. In M. Shibatani & S. A. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (pp. 235-288). Oxford: Oxford University Press.

Taylor, J. R. (2002). *Cognitive grammar*. Oxford: Oxford University Press.

Teddiman, L. (2009). *Conversion and the lexicon: Comparing evidence from corpora and experimentation*. Paper presented at the Fifth Corpus Linguistics Conference in Liverpool, UK.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: Cambridge University Press.

Tomasello, M. & Brooks, P. J. (1999). Early syntactic development: A construction grammar approach. In M. Barrett (Ed.), *The development of language* (pp. 161-190). London: University College London Press.

Tomasello, M. (2003). *Constructing a language. A usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Tomasello, M. & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language, 31,* 101-121.

Tremblay, A. (2009). *Processing Advantages of Lexical Bundles: Evidence from self-paced reading, word and sentence recall, and free recall with event-related brain potential recordings*. Edmonton, Alberta: University of Alberta.

Van den Berg, B. (1974). Het suffix *-baar* [The suffix *-baar*]. In F. F. Odendal, D. C. Hauptfleisch, & J. E. Loubser (Eds.), *Taalkunde - 'n lewe. Studies opgedra aan prof. dr. W. Kempen by geleentheid van sy 65ste verjaardag* (pp. 128-134). Kaapstad: Tafelberg-Uitgewers.

Van den Bosch, A. (2005). Scalable classification-based word prediction and confusible correction. *Traitement Automatique des Langues, 46,* 39-63.

Van den Bosch, A. & Berck, P. (2009). Memory-based machine translation and language modeling. *The Prague Bulletin of Mathematical Linguistics, 91,* 17-26.

Van den Branden, K., Bygate, M., & Norris, J. (2009). *Task-based language teaching: issues, research and practice*. Amsterdam: John Benjamins.

Van der Werf, R., Hootsen, G., & Vermeer, A. (2008). Automated user-centered task selection and input modification in language learning. *ITL Review of Applied Linguistics, 155,* 1-22.

Van Marle, J. (1990). A case of morphological elaboration: The history of Dutch -baar. *Folia Linguistica Historica, 9,* 213-234.

Verhoeven, L. & Vermeer, A. (1993). Taaltoets allochtone kinderen bovenbouw. Diagnostische toets voor de vaardigheid Nederlands bij allochtone en autochtone kinderen in de bovenbouw van het basisonderwijs [Language test migrant children grades 4-6. Diagnostic test for migrant and Dutch children's Dutch proficiency in the later grades of elementary school]. Tilburg: Zwijsen.

Wauters, L. N., Tellings, A. E. J. M., van Bon, W. H. J., & van Haaften, A. W. (2003). Mode of Acquisition of word meanings. *Applied Psycholinguistics, 24,* 385-406.

White, T. G., Power, M. A., & White, S. (1989). Morphological analysis: Implications for teaching and understanding vocabulary growth. *Journal of Memory and Language, 24,* 283-304.

Windsor, J. (1994). Children's comprehension and production of derivational suffixes. *Journal of Speech and Hearing Research, 37*, 408-417.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

Wulff, S. (2009). Converging evidence from corpus and experimental data to capture idiomaticity. *Corpus Linguistics and Linguistic Theory, 5,* 131-159.

Wulff, S., Ellis, N. C., Römer, U., Bardovi-Harlig, K., & Leblanc, C. J. (2009). The Acquisition of Tense-Aspect: Converging Evidence From Corpora and Telicity Ratings. *Modern Language Journal, 93,* 354-369.

Zavrel, J. & Daelemans, W. (1997). Memory-based learning: Using similarity for smoothing. In *Proceedings of the 35th annual meeting of the association for computational linguistics* (pp. 436-443).

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press.

Nederlandse samenvatting

Dit boek gaat over *Complexe Lexicale Items* (CLI's): sequenties die bestaan uit meer dan één betekenisdragend element, maar waarvoor goede redenen bestaan om aan te nemen dat ze als eenheid zijn opgenomen in het taalrepertoire van sprekers. De elementen in een CLI kunnen morfemen zijn, zoals *lees* en *baar* in *leesbaar*, of losse woorden, zoals *trots* en *op* in *trots op*. Voor dergelijke combinaties is het aannemelijk dat ze als eenheid worden opgeslagen omdat ze frequent samen voorkomen en/of omdat ze in combinatie een (deels) eigen betekenis hebben, die niet uit de delen is af te leiden. Vanwege deze tegenstelling tussen complexiteit en eenheid is de centrale onderzoeksvraag wat voor kennis mensen van CLI's hebben, en hoe ze deze kennis gebruiken.

Om een antwoord te kunnen geven op de onderzoeksvraag is een aantal experimenten uitgevoerd, die in de hoofdstukken 2, 3 en 4 beschreven worden. De deelnemers in deze experimenten zijn kinderen in de bovenbouw van het basisonderwijs. Uit eerder onderzoek is gebleken dat rond deze leeftijd de kennis over derivationele morfologie, de processen waarbij éénwoords-CLI's gevormd worden, zich sterk ontwikkelt (o.m. Carlisle, 2000; Nagy, Diakidoy & Anderson, 1993). Hierbij ga ik ervan uit dat taalverwerving gebaseerd is op de input die kinderen krijgen, en dat in het verwervingsproces algemene cognitieve mechanismen aan het werk zijn, zoals categorisatie (in navolging van o.m. Tomasello, 2003; Diessel, 2004).

De theorie waarbinnen de data beschreven en geanalyseerd worden, is die van de Cognitieve Taalkunde (Langacker 1987, 1991, 2008) en, meer specifiek, de Constructiegrammatica (Goldberg 1995, 2006; Boas 2003, 2008). In deze theorie zijn constructies de bouwstenen waaruit uitingen bestaan. Constructies zijn vorm-betekenis paren, waarbij de vorm *lexicaal specifiek* kan zijn, maar ook *ondergespecificeerd*. Een voorbeeld van een constructie met een lexicaal specifiek gedeelte en een ondergespecificeerd gedeelte is de V-BAAR constructie: het affix *–baar* is lexicaal specifiek, en de stam is ondergespecificeerd; deze kan met allerlei werkwoordstammen ingevuld worden (*leesbaar*, *vergelijkbaar*, *eetbaar*…).

