Morphological Effects in Danish Auditory Word Recognition

Laura Winther Balling

Ph.D. thesis
Department of English
Institute of Language, Literature and Culture
University of Aarhus

Morphological Effects in Danish Auditory Word Recognition

Morfologiske Effekter i Dansk Auditiv Ordgenkendelse

Laura Winther Balling

Ph.D. thesis
Department of English
Institute of Language, Literature and Culture
University of Aarhus

Supervisors

Professor Ocke-Schwen Bohn, Department of English, University of Aarhus
Professor Sten Vikner, Department of English, University of Aarhus
Professor Harald Baayen, Department of Linguistics, University of Alberta

Morphological Effects in Danish Auditory Word Recognition

Ph.D. thesis, Department of English, Institute of Language, Literature and Culture, University of Aarhus

© Copyright 2008 Laura Winther Balling.
E-mail: laura@wintherballing.dk

Acknowledgements

I am deeply grateful to my three supervisors: To Ocke-Schwen Bohn for encouragement and advice, for help with the experimental work, and for his interest in my work in spite of its sometimes peripheral connection to his own. To Sten Vikner for discussions on the Danish language and theoretical linguistics, and for his general academic advice. And to Harald Baayen for expert advice on the study of morphological processing and on statistics, and for friendship and hospitality.

Many people have contributed to the experimental work reported in this thesis. I am grateful to Rikke Slot Johnsen for her exemplary speaking of the auditory stimuli and for her ability to take the nonwords seriously; to Katrine Thielke and Johannes Kizach for checking the nonwords; and to all participants for sparing their time for the experiments.

Thanks to all colleagues at the English Department, and at the Institute of Language, Literature and Culture, particularly Lise Hedevang Nielsen and Ken Ramshøj Christensen. Also many thanks to the people at the IWTS at the Max Planck Institute in Nijmegen for their hospitality during my stays there; and particularly to Harald and the rest of the Baayen family for their support when I found myself away from home with an inflamed appendix. Thanks to Alex Klinge for hospitality at the Copenhagen Business School and to all at ISV.

My work has benefited from conversations with many people who have offered useful comments from a range of different perspectives. I am grateful to Raymond Bertram, Cathy Best, Mike Ford, Terry Gottfried, Christian Jensen, Alex Klinge, Victor Kuperman, Gary Libben, Petar Milin, Bolette Pedersen, Douglas Saddy, Lee Wurm and many others. I am indebted to Inger Mees, Arnt Lykke Jakobsen and Susanne Göpferich for comments on parts of chapter 4, and to Jonathan Grainger and two anonymous reviewers for comments on the article on which chapter 6 is based.

Parts of this work have been presented at various meetings and conferences. For questions and comments, I am grateful to the members of the network for Research on Cognition and Consciousness at the University of Aarhus; the members of the Linguistic Graduate School North; and to audiences at the 22nd Scandinavian Conference of Linguistics in Aalborg in June 2006; at Grammar in Focus in Lund in February 2007; at the 5th International Workshop on Morphological Processing in Marseilles in June 2007; at the

13th Conference on Architectures and Mechanisms for Language Processing in Turku in August 2007; at the Centre for Comparative Psycholinguistics at the University of Alberta in October 2007; at ISV-forum at the Copenhagen Business School in October 2007; and at the 12th Meeting about the Investigation of the Danish Language in Aarhus in October 2008.

I am grateful to my father for proof-reading, and to him and my mother for moral, nutritional and all manner of other support throughout.

Finally, thanks to my husband, Kristoffer Winther Balling, for showing me what computers can do and for helping me making them do it. Without this, and his continual support in all matters great and small, this thesis truly would not have come into existence.

Aarhus, November 2008.

Contents

List of Tables

List of Figures

List of abbreviations

ANOVA	Analysis Of Variance
CI	Credible Intervals
CRUP	Conditional Root Uniqueness Point
CUP	Complex Uniqueness Point
Df	Degrees of freedom
DP-Model	Declarative/Procedural Model of Ullman (2001)
ISI	Inter-Stimulus Interval
LDT	Lexical Decision Task
ms	Milliseconds
PCA	Principal Components Analysis
RT	Reaction Time
SD	Standard Deviation
SOA	Stimulus Onset Asynchrony
UP	Uniqueness Point

Danish words are italicized, with English translations between single quotation marks. English words used as examples are also italicized.

In complex words, hyphens mark morpheme boundaries.

Chapter 1

General introduction

1.1 Motivation and outline of the thesis

In normal speech, we are able to recognise two to three words per second (a range reported for production by Levelt, 1989: 22), identifying the vast majority correctly although each word has to be selected from a vocabulary of up to 150,000 words (Harley, 2008: 7). Both the speed and the accuracy of the word recognition process are fascinating, and it seems natural that the questions about how words are stored in long-term memory and how they come to be identified so quickly and accurately in speech and writing are important in psycholinguistics and cognitive psychology.

The study of morphologically complex words provides an interesting window on these questions, for a number of reasons: Firstly, and very basically, words consisting of several morphemes make up the majority of word types in many languages (Bertram, Baayen and Schreuder, 2000a). Secondly, morphology constitutes a level of structure between form and meaning, imposing regularity on the mapping between form and meaning in a way that may be understood as functional to word memory and word recognition in different ways (see section 1.4). The study of morphological processing thus provides a way of investigating both form and meaning and the relation between them. Thirdly, morphological operations are the main sources of new vocabulary, and thus one key to understanding productivity and creativity in language. Finally, the question of morphological processing is associated with several issues that are central to the study of cognition, including the contrast between symbolic and subsymbolic processing.

In the experiments reported in this thesis, morphologically complex words were presented to adult native speakers of Danish. The time it took participants to recognise these words was analysed to investigate the effects on processing time of different characteristics of the selected items, and in that way shed light on various aspects of the word recognition process. The focus of psycholinguistic experiments is necessarily narrow, relative to the breadth and variety of different processes that are involved in everyday language use — the demands of experimental control and set-up simply restrict the questions that can be addressed. The experiments reported in chapters 5 to 7 investigate the recognition of single words, largely ignoring sentential, pragmatic and contextual information. Most of these experiments investigate the auditory modality, mainly because this is ontogenetically and phylogenetically primary and therefore reflects more basic language processes. Also, some interesting aspects of processing pertain specifically to the auditory modality, especially relating to the temporal order in which the information becomes available in auditory word recognition. From a methodological point of view, listening skills are likely to differ less than reading skills, making participant groups more homogeneous for listening than for reading experiments. All experiments involve derivational morphology, but compare the recognition of derived words with simple, inflected and compound words. Like the experiments, the present chapter and the next are centred around derivational morphology, but necessarily also consider the processing of other types of words, particularly inflected words, which have formed a main empirical battleground for models of morphological processing over the last two decades (see section 1.3). The focus on Danish is unique: not only are there, to my knowledge, no previous studies of adult morphological processing in Danish, there are also no reaction time studies of adult word recognition in Danish. Although the focus of this thesis is on morphological processing, the experiments reported below were also constructed to be as informative as possible about general lexical processing in Danish. In addition to being a new language in this field, Danish has various interesting properties, including the function of particles as prefixes and the very productive use of compounding which results in large morphological families and in large cohorts of morphologically related words.