Hoofdstuk 1 begint met een beschrijving van onderzoek op het gebied van morfologie en langere lexicale eenheden. Onderzoek naar morfologische processen is bijzonder divers; onderzoekers concentreren zich op verschillende aspecten, afhankelijk van hun theoretische invalshoek. Voor langere lexicale eenheden, zoals *de grote vakantie*, en *een keuze maken* zijn er veel bewijzen dat deze als vaste combinatie opgeslagen zijn in het constructicon, de inventaris aan taalstructuren van een spreker. CLI's kunnen op verschillende niveaus beschreven worden: als lexicaal specifieke combinaties (*leesbaar*), maar ook als instantiaties van een gedeeltelijk ondergespecificeerd patroon: V-BAAR. Hierin is er geen verschil tussen CLI's en constructies op zinsniveau, zoals de CAUSED MOTION constructie. Daarvan is *zij niesde het schuim van haar cappuccino* een lexicaal specifiek voorbeeld, terwijl de ondergespecificeerde constructie weergegeven kan worden als SUBJ. – V – OBJ. –LOCATIE/PAD.

De verwerving en kennis van CLI's en de ondergespecificeerde patronen worden weergegeven in een model van Meervoudige Representaties (Multiple Representations, kortweg MultiRep). Verwerving begint met het leren van een vorm-betekenis paar. In de loop van de tijd komt een kind verschillende instantiaties van hetzelfde patroon tegen. Op basis van overeenkomsten in vorm, betekenis en gebruik kunnen dan een representatie van de interne structuur en van het abstracte patroon ontstaan, naast de opgeslagen lexicaal specifieke representatie.

Hoofdstuk 2 beschrijft een drietal experimenten. Deze verschillen van elkaar in de mate waarin deelnemers gebruik kunnen maken van expliciete kennis (kennis waar ze zich bewust van zijn) en van metalinguïstische kennis (kennis *over* taal), waarbij de hypothese is dat mensen bij voorkeur meer specifieke representaties gebruiken (Dąbrowska, 2008). Deze komen immers het meest overeen met de vorm die ze horen of de betekenis die ze willen weergeven. De 69 deelnemers zijn leerlingen van drie Amsterdamse basisscholen in groep 6, gemiddelde leeftijd 9;4 jaar. Ruim de helft van deze kinderen (N = 38) spreekt naast het Nederlands nog een andere (moeder)taal. Van alle leerlingen werd ook de woordenschatomvang gemeten.

In een woordvormingstaak werd leerlingen gevraagd een afleiding van een gegeven woord in een zin in te vullen, zoals in (1).

(1) **Fantasie**. Ze zegt dat haar vader drie auto's heeft, maar volgens mij zit ze maar een beetje te….

Vijf verschillende affixen (*on–, –heid, –baar, –eren* en *–er*) werden elk in drie toetsitems getest (zie Appendix 1 voor een overzicht). Uit de resultaten blijkt dat er grote verschillen zijn op individueel en op toetsitemniveau. Het percentage correcte antwoorden correleert sterk met de relatieve frequentie van het afgeleide woord (r = .54): woorden die in vergelijking met hun stam vaak voorkomen, worden vaak correct ingevuld. Op individueel niveau is een sterke samenhang met de woordenschatomvang zichtbaar (r = .70). Bovendien scoren kinderen vaak verschillend op items met hetzelfde affix. Dit wijst erop dat verwerving van de affixconstructies geleidelijk verloopt. De frequentiecorrelaties maken het aannemelijk dat kinderen waar mogelijk gebruik maken van lexicaal specifieke representaties.

In een definitietaak moest een deel van dezelfde leerlingen (N = 47, 29 eentalige, 18 tweetalige kinderen) een definitie geven van morfologisch complexe woorden die ze voorgelegd kregen (*wraakbaar, klantentrekker*, zie Appendix 2). Dezelfde vijf affixen kwamen weer elk drie keer voor in deze taak, die expliciete en metalinguïstische kennis meet. Ook hier correleren de individuele scores sterk met de woordenschatomvang (r = .49). Er zijn duidelijke verschillen tussen de affixen in het percentage antwoorden waarin de betekenis van het affix expliciet genoemd is. Dit is waarschijnlijk een taakeffect: voor een affix als *on–* is het eenvoudiger de betekenis te verwoorden dan voor *–heid*. Omdat de toetsitems hier voor de kinderen onbekende woorden waren, konden zij de betekenis alleen achterhalen door gebruik te maken van representaties van abstracte patronen.

Met een lexicale decisietaak is gemeten of kinderen een effect van morfologische familiegrootte lieten zien in hun reacties. In een lexicale decisietaak moet een deelnemer zo snel mogelijk aangeven of een bepaalde letterreeks een bestaand woord vormt (*stoep*) of niet (*smoer*). Uit onderzoek (De Jong et al., 2000) is gebleken dat een woord dat veel voorkomt in afleidingen en samenstellingen, sneller herkend wordt dan een op zichzelf even frequent woord waar dit minder voor geldt (zie Appendix 3 voor een overzicht van de stimuli). De 69 leerlingen uit het eerste experiment waren ook hier de deelnemers. In een variantieanalyse zijn de verschillen in reactietijd tussen woorden met een grote familie (gemiddeld 1171 ms., sd. 276) en die met een kleine familie (gem. 1138 ms., sd. 262) significant ($F_{(1, 1968)} = 7.05$, $p < .01$). Dat het bestaan van morfologische familieleden helpt bij het herkennen van een woord, kan alleen worden verklaard in een model waarin morfemen binnen een complex woord gerepresenteerd zijn. Alleen dan kan immers *werk* in *werkster* en *werkbriefje* faciliterend zijn voor het herkennen van *werk* als Nederlands woord. Het effect van familiegrootte bewijst dat we representaties hebben die weliswaar lexicaal specifiek zijn, maar waarin de morfologische structuur ook aanwezig is. Deze online taak meet impliciete kennis: deelnemers hebben geen tijd om bewust te reflecteren op hun reacties.

De uitkomsten van deze drie experimenten laten zien dat het niveau van een gebruikte representatie afhankelijk is van de eisen die de taak stelt. Waar mogelijk (woordvormingstaak) gebruiken de kinderen lexicaal specifieke representaties, maar als dit niet kan (definitietaak) maken ze gebruik van ondergespecificeerde representaties. De scores van kinderen verschillen sterk van elkaar bij alle drie de experimenten. Deze verschillen hangen samen met de woordenschatomvang, wat wijst op een ontwikkeling die gaande is: voor de meeste kinderen is de abstracte kennis nog niet zó ver ontwikkeld, dat ze hiervan bij alle toetsitems gebruik kunnen maken. De frequentie-effecten van de toetsitems passen bij dit beeld van graduele verwerving, dat volledig overeenstemt met het pad van ontwikkeling zoals dit voorgesteld wordt in het MultiRep-model.

Hoofdstuk 3 is een *case study* van twee constructies die in betekenis veel op elkaar lijken, maar niet in vorm: de V-BAAR constructie (2) en de IS TE V constructie (3).

(2) ik ben eigenlijk alleen overdag bereikbaar (fn006994.85)[1]

(3) de expositie **is** nog **te bezichtigen** tot zeven mei (fv600185.6)

Beide constructies geven weer dat iets mogelijk is. In het geval van de V-BAAR constructie gebeurt dit met een werkwoordstam en een derivationeel morfeem, en in de IS TE V constructie met een koppelwerkwoord, *te* en een infinitief. De vraag is nu of deze constructies productief zijn. De definitie van productiviteit is daarbij: een constructie is productief wanneer sprekers (deels) ondergespecificeerde representaties van deze constructies hebben, waarmee nieuwe instantiaties te maken zijn. Productiviteit is waarschijnlijker voor constructies die frequent zijn, *salient* (opvallend) zijn en decompositioneel zijn (d.i. waarbij de delen duidelijk herkenbaar zijn). Deze eigenschappen zijn alle drie gradueel en relatief, en geven daarmee onvoldoende houvast om voor een constructie eenvoudig de productiviteit te bepalen.

[1] Deze voorbeelden komen uit het Corpus Gesproken Nederlands. De code verwijst naar de vindplaats in dat corpus.

Om de betekenis van de constructies specifieker vast te stellen is een analyse van instantiaties in het Corpus Gesproken Nederlands (CGN) gedaan. Hieruit blijkt dat beide constructies ongeveer even vaak voorkomen (1282 tokens, 206 types voor V-BAAR en 1108 tokens, 211 types voor de IS TE V constructie). Met behulp van een collostructionele analyse (Stefanowitsch & Gries, 2003) kan de betekenis van de constructies meer specifiek omschreven worden. De distinctieve collexeemanalyse (Gries & Stefanowitsch, 2004) laat zien dat het verschil tussen beide constructies met name zit in het onderscheid tussen objectieve mogelijkheid (V-BAAR) en subjectieve mogelijkheid (IS TE V). In de V-BAAR constructie heeft “het meest Patiens-achtige argument van het werkwoord de eigenschap de gebeurtenis die het werkwoord beschrijft te kunnen ondergaan”. De betekenis van de IS TE V constructie is samen te vatten als “in de ogen van de spreker heeft het meest Patiens-achtige argument van het werkwoord de eigenschap de gebeurtenis die het werkwoord beschrijft te kunnen ondergaan”. Het verschil tussen de constructies komt bijvoorbeeld naar voren in de sterke vertegenwoordiging van werkwoorden die een mentale activiteit uitdrukken (*hopen, geloven, verwachten* etc.) in de IS TE V constructie. Dit zijn werkwoorden die door hun betekenis altijd een zekere mate van subjectiviteit uitdrukken. De corpusanalyse geeft aanleiding voor de aanname dat veel lexicaal specifieke instantiaties als zodanig opgeslagen zijn. Ze hebben een eigen, specifieke betekenis, een eigen vorm of een eigen distributie (collocaties met bijwoorden van gradatie zoals *best wel, niet, prima*). Deze analyses geven echter nog geen antwoord op de vraag of de constructies productief zijn.

De productiviteit van de constructies is gemeten met een acceptabiliteitstaak. In deze Magnitude Estimation-taak (o.m. Sorace, 1996) geven deelnemers met getallen aan in hoeverre een zin acceptabel is. Door nieuwvormingen met beide constructies aan te bieden, en daarin te variëren in het type werkwoord, kan worden nagegaan in hoeverre het type werkwoord de acceptabiliteit beïnvloedt. Vijf typen werkwoorden kwamen voor, die van elkaar verschilden in hun argumentstructuur (zie Appendix 4 voor een overzicht van de testzinnen). Bovendien waren zinnen zodanig geformuleerd, dat ze een objectieve of subjectieve mogelijkheid uitdrukten. Deelnemers waren 148 leerlingen van drie Tilburgse basisscholen (gemiddelde leeftijd 12;3 jaar, 110 eentalige en 38 tweetalige leerlingen). Daarnaast deden ook 72 volwassen moedertaalsprekers (gemiddelde leeftijd 43 jaar) mee aan het experiment, om de vraag te kunnen beantwoorden of zij een ondergespecificeerde representatie van de constructies hebben.

De verschillen in acceptabiliteit tussen de vijf typen werkwoorden zijn significant voor zowel de kinderen ($F_{(3.58,\ 526.71)} = 59.42$, $p < .001$) als de volwassenen ($F_{(2.63,\ 186.59)} = 189.07$, $p < .001$)[2], waarbij alle paarsgewijze vergelijkingen significant zijn. Het voornaamste verschil hierbij is tussen werkwoorden met een Agens-Patiens structuur enerzijds en een Stimulus-Ervaarder structuur anderzijds. Deze laatste worden beduidend lager gewaardeerd. Het verschil tussen een objectieve en subjectieve mogelijkheid voor V-BAAR en IS TE V is voor kinderen niet significant, maar voor volwassenen wel ($F_{(1,71)} = 7.81$, $p < .01$ voor het type mogelijkheid, en $F_{(1,\ 71)} = 51.83$, $p < .001$ voor de interactie met constructie). Daarbij haalde de V-BAAR constructie in combinatie met de weergave van een subjectieve mogelijkheid de laagste score.

Deze significante verschillen geven aan dat sprekers een ondergespecificeerde representatie van de constructies tot hun beschikking hebben. Alleen hiermee kunnen ze immers een consistent onderscheid in acceptabiliteit maken. De conclusie is dan ook, dat beide constructies productief zijn en dat er een subtiel onderscheid is in betekenis.

Ook **Hoofdstuk 4** behandelt enkele specifieke CLI's die twee lexicaal specifieke elementen bevatten. Het gaat hierbij om combinaties van het type *trots op*, hier FAP's genoemd: Fixed Adjective-Preposition-pairs (vaste Adjectief-Prepositie-paren). Deze combinaties zijn conventioneel; het ligt niet in de betekenis van de elementen besloten dat *trots* nu juist met *op* gecombineerd wordt, en niet met *bij* of *van*. Om die reden zullen FAP's in het constructicon opgeslagen moeten zijn. De onderzoeksvraag is nu in hoeverre deze combinaties als eenheid worden verwerkt.

In tegenstelling tot de voorgaande hoofdstukken ligt de focus juist op het actieve gebruik van de lexicaal specifieke representatie. Daarbij wordt ook een extra factor in de analyses betrokken: de waarschijnlijkheid van het voorkomen van een woord (in de tekst *perplexiteit* genoemd, wat naar het omgekeerde verwijst: de onverwachtheid van een woord). Deze *perplexiteit* (Van den Bosch, 2005) wordt berekend op basis van woorddistributies in een corpus, en geeft de (on)waarschijnlijkheid van voorkomen weer, gegeven de drie voorgaande woorden. Voor sommige sequenties bestaat er één duidelijke kandidaat, zoals *nog lang en…*(*gelukkig*), maar vaker zijn er enkele

[2] Gecorrigeerde F-waardes worden gerapporteerd (Huynh-Feldt correctie), omdat Mauchly's test significant is.

waarschijnlijke kandidaten, of een grotere set mogelijke woorden, zoals *heel trots op…* (*hem, haar, de ….*).