This chapter and the next describe the theoretical and empirical background for the experiments on morphological processing in Danish that are reported in chapters 5 to 7. The present chapter introduces the main theoretical questions and the most important psychological models of morphological processing. Chapter 2 presents and discusses main patterns in the experimental literature and evaluates the different models based on the empirical results. In both chapters, the focus is on recognition data and models, though other domains are also drawn in.

The next two chapters, 3 and 4, describe considerations and practical matters that were relevant for all experiments: Chapter 3 begins with a brief description of the morphological system and other aspects of Danish that are relevant to item choice and lexical variables. Then follows a description of the corpus of written Danish that was manipulated in order to extract the relevant lexical variables for the experimental items, including various frequencies, morphological family measures and uniqueness points. Chapter 4 motivates the choice of task, design and statistics for the experiments, describing the advantages of lexical decision, of a regression approach to experimental design and of linear mixed-effects modelling for data analysis.

After these general methodological chapters, chapters 5 to 7 report three experiments on Danish morphological processing. All three experiments include simple as well as morphologically complex words, partly because the comparison of the simple and complex words is informative with respect to the functionality of morphology. The study of simple words may also provide a baseline against which various effects for the morphologically complex words may be evaluated; this is relevant because of the lack of on-line studies of adult word recognition in Danish. The first experiment (see chapter 5) investigates the processing of suffixed, prefixed and particle prefixed words in both an auditory (Experiment 1a) and a visual (Experiment 1b) lexical decision task. The auditory Experiment 2 (see chapter 6) also includes derived forms but compares them to inflected as well as simple words. Experiment 3 in chapter 7 is also auditory and compares two types of prefixed derived words to compounds and simple words. In this way, the three experiments cover the three main types of morphological operations: inflection, derivation and compounding, as well as the proposed intermediate category between derivation and compounding, the particle prefixed words.

A main feature of all three experiments is the investigation of two new uniqueness points (UPs) for morphologically complex words. The traditional UP (see section 2.3.2) defines the point in time where a target word becomes uniquely distinguishable from other words in the given language. The new UPs describe two distinct types of lexical competition: first, competition from morphologically unrelated words, then competition between morphologically related words which is conditional on the first constituent having occurred. The way lexical processing is affected by frequency of occurrence is also investigated in all three experiments. Although the stem frequency of complex words has been seen as the hallmark of morpheme-based processing in the literature, both affix and whole-word frequencies are proposed as possible alternatives, based on an information-theoretical understanding of frequency effects. Finally, Experiment 1a showed a surprising inhibitory effect of morphological family size, which was therefore also investigated in Experiment 2 and, especially, Experiments 3a and 3b. Chapter 8 discusses the findings of all experiments jointly and relates the results to the existing literature.

In this chapter, section 1.2 introduces necessary terms and concepts. Section 1.3 discusses the role of morphology in word recognition and the broader questions about language and cognition that have been addressed in studies of morphological processing. The main models of morphological processing are introduced, both in terms of a continuum from full parsing to full storage and in terms of the related binary contrast between single- and dual-mechanism models. Section 1.4 describes ways in which morphology may be functional in language processing.

1.2 Terminology and definitions

What a word is seems intuitively obvious but is surprisingly difficult to define in a simple way that applies generally (cf. e.g. Chalker and Weiner, 1994: 426). Psycholinguistic studies usually involve words that are easily defined as such, being minimum free forms (a key definition according to Bloomfield, 1933), and the problems of definition are generally ignored. In the study of morphological processing, the word contrasts with the morpheme, which is defined as the smallest unit of meaning in the language. The theoretical construct of the morpheme as a discrete unit is problematic, as discussed in the next section, but it is descriptively useful. The term morpheme is used here to refer to both actual occurrences and the more abstract unit, while the term morphemic representation refers to the possible mental representation corresponding to this unit.

For a given morphologically simple or complex word, I use the term base form to refer to the citation form of the whole derived or simple word. The term lemma refers to all inflectional forms of a given word. In bimorphemic affixed words, root and stem can generally be used synonymously; the term stem is used here. Compounds consist of two immediate constituents, which can almost always function as independent words. In Danish, the first of the two immediate constituents is the modifier, while the second is the head, which determines the syntactic and semantic category of the whole compound. The terms first and second constituent are used rather than modifier and head for the bimorphemic compounds investigated in Experiment 3, because these terms may also refer to the affixed words in that experiment.

Following Jarema and Libben (2007), the mental lexicon is understood as the "cognitive system that constitutes the capacity for [...] lexical activity" (p. 2). The mental lexicon stores information about words, their form and meaning and syntactic, pragmatic and collocational properties. The role of morphological structure in this mental lexicon is the object of the present study. The nature of lexical representations is debated, as outlined in the next section. The process of activating entries in the mental lexicon is lexical access. The term lexical access is often used with the implicit assumption of a magic moment of word recognition (Balota, 1990) and a strict division into prelexical processing of the form and a postlexical stage at which meaning of the word is suddenly available. Lexical processing refers more broadly to processes involving the mental lexicon. Morphological processing is lexical processing of words that are morphologically complex; if such processing relies specifically on mental representations of morphemes or on morphemic information, the term morphemic processing is used.

1.3 The role of morphology in psycholinguistics

At its most basic, the question asked in research on morphological processing is what the role of morphology is. This question is usually framed in terms of a contrast between storage and computation, i.e. between whether morphologically complex words are stored as wholes in the mental lexicon or computed from representations of their constituent morphemes. This section introduces and discusses the ways different models describe the role of morphology in word recognition, based on a contrast between storage and computation.

The storage vs. computation contrast defines a continuum of models of morphological processing in word recognition, ranging from full-parsing models that favour obligatory decomposition of morphologically complex words, to full-storage models that argue in favour of storage of all words whether they are morphologically simple or complex. Full parsing means that a morphologically complex word like *govern-ment* is recognised through the morphemes *govern* and *-ment*, while full-storage models argue that recognition relies on a stored representation of the whole word *government*. Between these two extremes are various dual models which involve both morphemic and whole-word based processing.