FAP's bestaan uit twee lexicaal specifieke elementen. Zin (4a) geeft een lexicaal specifiek voorbeeld, waarbij de ondergespecificeerde structuur in (4b) weergegeven staat.

(4) a. de boer is **trots op** zijn auto (fn001204.26)
b. NP V_{fin} [Adj [Prep $[N]_{NP}]_{PP}]_{AP}$

In de instantiaties van de constructie wordt steeds met een koppelwerkwoord aan een nominale constituent (het subject) een eigenschap toegewezen (het adjectief) die veroorzaakt wordt door of gericht is op de nominale constituent in de voorzetselgroep. Deze analyse is niet alleen van toepassing op *trots op*, maar op de meeste FAP's (vergelijk *blij met, tevreden over, boos op* etc.). Veel FAP's komen vooral met één werkwoord voor (*zijn*, in het geval van *trots op*).

De combinatie van de twee woorden die de lexicaal specifieke elementen van FAP's vormen, kan ook voorkomen zonder dat er sprake is van een FAP, zoals in (5), waar de voorzetselgroep een locatie aanduidt. In dergelijke contexten heeft de Adjectief-Prepositie combinatie dus geen FAP-interpretatie.

(5) Daan stond **trots op** de carnavalswagen

Dit maakt het mogelijk om te toetsen of sprekers alleen de vormsequentie Adjectief-Prepositie opgeslagen hebben, dan wel of de specifieke betekenis van de sequentie deel uitmaakt van de representatie. In het eerste geval is er geen verschil tussen de verwerking van *trots op* in (4) en (5), en in het tweede geval wel.

De vorm- of vorm+betekenisopslag en de invloed van het werkwoord zijn gemeten in een zogeheten kopieertaak. Deelnemers krijgen een zin te zien op een computerscherm. Ze klikken op een knop, waarna de zin vervangen wordt door een scherm met de losse woorden en een lege balk. Hierin moeten ze de zin reconstrueren in de juiste volgorde. Wanneer ze niet meer weten hoe de zin verder gaat, kunnen ze terugkijken met een klik op een knop. Deze momenten van terugklikken, *switches*, worden opgeslagen. De distributies van switches zijn informatief: ze zullen plaatshebben op grenzen van tijdelijk opgeslagen lexicale sequenties. De deelnemers vormden een subset van de deelnemers aan de acceptabiliteitstaak (N = 35, gem. leeftijd 12;5). Zij kopieerden zinnen waarin steeds één van zes FAP's voorkwam (*trots op, voorzichtig met, jaloers op geïnteresseerd in, enthousiast over* en *boos op*), elk in vier contexten: met en zonder FAP-interpretatie, en met werkwoord dat er veel of weinig mee voorkomt (zie Appendix 5 voor een overzicht van de testzinnen).

De switches van alle kinderen voor elke woordgrens werden opgeteld, waarna het aantal switches op dezelfde positie (bijvoorbeeld tussen het adjectief en de prepositie) in de vier verschillende contexten werd vergeleken. Dit levert significante verschillen op voor het effect van het werkwoord tussen het subject en het adjectief ($t_{(22)} = 3.21$, $p < .01$), met minder switches (10.33) bij zinnen met een werkwoord dat veel voorkomt met deze constructie, dan bij zinnen met een minder frequent werkwoord (14.00). Het hoofdeffect van de FAP-interpretatie is niet significant, maar er zijn wel complexe interacties. De woordperplexiteitsdata laten alleen een significant effect van het werkwoord zien ($t_{(22)} = 3.21$, $p <.01$). De correlaties tussen beide maten zijn sterk: $r > .400$ voor 23 van de 24 zinnen. Verschillen tussen beide maten zijn vooral zichtbaar bij de sequentie subject-werkwoord, waartussen weinig switches gemaakt worden, maar waar de perplexiteit relatief hoog is, en bij de sequenties prepositie-determinant-naamwoord, waar de perplexiteit vaak hoger is tussen het tweede en het derde woord, terwijl er meer switches zijn tussen het eerste en het tweede woord. Dit lijkt er op te wijzen dat de voorspelbaarheid van woorden weliswaar een belangrijke determinant is voor het tijdelijk opslaan van sequenties (sterke correlaties), maar dat betekeniseenheden toch ook een rol spelen. Voor het MultiRep-model betekenen deze resultaten dat context en interpretatie van een sequentie de waarschijnlijkheid van het gebruik van een specifieke representatie beïnvloeden. Omdat het model uitgaat van vorm-betekenis paren, is de invloed van interpretatie te voorspellen. De rol die de context speelt, krijgt vorm door uit te gaan van activering door collocaties.

In **Hoofdstuk 5** worden tenslotte de resultaten van de experimenten teruggekoppeld naar het MultiRep-model. Hier zet ik uiteen dat het model representaties van constructies beschrijft in termen van lexicale specificiteit (een element is specifiek of ondergerepresenteerd) en van waarschijnlijkheid (een element komt in elke instantiatie van een constructie voor of is een waarschijnlijk element, een collocatie). Kinderen beginnen met specifieke representaties, en ontwikkelen op basis van overeenkomsten in vorm, betekenis en distributie ondergespecificeerde representaties. Deze specifieke en ondergespecificeerde representaties bestaan naast elkaar. Onderzoekers moeten daarom in principe uitgaan van het gebruik van specifieke representaties, totdat bewezen is dat sprekers een ondergespecificeerde representatie hebben van een constructie. Mogelijk herkennen taalkundigen abstracte patronen die niet als zodanig voorkomen in het constructicon van sprekers.

Voor taalleerders is het herkennen van patronen de weg waarlangs generalisaties kunnen ontstaan, maar hoe dit proces precies te werk gaat, is nog niet duidelijk. De distributie van types en tokens is daarbij een eerste punt van aandacht. Goldberg (2006) zegt dat constructies vaak één veelvoorkomende instantiatie hebben die een prototypische betekenis heeft en daarom een aanknopingspunt vormt voor de verwerving van de constructie. Het is echter maar de vraag of dit voor CLI's en hun onderliggende patronen ook geldt. Daarnaast is het interessant te kijken in hoeverre het expliciet aandacht besteden aan taalstructuren op school van invloed is op de kennis van kinderen. Ook de rol van input op het constructicon dat kinderen ontwikkelen en de effecten van frequentie verdienen nadere aandacht.