The full-parsing position is usually exemplified by the prefix-stripping model of Taft and Forster (1975). In this model, lexical access is a decompositional process in which the prefixes of prefixed words are stripped off and recognition is based on the stem. Although the focus is on prefixed words, obligatory decomposition is also assumed for suffixed words. After the complex words have been decomposed and recognised through their morphemes, the morphemes are recombined for interpretation of the whole word, so there is an element of dual representation in this model, which is most explicit in later versions (e.g. Taft, 1994, 2004).

For the opposite end of the continuum, the full-storage position, the classic reference is Butterworth (1983) who argues that morphologically complex words are listed as full forms in the mental lexicon with no internal morphological structure. However, connectionist single-mechanism models, which are discussed more extensively below, represent a more contemporary version of this view. According to these models (e.g. Seidenberg and McClelland, 1989; Joanisse and Seidenberg, 1999; Seidenberg and Gonnerman, 2000), morphologically complex words are stored as whole words, just like simple words. These models do not deny the existence of morphological effects in word recognition, but argue that they emerge because morphologically related words tend to overlap both in meaning and in phonological and orthographic form. In other words, morphological effects are seen as epiphenomena of semantic and form-related factors.

The decompositional model of Taft (1994) may also be contrasted with the Cohort Model of Marslen-Wilson and colleagues (Marslen-Wilson and Welsh, 1978; Marslen-Wilson, 1987, 1990; Gaskell and Marslen-Wilson, 1997). Taft's work has mainly been concerned with the visual modality, but some of the predictions of his prefix-stripping model were contrasted for auditory recognition with the predictions of the Cohort Model by Tyler, Marslen-Wilson, Rentoul and Hanney (1988) and Schriefers, Zwitserlood and Roelofs (1991; see section 2.3.2). The Cohort Model is not primarily a model of morphological processing, but it is an influential model of spoken word recognition. It is relevant for morphological processing because it operates with continuous processing of the incoming auditory signal, with no role for decomposition into morphemes. The continuous access process ignores the morphological structure of words, but the uniqueness point construct, which originates within the paradigm of the Cohort Model, may be adapted to take morphological structure into account. The most important insight of the Cohort Model is that lexical candidates are activated in parallel and that competition between them is extensive.

Between these extreme models, there are various hybrid positions. These include the class of so-called dual-route models, as well as other dual positions. The term dual is used in different ways in the literature. The most general use, adopted here, is for models that operate with both morphemic and whole-word representations. Dual in this sense distinguishes between different processes and/or representations, which may be simultaneously active, while "dual" in "dual-mechanism" refers to two distinct cognitive systems. Finally, single vs. dual-route models of reading are concerned with the question of whether reading must involve phonological representations, a single reading route as defended by e.g. Lukatela and Turvey (1994a,b) and Plaut, McClelland, Seidenberg and Patterson (1996), or can also proceed via a direct route from orthographic to semantic representations, i.e. dual reading routes as posited by e.g. Coltheart, Rastle, Perry, Langdon and Ziegler (2001). Like the models of morphological processing, the reading models are concerned with the question of rule-based vs. statistical mapping. The reading models are of peripheral relevance here.

Dual-route models of morphological processing argue that word recognition can proceed via a whole-word route which processes complex words based on whole-word representations or via a decompositional route which achieves recognition through decomposition. The route metaphor implies a focus on lexical access: the different types of representations are typically associated with the process of accessing more central semantic representations. Dual-route models can be divided into at least three types: the routes may operate consecutively, they may race or they may interact. In consecutive dual-route models like the Augmented Addressed Morphology Model of Caramazza, Laudanna and Romani (1988), the morphological route is secondary, operating only if the whole-word route fails as it is predicted to do for morphologically complex words that are novel or of low frequency. In the Morphological Race Model of Frauenfelder and Schreuder (1992), the two routes are activated in parallel and the fastest route achieves word recognition. A variety of characteristics of morphologically complex words (reviewed in chapter 2) determine whether both routes are activated and which route wins the recognition race, see e.g. Schreuder and Baayen (1995). In race models, words that are processed simultaneously via both routes may be at an advantage, due to statistical facilitation (Frauenfelder and Schreuder, 1992 based on Raab, 1962): if one route happens to be slow, the other may be fast. Finally, the two routes may interact in dual-route models, such that the operation of one route is not independent of the other route. For instance, Wurm (1997) suggests that the morphemic route cannot commit to a lexical candidate but that it operates in parallel with the whole-word route and may help the whole-word route process the correct candidate.

Hybrid models on the continuum may be dual without operating with dual routes: As mentioned, the model of Taft (1994, 2004) operates with obligatory morphological decomposition, but also with some element of whole-word representation at a later stage of recognition. In contrast, the supralexical model of Giraudo and Grainger (2000, 2001, 2003b) posits whole-word representations at an early stage of processing, the lexical level, while morphemic representations come into play on a later, supralexical level. The supralexical representation of morphological structure contrasts not only with Taft's model, in which morphology is sublexical, but also with the dual-route models described above in which morphemic representations are activated simultaneously with whole-word representations. Another example of dual representation is the position of Marslen-Wilson, Tyler, Waksler and Older (1994) and Marslen-Wilson and Zhou (1999) which includes both whole-word and morphemic representations on a central, semantic level of the mental lexicon, based on a very simple binary distinction: Words that are semantically transparent (i.e. cases such as *govern-ment*, see further in section 2.3.3) are represented as the morphemes *govern* and *-ment* in the central lexicon, while semantically opaque words (e.g. *depart-ment*) are represented as whole words with no internal morphological structure.

Several of the models described are concerned with the relation between different levels of the mental lexicon, distinguishing between access representations of the form of words and central, more semantic representations which may be modality-independent (Marslen-Wilson et al., 1994). A more neutral, model-independent way of addressing these levels is in terms of different stages of word recognition, early processes being more form-based and later processes activating semantics. Many researchers have argued in favour of specific access representations in the visual modality (e.g. Schreuder and Baayen, 1995; Caramazza et al., 1988). In the auditory modality, the idea of specific access representations is problematic (Marslen-Wilson and Zhou, 1999). When a word becomes gradually available over time, semantic processing is unlikely to be deferred until the form of the whole word has been processed. The time-course of processing can be straightforwardly studied for the visual modality by eye-tracking participants while they are reading. Eye-tracking can also be used to study auditory processing, e.g. in the visual world paradigm which is described in section 2.2, but this task has a narrower applicability. The time-course of processing is not directly studied in the auditory experiments on Danish reported below, but the time-course of lexical competition is considered by investigating the state of the competition cohort at different points in the word. Some evidence of the temporal dynamics of visual word recognition is discussed in chapter 2, primarily to address the question of sublexical vs. supralexical morphemic representations.