In *Complex Lexical Items* breng ik verschillende subdisciplines van de taalwetenschap samen. Door gebruik te maken van de theoretische inzichten van de Constructiegrammatica, door taalverwerving vanuit een cognitief-taalkundige invalshoek te benaderen, en door te kiezen voor een combinatie van experimenten en corpusonderzoek, is een duidelijker beeld ontstaan van de kennis die sprekers hebben van CLI's en hoe ze deze kennis gebruiken. Sprekers hebben zowel lexicaal specifieke representaties van CLI's als ondergespecificeerde representaties van de onderliggende patronen in hun constructicon. Welke representaties geactiveerd worden, is afhankelijk van de context en de taaltaak die ze uitvoeren. CLI's zijn dus niet eenvoudig lexicaal, en meer dan alleen syntactische patronen; het zijn *Complexe Lexicale Items*.

Appendix 1: Test items word formation task (Chapter 2)

Below, the task is replicated in its original format (not including the correct answers in the last column), followed by an English translation.

In de zinnen hieronder is steeds een woord weggehaald.
Vul het woord in dat het beste past.
Let op: het gaat steeds om maar één woord.

Bijvoorbeeld:	Lopen	Anne heeft tien kilometer …	gelopen
	Tekenen	Zij maakt een hele mooie …	tekening

1.Fantasie	Ze zegt dat haar vader drie auto's heeft, maar volgens mij zit ze maar een beetje te …	Fantaseren
2.Advies	Ik weet niet wat ik moet doen. Wat zou je me …?	Adviseren
3.Vangen	Sommige indianen hangen een netje boven hun hoofd om dromen in te vangen. Dat netje heet een dromen….	Vanger
4.Verkouden	Mo is anders nooit ziek, maar nu heeft hij een ….	Verkoudheid
5.Aardig	Je kan het ook wel vriendelijk vragen! Doe niet zo ….	Onaardig
6.Bijten	Ze zeggen dat blaffende honden niet bijten. Maar als een hond niet blaft, is het dan altijd een…?	Bijter
7.Duidelijk	Ik begrijp niets van de uitleg! Misschien geeft een tekening meer … over wat we moeten doen.	Duidelijkheid
8.Lezen	Wat kan Max mooi schrijven! Zijn handschrift is heel ……	Leesbaar
9.Geduld	Ad wil meteen beginnen. Hij wordt van al dat wachten super …..	Ongeduldig
10.Hard	Door een luchtbed minder vol te blazen, zorg je ervoor dat de … minder wordt.	Hardheid
11.Herhalen	Dat is een heel moeilijk woord om uit te spreken. Eén keer lukt wel, maar het is bijna niet …	Herhaalbaar
12.Horen	Met de ramen dicht merk je niets van het lawaai. De geluiden zijn bijna …	Onhoorbaar
13.Inbreken	De televisie van de buren is gestolen door een ….	Inbreker
14.Kopie	Wil je deze brief voor mij een keer ….?	Kopiëren

Complex Lexical Items

In each sentence below, a word has been deleted.
Fill in the word that fits best.
Pay attention: you should always add just one word

For example:	Walk	Anne has ... ten kilometers	walked
	Draw	She makes a very pretty …	drawing

1.Fantasy (N)	She says that her father owns three cars, but I think she's just ...	Fantasizing
2.Advise (N)	I don't know what to do. What would you ... me?	Advise (V)
3.Catch (V)	Some Indians hang a net over their head to catch dreams in. That net is called a dream...	Catcher
4.Bad cold (Adj)	Mo is never sick, but now he has a ...	Bad cold (N)
5.Kind (Adj)	You could have asked it a little nicer! Don't be so ...	Unkind
6.Bite (V)	Ze zeggen They say that a barking dog won't bite. But if a dog doesn't bark, is it always a ...	Biter
7. Clear (Adj)	I don't understand the explanation at all! Perhaps a drawing gives more ... about what we have to do	Clarity
8.Read (V)	Max writes beautifully! His hand writing is very …	Legible
9.Patience (N)	Ad wants to start right away All the waiting makes him super...	Impatient
10.Hard (Adj)	By putting less air in an air bed, you make sure that the ... is less.	Hardness
11.Repeat (V)	That is a very difficult word to pronounce. To do it once is ok, but it's almost not...	Repeatable
12.Hear (V)	With the windows closed you hardly notice the noise. The sounds are almost...	Inaudible
13.Break in (V)	The neighbors' TV was stolen by a ...	Burglar
14 Copy (N)	Would you ... this letter for me?	Copy (V)

Appendix 2: Test items word definition task (Chapter 2)

Test item	Morpheme-by-morpheme translation	Word translation
Maxim**eren**	Maxim-ize	Maximize
Meubil**eren**	Piece_of_furniture-ize	Furnish
Adress**eerbaar**	Address-ize-able	Addressable
Plooi**baar**	Pleat-able	Pliable
Wraak**baar**	Revenge-able	Objectionable
Overdraag**baarheid**	Over-carry-able-ness	Infectiousness
Breedsprakig**heid**	Broad-speech-y-ness	Verboseness
Smal**heid**	Narrow-ness	Narrowness
Onaanzienlijk**heid**	Un-on-see-ly-ness	Insignificance
Onaangedaan	Un-on-put	Unmoved
Onvrijwillig**er**	Un-free-will-y-er	More involuntarily
Onruststok**er**	Un-rest-stoke-er	Troublemaker
Herintred**er**	Re-entry-er	Returnee (to the workforce, after period of absence)
Klantentrekk**er**	Customer-pull-er	Sale item that attracts costumers

NB: order of presentation was randomized for each participant.

Appendix 3: Test items lexical decision task (Chapter 2)

Item	Family size	Mean RT (1) (ms.)	Mean RT (2) (ms.)	FS (1)	FS (2)	f per million (CELEX)	<70% correct
Kerel	S	1197	465	1	5	62	
Sofa	S	*	506	1	1	9	X
Maïzena	S	*	675	1	0	6	X
Villa	S	1235	488	1	3	25	
Kolonel	S	*	534	1	5	66	X
Neef	S	1067	522	1	3	27	
Reeks	S	*	497	1	7	59	X
Hiel	S	1270	n.a.	1	3	13	
Veranda	S	*	578	2	1	7	X
Gazon	S	*	535	2	1	9	X
Term	S	*	491	2	1	88	X
Humor	S	1227	500	2	7	17	
Prooi	S	1193	500	2	1	16	
Atlas	S	1231	532	2	2	3	
Teen	S	1123	n.a.	2	2	34	
Dal	S	1243	506	2	6	36	
Stoep	S	1116	n.a.	3	3	17	
Ellende	S	1435	476	3	4	36	
Broer	S	1119	456	4	6	128	
Tante	S	1066	493	4	5	104	
Mode	L	1121	483	5	27	17	
Berk	L	*	543	6	10	4	X
Spion	L	1210	489	7	20	7	
Alarm	L	1184	456	9	19	6	
Rapport	L	1157	n.a.	9	24	67	
Dief	L	1033	n.a.	10	20	13	
Machine	L	1232	n.a.	39	85	61	
Ketel	L	1241	554	10	19	10	
Schema	L	1288	467	11	31	25	
Muts	L	1128	498	11	19	7	
Vee	L	1312	512	11	35	17	
Gast	L	1191	n.a.	11	38	56	
Plein	L	1039	502	12	19	33	
Lente	L	1135	n.a.	12	20	19	
Stroom	L	1179	n.a.	17	67	47	
Spiegel	L	1116	n.a.	18	47	51	
Park	L	1039	466	19	40	38	
Broek	L	1134	487	24	37	61	
Band	L	1141	557	32	79	75	
Koning	L	1136	486	36	54	100	