The models on this continuum, particularly the dual models, are relatively broad in the sense that they investigate different types of morphology in a range of different languages. In another sense, they are narrow, because they focus exclusively on word recognition. The opposite can be said about the contrast between single- and dual-mechanism models: Like the continuum, these models can be understood in terms of the contrast between storage and computation, but they differ from the models described above in focusing more narrowly on inflectional morphology and the dichotomy between regular and irregular. The single- and dual-mechanism models also rely more on a single language, namely English, though other languages are studied within this perspective, for instance German by Clahsen (1999) and Hahn and Nakisa (2000). The label "past-tense debate", which is often used for this contrast (Pinker and Ullman, 2002; McClelland and Patterson, 2002), derives from this focus on contrasts between regular and irregular English past-tense forms. In this sense, the contrast between the single- and dual-mechanism models can be understood as a subdivision of the full-storage to full-parsing continuum. However, in another sense the single- and dual-mechanism models are much broader than the word recognition models described above: they include word production as well as recognition, and draw in neurolinguistics, language acquisition and developmental and acquired language deficits to a much larger extent. Moreover, these models make far-reaching claims about the nature of language and cognition based on morphological processing. The word recognition models described above are primarily concerned with investigating how morphological processing works in word recognition and secondarily with the wider implications for language and cognition. In contrast, the single- vs. dual-mechanism models focus on the broad implications of storage vs. computation in language and cognition. Within the dual-mechanism model, the distinction between regular and irregular processes is held to be absolutely central to the language faculty, and the processing of regular and irregular past tense forms is argued to be an ideal way of demonstrating these differences.

In this way, the continuum of word recognition models described above and the single- vs. dual-mechanism models discussed below are to some extent two traditions, but they are related in many ways: Firstly, both traditions can be framed in terms of the contrast between storage and computation of morphologically complex words. Secondly, the single- and dual-mechanism models draw on evidence from studies that focus on word recognition, while the narrower word recognition models sometimes rely on the sharp binary distinctions of the past-tense debate. And thirdly, the single-mechanism models function both as the contemporary endpoint of the word recognition continuum and as one pole of the past-tense debate. The data as well as the similarities and differences between the models are nonetheless clearer if they are understood as two traditions because of the difference in the aim and scope of the different models.

On the computation side of the past-tense debate is the model variously known as the dual-mechanism model (Pinker, 1999), Words and Rules theory (Pinker and Ullman, 2002) and the Declarative/Procedural Model (Ullman, 2001, 2004). The focus here is on the Declarative/Procedural Model (henceforth abbreviated to the DP-Model) because this is most recent and developed in most detail, but the main points are relevant to all versions of the dual-mechanism approach.

The core of the DP-Model is a distinction between a mental lexicon of words with arbitrary form-meaning relations and a grammar of rules. The items in the mental lexicon are stored in declarative memory and processing is associative, i.e. based on similarity. The grammatical rules belong to the procedural memory system and operate independently of similarity. Words and grammatical rules are generally difficult to compare directly, but it is argued by proponents of the DP-Model that the distinction between regular and irregular inflectional morphology provides a testing ground for the hypothesis of separate systems. Regular and irregular past-tense forms share many properties — being words, being verbs and expressing past tense — but processing of regular forms may be rule-based while processing of irregular forms cannot be rule-based. The DP-Model posits that regular morphologically complex words are processed by the application of a combinatorial rule in procedural memory, e.g. processing the regular past tense *walked* proceeds through the morpheme *walk* and the -*ed*-rule. Rule-based processing is symbolic in the sense that the present tense from which a past tense is assumed to be formed is treated as a symbol: the rule forms the past tense without reference to the form or meaning of the present tense. Irregular past tense forms such as *came* are stored in declarative memory, like simple words, and linked to their present-tense forms, i.e. *come*, by formal and semantic similarity rather than by rule. The past-tense rule functions as a default: if no specific past-tense form is stored in declarative memory, the rule applies by default. The distinction between regular and irregular is categorical.

Proponents of dual-mechanism models (e.g. Pinker, 1999) sometimes describe the dual-mechanism models as the intermediate position on a continuum ranging from exclusively associative models to the entirely rule-based account of Generative Phonology (Chomsky and Halle, 1968; Halle and

Mohanan, 1985) according to which both regular and irregular forms are processed based on rules. In psycholinguistics, however, the DP-Model constitutes the extreme opposite of the associative models, rather than an intermediate position, since Generative Phonology is a linguistic rather than a psycholinguistic model. The very abstract underlying representations that are necessary within Generative Phonology seem psycholinguistically implausible (Pinker and Ullman, 2002).

Contrasting with the DP-Model, there are various associative single-mechanism models such as the connectionist Triangle Model of Seidenberg and McClelland (1989) and Joanisse and Seidenberg (1999), and the Convergence Account of Seidenberg and Gonnerman (2000) and Gonnerman, Seidenberg and Andersen (2007); the latter is inspired by connectionism, but not computationally implemented. Connectionist models (for a brief description, see for instance McClelland, 1999) operate with a large number of simple processing units which are inspired by, but do not correspond to neurons in the brain. In distributed models, mental representations — such as words in the mental lexicon — are represented as patterns of activation across these simple processing units, rather than as discrete representations of the individual words. In models of the mental lexicon, groups of processing units represent form and meaning, which are linked through connections with different weights that determine the associations of specific forms with specific meanings. In the connectionist single-mechanism models, all morphologically complex words are fundamentally processed in the same way as simple words, and the mapping from present to past tense forms is the same for regular and irregular verbs. All past-tense forms are stored in associative memory, rather than formed by a rule, and the production and recognition of new past-tense forms are based on formal and semantic similarities to existing stored forms, irrespective of regularity. Processing differences between regular and irregular complex words, which are claimed by the DP-Model to be evidence of two distinct mechanisms, are argued within the single-mechanism models to arise from the graded effects of semantic and formal similarity. Like the DP-Model, many single-mechanism models regard inflectional morphology as a testing ground for broader questions: the rejection of symbol-manipulating rules for regular inflectional forms generalises to other areas of language and cognition (McClelland and Patterson, 2002), notably syntax.

The main empirical question of the past-tense debate is whether the difference between regular and irregular morphology is categorical or graded. This is assessed in a range of different domains. In word recognition, the focus is on different frequency and priming effects; these effects are reviewed in chapter 2 and the evaluation of the models in section 2.7 is based on that. Some of the results of the word recognition experiments reported in chapters

5 to 7 are also relevant to the past-tense debate; the implications of the results from Danish for these models are discussed in the relevant experiment chapters and synthesised in chapter 8.