FS: test items either had a small (S) or a large (L) morphological family
RT: reaction times in the present experiment (1) and in De Jong et al. (2000) (2)
FS: morphological family size in the *Woordwerken* corpus (1) and in CELEX (2)
Test items with less than 70% correct responses were removed from the analyses

Appendix 4: Test items Magnitude estimation task (Chapter 3)

The test items below are listed by verb category. In the experimental setting, the presentation order was randomized for each participant. The children who participated rated a subset of these items. Test items that were only included in the adults' version are indicated with *.

I. Prototypical transitive verbs, with Agent-Subject and Patient-Object

A. V–BAAR, factual
B. IS TE V, subjective
C. –BAAR, subjective
D. IS TE V, factual

1. SCHRIJVEN
   A. 0.25 is ook schrijfbaar als ¼.
   B. Deze brief is makkelijk te schrijven in een half uur, lijkt me.
   C. Zo'n werkstuk is, denk ik, prima schrijfbaar voor brugklassers.
   D. Je klacht is te schrijven aan de afdeling klanteninformatie.

2. SCHILDEREN
   A. Deze gladde muur is eenvoudig schilderbaar met een roller.
   B. Je nieuwe kamer is vast wel te schilderen in een kleur die jij leuk vindt.
   C. Een portret zal wel schilderbaar zijn met een achtergrond naar keuze.
   D. De buitenmuren zijn alleen te schilderen met een speciaal soort verf.

3. COMBINEREN
   A. Deze aanbieding is niet combineerbaar met andere kortingen.
   B. Pannenkoeken met stroop zijn goed te combineren met een glas koude melk, vind ik.
   C. Pannenkoeken en mayonaise zijn niet combineerbaar, lijkt me.
   D. Deze korting is niet te combineren met andere aanbiedingen.

4. MAAIEN*
   A. Dat lange gras is alleen maaibaar met een elektrische grasmaaier.
   B. Die jungle die de buren hun achtertuin noemen is niet te maaien.
   C. Het lijkt me dat het gras van de buren maaibaar is in een middag.
   D. In het voorjaar is het gras op sommige plekken niet te maaien vanwege vogelnesten.

II. Optionally transitive verbs with Agent-Subject and Patient-Object or Undergoer-Subject

A. V–BAAR
B. IS TE V

5. SMELTEN
   A. IJzer is alleen smeltbaar op een hele hoge temperatuur.
   B. Chocolade is makkelijk te smelten in de magnetron.

6. SCHEUREN
   A. Dit extra sterke plakband is toch met de hand scheurbaar.
   B. Als je het eerst omvouwt, is inpakpapier best te scheuren.

7. DROGEN*
   A. Een wollen trui is niet droogbaar in de machine.
   B. Pas geschilderde muren zijn sneller te drogen door de verwarming hoger te zetten.

8. BREKEN
   A. Zelfs de hardste schelp is breekbaar.
   B. Een chocoladeletter is moeilijk in gelijke stukjes te breken.

III Optionally transitive verbs with a (generally) implicit second argument

A. V–BAAR
B. IS TE V

9. ROKEN*
   A. Sigaren moeten eerst rijpen voordat ze rookbaar zijn.
   B. Zware tabak is niet te roken als je er niet aan gewend bent.

10. SCHREEUWEN
   A. Een goede yell is makkelijk schreeuwbaar.
   B. Een korte boodschap is nog wel naar de overkant van de rivier te schreeuwen.

11. WANDELEN
   A. De afstand van huis naar de middelbare school is voor de meeste leerlingen niet meer wandelbaar.
   B. Elke etappe van de avondvierdaagse is in twee uur te wandelen.

12. ZINGEN
    A. De tekst van dit liedje is zo lastig dat het bijna niet zingbaar is.
    B. Als een liedje in een andere taal is, is de tekst vaak bijna niet te zingen.

IV Non-prototypical transitive verbs, with Stimulus-Experiencer roles and BE-perfect

A. V–BAAR
B. IS TE V

13. FASCINEREN
    A. Een spannend boek is voor mij helemaal fascineerbaar.
    B. Een enge film is enorm te fascineren voor mij.

14. VERBAZEN
    A. Eerst was iedereen nog enthousiast, dus deze onverwachte kritiek is verbaasbaar.
    B. Hoewel hij altijd hard traint, is deze plotselinge vooruitgang wel te verbazen.

15. AFSCHRIKKEN
    A. Niemand weet hoe een vogel afschrikbaar is van een kersenboom.
    B. Ongewenst bezoek is af te schrikken met een alarmsignaal.

16. VERRASSEN*
    A. De tegenstander was simpel verrasbaar met een onverwachte beweging.
    B. Mijn oma is altijd te verrassen met een mooie bos bloemen.

V. Non-prototypical transitive verbs, with Stimulus-Experiencer roles and HAVE-perfect

A. V–BAAR
B. IS TE V

17. LUKKEN
    A. Het is een ambitieus plan, maar als iedereen helpt is het zeker lukbaar.
    B. Dankzij haar slimme idee is het plan zeker te lukken.

18. MEEVALLEN
    A. Als je wacht tot de uitverkoop, is de prijs zelfs in dure winkels soms meevalbaar.
    B. Met vijftig procent korting is de prijs van een winterjas best mee te vallen.

19. ONTGAAN
    A. Als de juf er is, weet je dat meteen: haar stem is onontgaanbaar.
    B. Omdat mijn tante altijd zo hard praat is haar aanwezigheid niet te ontgaan.

20. ONTSCHIETEN*
    A. Al gebruik je een woord nog zo vaak, het blijft toch plotseling ontschietbaar.
    B. Met zoveel punten voorsprong is de Trappers het kampioenschap niet meer te ontschieten.