Because the processing of inflectional morphology is regarded as a testing ground for general models of language and cognition, the arguments for categorical vs. graded effects become associated with specific theoretical positions. In terms of linguistic theory, the dual-mechanism approach is associated with generative linguistics; the dual-mechanism distinction between regular and irregular arguably originates in the Lexicalist Hypothesis first put forward by Chomsky (1970), according to which there are separate systems for syntax, including inflection, and the lexicon including derivation. Single-mechanism models are more in line with the tenets of functionalist linguistics (exemplified for instance by the schema-based model of Bybee, 1995). The issue of categorical vs. graded effects is also closely connected with symbolic vs. subsymbolic representation that characterises the dual- and single-mechanism models, respectively. The dual-mechanism model is symbolic because the rule operates on a single lexical entry independently of similarity, while the single-models are subsymbolic because representations are patterns of activation across several simple processing units. However, as argued by Hay and Baayen (2005), the association of graded effects of morphological structure with subsymbolic representation is not a necessary one (a related point is made by Marslen-Wilson and Tyler (2003)): There is strong empirical evidence of graded effects of morphological structure, as shown in the literature review in chapter 2 and in the experiments reported in chapters 5 to 7, but this does not necessarily mean that processing and representation is subsymbolic and analogy-based. For instance, Albright and Hayes (2003) propose a model which is neither analogy-based, nor involves a single symbolic rule, but instead operates with multiple stochastic rules with different confidence scores depending on probabilities. Thus, Albright and Hayes operate with rules, but probabilistic rules that can account for graded effects. Connectionist single-mechanism models are only one way of computationally implementing the probabilistic nature of morphological processing (for alternatives, see also for instance Skousen, Lonsdale and Parkinson, 2002; Daelemans, Zavrel, van der Sloot and van den Bosch, 2003; Keuleers, 2008). Following Hay and Baayen (2005) and Marslen-Wilson and Tyler (2003), I distinguish the empirical questions from those theoretical issues with which the empirical questions are customarily but not necessarily associated.

Graded effects of morphological structure in empirical studies are often equated with the rejection of any independent role for morphology in language processing (e.g. Seidenberg and Gonnerman, 2000). However, this challenge to the role of morphology is based on morphemes as discrete units, a position which is somewhat problematic within morphological theory as well as in psycholinguistics. In structuralist linguistics (Bloomfield, 1933), the morpheme is standardly defined as the smallest meaningful unit of language. The validity of the morpheme-construct is most intuitively convincing for regularly inflected words in agglutinating languages: For example, in a Danish plural form such as *blyant-er* ('pencil-s'), two discrete morphemes can be identified and, at least for the written form *blyanter*, each of these morphemes represents a clearly defined form and a clearly delimited meaning. The morpheme-construct becomes more problematic when considering phenomena such as phonaesthemes. These are forms which recur in clusters of words with similar meanings, such as the words starting with *gl-* in English which have meanings relating to light or vision (Bergen, 2004). They cannot be defined as morphemes in the traditional sense, since no complete morpheme-parse is possible for words such as *glimmer* or *glow*. Nonetheless, Bergen (2004) reports priming effects that are similar to morphological priming effects for pairs of words sharing a phonaestheme. Within theoretical morphology, there is a range of other exceptions from the clear one-to-one mapping of form and meaning in individual morphemes which are necessary for a morpheme-based account (see e.g. Bauer, 1999; Spencer, 2001; Haspelmath, 2002).

An alternative understanding of morphology is based on paradigms: Morphology is not understood in terms of combination of discrete morphemes, but in terms of whole complex words and the relations between these complex words within morphological paradigms. Since the role of morphology rests on relations within paradigms, paradigms become theoretically important rather than descriptive tools. Both inflection and derivation are understood as operations on whole words, agglutination being one such operation, but not the one operation from which all others must be derived. Word-and-Paradigm approaches, to use the term of Hockett (1954), have become increasingly popular in theoretical linguistics, cf. for instance Matthews (1991), Anderson (1992), Spencer (2001) and Stump (2001). The paradigm-approach not only avoids the problem of defining discrete morphemes, but also the problem of zero-morphemes that are posited in morpheme-based theories (e.g. in the Word Syntax approach, cf. Lieber, 1992; Toman, 1998), and whose psychological and scientific validity may often be questionable. Instead, within a Word-and-Paradigm approach, all forms, including those that do not carry an overt morpheme, can be defined relative to other forms in the paradigm.

Paradigm effects are also emerging in the experimental literature, most prominently in the shape of various effects of morphological family (see section 2.5). Word-and-Paradigm approaches can account for the role of morphology in psycholinguistics without relying on discrete morphemes that may make it difficult to understand and explain graded effects. Within a paradigmatic framework, morphological effects in processing are not necessarily evidence of separate mental representations of morphemes. Such mental representations may develop, but depend on paradigmatic support from whole-word exemplars (Hay and Baayen, 2005). An exemplar is a memory trace of an experience with a given word, including often quite subtle details of the instance and its context. In exemplar-based theories (e.g. Goldinger, 1996b), all experiences leave traces which decay at some rate; processing is based on generalisations across these exemplars. Storage of exemplars and paradigmatic generalisations across these exemplars may account for morphological processing without postulating abstract morpheme or word representations. The morphological effects discussed in this thesis may be understood in terms of paradigmatic relations between whole words, but the term morpheme remains a useful shorthand for referring to the morphological structure of complex words.

One way of understanding morphological processing as graded and probabilistic relies on information theory (e.g. Moscoso del Prado Martín, Kostic and Baayen, 2004b). Information theory does not in itself provide a cognitive model of morphological processing, but a way of understanding a range of phenomena including graded effects of morphological structure and interactions between morphemic and whole-word information discussed in the next chapter. A basic insight of information theory is that less probable events are more informative; information-theoretical accounts of word recognition are based on the assumption that less probable and hence more informative words or constituents take longer to process (Kuperman, Bertram and Baayen, 2008a). The Probabilistic Model of Information Sources (PROMISE) of Kuperman et al. (2008a) combines absolute probabilities of whole words and their constituents with conditional probabilities of different constituents, allowing interactions between these different sources of information. In terms of the continuum of word recognition models, PROMISE is an interactive simultaneous dual-route. The Bayesian Information-Theoretical Model of Moscoso del Prado Martín (2007) attempts to link the effects of information-theoretical measures on behaviour to the neurological model developed by Pulvermüller (1999). These phenomena and the information-theoretical interpretation of them are outlined in chapter 2.

1.4 The functionality of morphology in lexical processing

Taking the contrast between storage and computation as a starting point, morphological structure in the mental lexicon can be understood as functional because it saves storage space: in principle, morphemes that recur in many words need to be stored only once, so morphological structuring of the vocabulary may reduce storage. This metaphor of storage economy has obvious roots in the limited storage space of the computers of the 1970's and 80's (Libben, 2006). Another source of this understanding of morphology as storage-saving could be the drive of linguistic morphology to provide the most parsimonious analysis of the morphological systems of different languages, in accordance with the demands of Occam's razor. Parsimony of storage results in increased demands on computation: the morphologically complex words that are stored as morphemes rather than through whole-word representations must be assembled for interpretation in recognition, and for pronunciation in production. Conversely, computational demands can be reduced if all complex words are stored, but in that case storage efficiency is low.

As is clear from the previous section, morphological processing is often understood in terms of storage vs. computation, but it is becoming increasingly questionable whether this computer metaphor is appropriate for the way the human mind and human language work (Baayen, 2007). There is evidence of parallel processing in word recognition (McQueen and Cutler, 2001) and redundancy in language is extensive (cf. e.g. Pinker and Jackendoff, 2005). Moreover, evidence of simultaneous effects of whole-word properties, indicating storage, and morphological properties, indicating computation, is presented in chapter 2. This evidence leads to a new understanding of the functionality of morphology in the mental lexicon, not in terms of storage efficiency or computational efficiency, but in terms of processing efficiency and robustness, what Libben (2006) terms maximisation of opportunity. In this view, the recognition of morphologically complex words does not proceed either through morpheme or whole-word representations but is based on both morphemic and whole-word information, maximising the opportunity for word recognition by employing as many sources of information as are available. This may lead to redundancy in both processing and representation but it should also lead to more robust processing. The dual-route models which can be interpreted within the context of the storage vs. computation contrast may account for this functionality, but only if the routes are understood as simultaneously active. The idea of maximisation of opportunity accounts for the functionality of morphology in existing vocabulary and also meets the productivity constraint (Frauenfelder and Schreuder, 1992): like existing complex words, the recognition of novel complex words draws on all available resources.

The processing efficiency provided by morphological structure of the vocabulary can be understood in terms of both global and local functionality. Firstly, morphological structure may be globally efficient in the mental lexicon because it contributes to a partial alleviation of the arbitrariness of the linguistic sign. That the word *bog* means 'book' in Danish is, of course, arbitrary (De Saussure, 1916/1966). However, once the meaning of *bog* is established, the meaning of a morphologically complex word like *bog-handel* ('bookshop') is no longer entirely arbitrary. The connectivity in the mental lexicon between morphologically related words reflects relations between their referents. The alleviation of arbitrariness may result in a reduced memory load and in a mental lexicon which is globally more efficient.

Secondly, morphology may be locally efficient in that the individual acts of recognising morphologically complex words can rely on both morphemic and whole-word information. The latter seems to be closest to Libben's idea of maximisation of opportunity and it is also the kind of functionality that can be established based on experimental work, though effects such as the morphological family size effect may also reflect a global functionality. Further, recognition of individual morphologically complex word may also benefit from the possibility of processing a second constituent of a bimorphemic word given the first constituent that has already been processed.

This view of the functionality of morphology is investigated in the experiments reported in chapters 5 to 7 below, in three main ways: First, the processing of morphologically simple and complex words are compared. If morphological structure is functional, the recognition of morphologically complex words should be more robust and possibly also faster than the recognition of simple words. Secondly, two new UPs are introduced, to investigate the extent to which the processing of second constituents is conditional on the processing of the first constituent for all types of morphologically complex words. Thirdly, the simultaneous influence of whole-word and morphemic information is assessed, in the shape of whole-word and constituent frequencies and interactions between them.

Chapter 2

Characteristics of morphological processing

2.1 Introduction

This chapter reviews current knowledge of morphological processing in sections 2.3 to 2.6. Throughout, the focus is mainly on those aspects that are investigated in the experiments reported below. This chapter is intended as a relatively broad background for the experiments on Danish; more specific details that are relevant to the interpretation of the experimental results are brought up in the different results-sections and in the general discussion in chapter 8. In section 2.7, the experimental evidence discussed in this chapter is used to evaluate the different models that were outlined in chapter 1. As a background to the discussion of the experimental literature, the methods used for studying word recognition are briefly reviewed in section 2.2; this also to some extent forms the background for the choice of task for the present experiments, which is discussed in chapter 4.

An important distinction in language use and, more or less explicitly, in the study of lexical processing is between auditory and visual word recognition. However, a number of empirical effects are the same or similar in the two modalities, so the present discussion is not strictly divided into an auditory and a visual part; instead the differences between the modalities are considered wherever relevant. A main division applied here is between effects that are mainly related to the form of words and effects that are mainly related to the meaning of words. The effects of formal similarity probably constitute the area where differences between auditory and visual processing are most salient, and special attention is paid to modality differences in section 2.3. Effects of semantic transparency and relatedness are presented in section 2.3.3. Section 2.4 discusses frequency effects, focusing on the distinction between stem and whole-word frequency. The effects of morphological family measures are discussed in section 2.5. The last section that deals explicitly with a set of empirical effects, section 2.6, is the most diverse, describing differences and similarities between different types of words and different types of morphological operations.

The present chapter does not constitute a general review of word recognition but rather a review of those aspects which are particularly relevant to the recognition of morphologically complex words. These include specifically morpheme-related properties like for instance the semantic transparency of complex words, as well as those effects for which non-morphological measures contrast with morphological measures as predictors of the recognition of morphologically complex words. Examples of such contrasts are the distinction between whole-word (i.e. non-morphological) and constituent frequencies, and the distinction between the traditional, a-morphological uniqueness point (UP) and the elaborations of the UP here and elsewhere which draw in the morphological structure of complex words. Finally, non-morphological effects are also reviewed when they are specifically relevant as control variables in the experiments described in chapters 5 to 7.

2.2 Methods for studying lexical processing

This section provides a brief review of the most important tasks used to study lexical processing, as a background for understanding the empirical effects discussed in the subsequent sections. The focus is on how the different tasks work and what they measure, while the specific experimental manipulations of for instance various item characteristics are discussed in the sections devoted to the different experimental effects. One exception to this is the experimental manipulation known as priming; this is presented in the present section because it is a very general manipulation which can be used to assess both formal and semantic effects.