Filler items:

1. Vier pagina's kunnen makkelijk geschreven worden in een uur.
2. Een brief is mogelijk te schrijven naar de directeur van de school.
3. Een muur schilderen kun je zeker in je eentje.
4. Voor het schilderen van je nieuwe kamer is het mogelijk alles blauw te erven.
5. Deze broek is mogelijk gecombineerd te worden met zwarte of bruine schoenen.
6. Je kunt het bezoek aan je oma meteen combineren met een vakantie.
7. De grasmaaier kan gemaaid worden over het gras van de buren.*
8. Met een elektrische grasmaaier is het mogelijk heel snel het gras te maaien.*
9. Dat plan kan alleen gelukt worden als iedereen meewerkt.
10. Het kan hem vast wel lukken om op tijd thuis te zijn.
11. De teleurstelling kan meegevallen worden, omdat ze toch niet hadden verwacht te winnen.
12. Omdat ze zich op het ergste had voorbereid, kon de vertraging alleen maar meevallen.
13. Zijn nieuwe jas is zo opvallend, dat hij niet kan worden ontgaan.
14. Het grote reclamebord tegenover de school kan me niet ontgaan.
15. De lange naam van die nieuwe jongen is erg ontschietelijk.*
16. Het is goed mogelijk dat de naam van die straat je ontschiet.*
17. Ik kan helemaal gefascineerd worden door een eng verhaal.
18. Het is mogelijk hem te fascineren met een vies verhaal over wormen.

19. Jouw onverwachte reacties konden hem echt verbazen.
20. De beroemde voetballer kon toch verbaasd worden door de televisiecamera’s.
21. Zelfs met geluiden van een blaffende hond kon de inbreker niet worden afgeschrikt.
22. Ook de prik bij de dokter was niet mogelijk haar af te schrikken van die verre reis.
23. Hem verrassen met kaartjes voor de bioscoop is zeker mogelijk.*
24. Wij konden niet meer verrast worden omdat iemand anders het geheim al had verklapt. *
25. Een waterpijp kan alleen gerookt worden door mensen met ervaring.*
26. Een zwangere vrouw is niet mogelijk veel sigaretten te roken.*
27. Het is niet mogelijk urenlang te schreeuwen zonder schor te worden.
28. Steeds dezelfde protesten kunnen niet eindeloos geschreeuwd worden.
29. Het is mogelijk van Tilburg naar Breda te wandelen.
30. Dat korte stukje van school naar huis kan prima gewandeld worden.
31. Dat makkelijke refrein kan ook door jou gezongen worden.
32. Jou is mogelijk dat leuke liedje te zingen.
33. IJs kan gesmolten worden door het in de zon te leggen.
34. Het is mogelijk boter te smelten op een laag vuur.
35. Je kunt een heel telefoonboek nooit in één keer scheuren.
36. Dik karton is bijna onmogelijk gescheurd te worden.

Appendix 5: Test sentences copy task (Chapter 4)

Sentences are grouped by adjective-preposition pair. For each pair, there are four test sentences:

A: frequent verb, FAP interpretation
B: frequent verb, AP is a coincidental sequence
C: infrequent verb, AP is a coincidental sequence
D: infrequent verb, FAP interpretation

***Boos op*, 'angry at'**

A. Vanwege die beslissing was Fatima boos op haar school en de nieuwe leraar
   'Because of that decision, Fatima was angry at her school and the new teacher'
B. Ondanks de toestemming was Fatima boos op haar school brood aan het eten
   'In spite of the permission, Fatima was angrily eating a sandwich at school'
C. Om te protesteren stond Fatima boos op haar school folders uit te delen
   'In order to protest, Fatima angrily handed out flyers at her school'
D. Door de veranderingen stond Fatima boos op haar school in een volle trein
   'Because of the changes, Fatima, angry at her school, was standing in a full train'

***Enthousiast over*, 'enthusiastic about'**

A. Al in april was Esra enthousiast over de vakantie naar haar familie in het buitenland
   'Already in April, Esra was enthusiastic about the vacation to her family abroad'
B. Voor de pauze was Esra enthousiast over de vakantie aan het kletsen met haar vriendin
   'Before the break, Esra was chatting enthusiastically about the vacation to her friend'
C. Na het weekend begon Esra enthousiast over de vakantie te vertellen aan haar hele klas
   'After the weekend, Esra enthusiastically started to tell her whole class about the vacation'
D. Lang voor vertrek begon Esra enthousiast over de vakantie alvast haar tas in te pakken

'Long before departure, Esra, enthusiastic about the vacation, already started packing her bag'

***Geïnteresseerd in*, 'interested in'**

A. Vanwege zijn loopneus was Tim geïnteresseerd in een folder van de huisarts over griep en snot
   'Because of his runny nose, Tim was interested in a brochure from the doctor about the flue and mucus'
B. Met zware hoofdpijn was Tim geïnteresseerd in een folder van de huisarts tips aan het lezen
   'With a bad headache, Tim was reading a brochure from the doctor with interest'
C. Met hoge koorts stond Tim geïnteresseerd in een folder van de huisarts te lezen over ziektes
   'With a high fever, Tim stood reading with interest a brochure from the doctor about diseases'
D. Met erge buikpijn stond Tim geïnteresseerd in een folder van de huisarts te wachten op hulp
   'With severe stomachache, Tim, interested in a brochure from the doctor, stood waiting for help'

***Jaloers op*, 'jealous of'**

A. Zonder enige twijfel was Sophie jaloers op de tas van haar nichtjes uit Breda
   'Without any doubt, Sophie was jealous of her nieces' bag from Breda'
B. Zonder te aarzelen was Sophie jaloers op de tas van haar buurvrouw gaan staan
   'Without hesitating, Sophie had jealously stepped on her neighbor's bag'
C. Zonder enige aanleiding begon Sophie jaloers op de tas van haar zus te spugen
   'Without any reason, Sophie started to spit jealously on her sister's bag'
D. Zonder goede reden begon Sophie jaloers op de tas van haar vriendin te worden
   'Without a good reason, Sophie started to get jealous of her friend's bag'

Voorzichtig met, 'careful with'

A. Na dat ongelukje deed Fleur voorzichtig met de vaas die enorm veel geld gekost had
'After that accident, Fleur was careful with the vase that had cost an enormous amount of money'

B. Na het schoonmaken deed Fleur voorzichtig met de vaas in haar handen een stap opzij
'After cleaning, Fleur carefully stepped aside with the vase in her hands'

C. Na het uitpakken liep Fleur voorzichtig met de vaas naar de vensterbank in de kamer
'After unpacking, Fleur carefully walked with the vase to the windowsill in the room'

D. Na het verhuizen liep Fleur voorzichtig met de vaas maar druk pratend naar de gang
'After moving, Fleur walked to the hallway, careful with the vase, but talking busily'

***Trots op,'* proud of'**

A. Net zoals zijn broer was Daan trots op de carnavalswagen waaraan ze maandenlang hadden gewerkt
'Just like his brother, Daan was proud of the carnival float they had been working on for months'