An important distinction between different experimental tasks is whether they are on-line or off-line, i.e. whether they measure lexical processing mainly during recognition, as the lexical decision task does, or after recognition has been achieved, such as for instance rating tasks do. Off-line tasks are more susceptible to influence from processes which are not directly related to the word recognition process, but may nonetheless yield interesting insights into the structure of the mental lexicon. Also tasks characterised as on-line contain components which are not parts of word recognition proper, so the distinction between on-line and off-line is not strictly binary. In this thesis, the main experiments are relatively on-line (using lexical decision), while off-line rating tasks were used to assess background variables.

The dominant task used to study morphological processing is probably the lexical decision task (LDT); certainly, most of the results discussed in this chapter come from studies employing lexical decision. The task is also extensively used in the study of word recognition more generally. In LDT, a mixture of real words and nonwords are presented either auditorily or visually, and participants are asked to decide whether each token is a word in the given language or not. Lexical decision is generally speeded but may be more or less so, depending on whether the time available on each trial is relatively long or short (a factor which plays a role in the experiments reported below). Reaction time (RT) is generally the most important dependent variable. Response accuracy is also analysed, but errors are usually few, making error analyses less powerful. Error analyses are therefore mostly used to confirm general patterns in RT-analyses and to assess whether there are speed-accuracy trade-offs. In most cases (including the experiments reported below), it is the responses to real words that are studied, but some experiments focus instead on the processing of nonwords, studying the time it takes to reject different nonwords (e.g. Taft and Forster, 1975; Caramazza et al., 1988; Wurm, 2000).

Lexical decision is a relatively artificial task that does not directly correspond to any everyday act of language processing (Goldinger, 1996a), and it arguably includes a post-access decision stage (Seidenberg, Waters, Sanders and Langer, 1984) which is not part of word recognition proper. This decision stage of LDT may inflate frequency effects relative to other tasks (Balota and Chumbley, 1984), but frequency effects are also found in tasks such as naming (e.g. Baayen, Wurm and Aycock, 2007), gating (e.g. Wurm, 1997) and reading measured by eye-tracking (e.g. Kuperman et al., 2008a), so the frequency effects in lexical decision are not exclusively artefacts of the task. The LDT is sensitive to the nature of the nonwords and the composition of the experimental list, so it is necessary to consider carefully how to construct the nonwords and how to balance the experimental list. The task is widely used because it is relatively easy to implement and because it measures recognition processes on-line, though the decision stage probably has little to do with the on-line measurement of word recognition (Balota and Chumbley, 1984; Seidenberg et al., 1984).

Word recognition can rely on other decisions than whether a token is a word or a nonword. In languages with grammatical gender, one option is gender decision (Radeau and van Berkum, 1996). Another decision task is semantic categorisation, where participants decide whether words presented belong to a particular semantic category or not. This task assesses semantic competition from orthographic or phonological neighbours, a fact utilised in the study of Bowers, Davis and Hanley (2005), see section 2.3.1. Like lexical decision, gender decision and semantic categorisation are relatively artificial tasks, involving a decision stage which is not a part of normal word recognition.

The naming task can be both auditory and visual: participants either hear or read a word and then say the word out loud. In other words, visual naming is reading aloud, while auditory naming is simple repetition. Picture naming, where participants are required to supply words corresponding to pictures shown to them, is used to study language production and therefore not relevant here. The dependent variables in naming are RT and correctness. Naming involves both recognition and production of the word presented; this means that effects observed in naming tasks may arise from other processes than word recognition, but that problem can be solved by converging evidence from tasks that more exclusively tap recognition. Particularly visual naming has the advantage of being a relatively natural task. Relative to lexical decision, it tends to tap more form-related aspects of word recognition, though frequency effects are also found in naming and form effects also in lexical decision. Visual naming tends to be more informative than auditory naming (see for example the parallel visual and auditory naming experiments in Baayen et al., 2007), making naming most suitable for use in studies of visual word recognition.

Eye-tracking while participants are reading also involves a relatively natural task, particularly with modern eye-tracking apparatus which is less noticeable than older equipment. This can be combined with other tasks that measure word recognition, cf. Kuperman, Schreuder, Bertram and Baayen's (2008b) combination of lexical decision with eye-tracking. In addition to the advantage of being a natural task, eye-tracking has the advantage of providing a very detailed picture of the time-course of visual word recognition. Eye-tracking can also be used to study auditory recognition processes, in the so-called visual world paradigm. In this paradigm, participants hear a target word and are presented with a set of pictures on a screen, of which one corresponds to the target word while the other are distractors which may be related to the target in various ways. The participants are asked to click on the picture that corresponds to the target word with a computer mouse, and their eye-movements are tracked. This technique is for example used to study neighbourhood and uniqueness point effects by Magnuson, Dixon, Tanenhaus and Aslin (2007).

Gating is an off-line task which is only used auditorily, but it is in some ways parallel to progressive demasking of a visual stimulus. In gating, participants listen to a succession of gradually larger chunks of each target word, so-called gates (see Grosjean, 1980, 1996). After each gate has been played, participants respond which word they think they have heard and rate their confidence that their guess is correct; recognition is defined as the gate where participants with a certainty above some (high) percentage identify the correct word without changing this identification at subsequent gates. In progressive demasking (Grainger and Segui, 1990), a word and a mask of hash marks are presented alternately in a number of equally long cycles; gradually, the word is presented for longer and the mask for shorter parts of the cycle and the participants press a button whenever they recognise the word.

2.2.1 Priming

Priming is an experimental manipulation rather than a task, but it is so widespread in the study of morphological processing that it merits brief presentation in this more general section. Most of the priming studies discussed in the present chapter employ lexical decision, though the priming manipulation can also be used with other tasks (e.g. Rueckl, Mikolinski, Raveh, Miner and Mars, 1997). Priming studies measure the effect of the presentation of one token, the prime, on the recognition of another token, the target. Priming effects arise when the target is recognised faster when encountered after the prime, which is related to the target in some way, relative to a condition where the target is encountered after a control prime, typically an unrelated word (but see Pastizzo and Feldman, 2004 who argue in favour of orthographic baselines). The difference between the related and the control prime is usually manipulated between groups of participants, such that each participant only encounters each target once.

The relation between prime and target can be formal (orthographic and/or phonological), semantic, morphological or pseudo-morphological. Effects of these different types of overlap are assumed to indicate overlap or relatedness between the mental representations of the primes and targets. However, this assumption is not entirely unproblematic: facilitatory priming effects may be due to episodic memory traces left by the primes which do not reflect shared abstract representations (Feustel, Shiffrin and Salasoo, 1983). This problem is argued to be more serious when target presentation is delayed relative to prime presentation, while it is alleviated by the use of a cross-modal task where prime and target are presented in different modalities (e.g. Marslen-Wilson et al., 1994). Cross-modal priming also avoids the influence of overlap at a low level, i.e. letter overlap in visual studies.