B. Met zijn blauwe pruik was Daan trots op de carnavalswagen geklommen voor hij eindelijk vertrok
'With his blue wig, Daan had proudly climbed onto the carnival float before it finally took off'

C. Ondanks de koude regen stond Daan trots op de carnavalswagen te zwaaien naar iedereen onderweg
'In spite of the cold rain, Daan stood waiving proudly on the carnival float to everyone on the way'

D. In zijn nieuwe kostuum stond Daan trots op de carnavalswagen tussen de mensenmassa te wachten
'In his new costume, Daan, proud of the carnival float, stood waiting between the crowd'

Curriculum Vitae

Maria Baukje Johanna Mos was born in Amsterdam on September 11 1978. She learned to read at the *Vijfde Montessorischool De Zilvermeeuw* and attended secondary school at the *Ignatius Gymnasium*. She studied Spanish and Linguistics at the *University of Amsterdam*, spending one year as an Erasmus student in Madrid at the *Universidad Autónoma*. In 2003, she graduated from a one-year Master's program in Linguistics with a thesis entitled *Word by word: The influence of semantic and form-based clustering on the acquisition of Spanish verbs* and in 2004 from a two-year Research Master's program in Linguistics with *Cognitive task complexity and second language writing performance*, both *cum laude*.

Between 2004 and 2009, Maria Mos worked as a Ph.D. candidate at the Department of Language and Culture Studies at Tilburg University. While finishing this manuscript, she has been teaching courses on topics ranging from statistics to cognitive linguistics at the Department of Communication and Information Sciences at Tilburg University and the Department of Language and Communication at VU University Amsterdam. Maria Mos is also currently the secretary of Anéla, the Dutch Association for Applied Linguistics. She lives in Amsterdam.

TiCC Ph.D. Series

1. Pashiera Barkhuysen. *Audiovisual prosody in interaction.* Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, October 3, 2008.
2. Ben Torben-Nielsen. *Dendritic morphology: function shapes morphology.* Promotores: H.J. van den Herik, E.O. Postma. Co-promotor: K.P. Tuyls. Tilburg, December 3, 2008.
3. Hans Stol. *A framework for evidence-based policy making.* Promotor: H.J. van den Herik. Tilburg, January 21, 2009.
4. Jeroen Geertzen. *Act recognition and prediction. Explorations in computational dialogue modeling.* Promotor: H. Bunt. Co-promotor: J.M.B. Terken. Tilburg, February 11, 2009.
5. Sander Canisius. *Structured prediction for natural language processing: a constraint satisfaction approach.* Promotores: A.P.J. van den Bosch, W. Daelemans. Tilburg, Februari 13, 2009.
6. Fritz Reul. *New Architectures in Computer Chess.* Promotor: H.J. van den Herik. Co-promotor: J. Uiterwijk. Tilburg, June 17, 2009.
7. Laurens van der Maaten. *Feature Extraction from Visual Data.* Promotores: E.O. Postma, H.J. van den Herik. Co-promotor: A.G. Lange. Tilburg, June 23, 2009.
8. Stephan Raaijmakers. *Multinomial Language Learning: Investigations into the Geometry of Language.* Promotores: W. Daelemans, A.P.J. van den Bosch. Tilburg, December 1, 2009.
9. Igor Berezhnyy. *Digital Analysis of Paintings.* Promotores: E.O. Postma, H.J. van den Herik. Tilburg, December 7, 2009.
10. Toine Bogers. *Recommender Systems for Social Bookmarking.* Promotor: A.P.J. van den Bosch. Tilburg, December 8, 2009.
11. Sander Bakkes. *Rapid Adaptation of Video Game AI.* Promotor: H.J. van den Herik. Co-promotor: P. Spronck. Tilburg, March 3, 2010.
12. Maria Mos. *Complex Lexical Items.* Promotor: A.P.J. van den Bosch. Co-promotores: Dr. A.M. Backus, dr. A. R. Vermeer. Tilburg, May 12, 2010 (in collaboration with the Department of Language and Culture Studies).

Ph.D. Series in Language and Culture Studies, Tilburg University

Sander Bax. De taak van de schrijver. Het poëticale debat in de Nederlandse literatuur (1968-1985). Promotor: J.L. Goedegebuure. Co-promotor: O.M. Heynders. Tilburg, May 23, 2007.

Tamara van Schilt-Mol. Differential Item Functioning en itembias in de Cito-Eindtoets Basisonderwijs. Oorzaken van onbedoelde moeilijkheden in toetsopgaven voor leerlingen van Turkse en Marokkaanse afkomst. Promotor: A.L.M. Vallen. Co-promotor: J.H. Uiterwijk. Tilburg, June 20, 2007.

Mustafa Güleç. Differences in similarities. A comparative study on Turkish language achievement and proficiency in a Dutch migration context. Promotor: Guus Extra. Co-promotor: Kutlay Yağmur. Tilburg, June 25, 2007.

Massimiliano Spotti. Developing identities. Identity construction in multicultural primary classrooms in The Netherlands and Flanders. Promotores: Sjaak Kroon, Guus Extra. Tilburg, November 23, 2007.

A. Seza Doğruöz. Synchronic variation and diachronic change in Dutch Turkish. A corpus based analysis. Promotor: A.J.A.G. Extra. Co-promotor: A.M. Backus. Tilburg, December 12, 2007.

Daan van Bel. Het verklaren van leesgedrag met een impliciete attitudemeting. Promotores: H. Verdaasdonk, W.L.H. van Lierop. Co-promotor: M.J.W. Stokmans. Tilburg, March 28, 2008.

Sharda Roelsma-Somer. De kwaliteit van Hindoescholen. Promotores: R.S. Gowricharn, J.F.A. Braster. Tilburg, September 17, 2008.

Gürkan Çelik. The Gülen movement. Building social cohesion through dialogue and education. Promotores: R.S. Gowricharn, H.L. Beck.Tilburg, December 12, 2008.

Yonas Mesfun Asfaha. Literacy acquisition in multilingual Eritrea. A comparative study of reading across languages and scripts. Promotor: Sjaak Kroon. Co-promotor: Jeanne Kurvers. Tilburg, November 4, 2009.

Dong Jie. The making of migrant identities in Beijing. Scale, discourse, and diversity. Promotores: Jan Blommaert, Sjaak Kroon. Tilburg, November 4, 2009.

Elma Nap-Kolhoff. Second language acquisition in early childhood. A longitudinal multiple case study of Turkish-Dutch children. Promotor: Guus Extra. Co-promotor: Kutlay Yağmur. Tilburg, May 12, 2010.

Maria Mos. Complex lexical items. Promotor: A van den Bosch. Co-promotores: A. Backus, A. Vermeer. Tilburg, May 12, 2010 (In collaboration with TiCC).