An important variable in priming studies, which has practically spawned a whole literature of its own, is the delay between the presentation of prime and target (which also figures prominently in section 2.3.3 below). Priming may be immediate, delayed or long-term. In immediate priming tasks, where no items are presented between the prime and the target, the stimulus onset asynchrony (SOA) and prime duration have been the subject of considerable attention. Prime durations between 13 and 72 ms may be categorised as subliminal, often with the prime occurring at the end of a pattern mask. This technique is referred to as masked or subliminal priming in the following, while the term immediate is generally used to refer to cases where the prime presentation is long enough for conscious processing. In delayed priming tasks, several items intervene between prime and target, while long-term priming embeds the prime in one task and the target in another (e.g. Rueckl et al., 1997).

How a prime affects the recognition of a target is in a sense a context effect, but the context is very different from word context in everyday language use. Whether prime presentation is subliminal or not, the prime is processed by the word recognition system. This process has a wide range of consequences, only some of which are intended and have anything to do with the target recognition; essentially, the word recognition system is misled by the prime presentation. Moreover, the very different priming effects obtained at different SOAs (see section 2.3.3) suggest that not only differences between different stages of word recognition are involved, but also demands that are specific to the task. When corpus-based distributional variables such as frequency of occurrence are available, manipulating these is preferable, as is done in the experiments reported below.

2.3 Effects of formal and semantic similarity

Whenever a word is heard, a large set of possible lexical candidates are activated in parallel (McQueen and Cutler, 2001). In everyday language processing, this activation is co-determined by sentential, textual and pragmatic context, but also by the formal and semantic overlap between the target word and other lexical candidates. The influence of this overlap can be assessed in word recognition studies.

There are two key ways of defining the sets of words which are activated as a consequence of formal overlap with the target: One set is the orthographic or phonological neighbourhood, which is traditionally defined as the words that can be generated by changing one letter or one phoneme in the target word without changing the order of letters or phonemes (Coltheart, Davelaar, Jonasson and Besner, 1977). Different neighbourhood effects are discussed in section 2.3.1. Another set of formally related words is the cohort. The cohort at a given phoneme of a target word consists of the competitors that overlap with the target from the onset up to and including the given phoneme. The cohort changes over time: Early in the word, many candidates are compatible with the input, but these candidates gradually disappear from the cohort as they become incompatible with what has been heard. The point in the auditory signal at which only the actual target word remains compatible with the input is known as the uniqueness point (UP), a key point in auditory word recognition, as shown in section 2.3.2.

As outlined in section 1.3, a main question in the field is whether morphological effects on word recognition can be reduced to non-additive, i.e. interacting effects of semantics and form, or whether morphology plays a role independently of semantic and formal similarity. Morphologically related words are generally semantically and formally similar, and the perception of morphological structure in complex words relies at least to some extent on formal and semantic overlap with its morphological relatives in the mental lexicon. A large body of priming studies attempts to disentangle the effects of form, semantics and morphology, as discussed in section 2.3.3.

2.3.1 Neighbourhoods

Lexical neighbourhoods consist of words in a language which are similar to a given target word according to some definition of formal similarity. The most common definition is known as Coltheart's N (Coltheart et al., 1977). On this definition, neighbourhoods are all words that are at a Hamming distance of 1 from the target word, i.e. words that can be constructed by changing one letter (for an orthographic neighbourhood measure) or one phoneme (for a phonological neighbourhood measure) in the target word without changing the position of the letters or phonemes. Coltheart's N is a type count, also known as neighbourhood size or density; some researchers also draw in the token count of the neighbours (e.g. Grainger, 1990; Magnuson et al., 2007), on the rationale that more frequent neighbours will be more highly activated than less frequent neighbours and therefore affect recognition more. A slightly broader way of defining neighbourhoods, which is based on the same logic, is applied by for instance Vitevich and Luce (1998) who define the phonological neighbourhood of a target word as the words that can be constructed by changing, inserting or deleting one phoneme in the target.

In visual naming, orthographic neighbourhood density has a consistently facilitatory effect (Grainger, 1990; Andrews, 1989, 1992), also over and above effects of spelling-to-sound consistency (Balota, Cortese, Sergent-Marshall, Spieler and Yap, 2004; Baayen, Feldman and Schreuder, 2006). It is, however, questionable whether this neighbourhood effect has to do with the recognition component of the naming task, or whether words with larger neighbourhoods are easier to produce because they have stronger spelling-to-sound mappings. Similarly, in auditory naming, strings with high neighbourhood density involve frequent phoneme sequences; this correlation results in faster naming of nonwords with large neighbourhoods (Vitevich and Luce, 1998, 1999). At a lexical level, i.e. in word rather than nonword naming, Vitevich and Luce (1998, 1999) found inhibitory neighbourhood effects; they argue that the inhibition was caused by the activation of many phonologically similar candidates which cancelled out the facilitatory effect of frequent phoneme combinations.

A similar duality is evident in visual lexical decision, which more exclusively taps word recognition: On one hand, a dense orthographic neighbourhood supports the lexicality of a word, with the result that rejecting nonwords may become more difficult when they have larger lexical neighbourhoods (Coltheart et al., 1977; Balota et al., 2004), and low-frequency real words may become easier to accept (Andrews, 1989; Balota et al., 2004). On the other hand, when lexical representations are relatively well-supported, as they are for words of higher frequency, neighbourhood density is inhibitory, reflecting the activation of competing lexical candidates which are formally similar but semantically unrelated to the target word (Andrews, 1989; Balota et al., 2004). However, in their reanalysis of the data of Balota et al., Baayen et al. (2006) found no independent effect of orthographic neighbourhoods once the consistency of spelling-to-sound mappings were controlled, and no interaction with frequency.

Consistency measures

Various spelling and pronunciation consistency measures were significant in the analysis of Baayen et al. (2006) while standard neighbourhood density were not. These consistency measures can also be understood as measures of similarity neighbourhoods. Databases of phonologically transcribed words are required to extract such consistency measures, to assess whether the pronunciation of a word is consistent with most other words that are spelled in the same way (spelling-to-sound consistency) and whether the spelling is consistent with that of most other words that are pronounced in the same way (sound-to-spelling consistency). Such databases are not available for Danish, hence these measures could not be obtained for the Danish visual stimuli in Experiment 1b (see chapter 5) where they would be most relevant. Moreover, the appropriate consistency measures are difficult to determine and probably less informative for polysyllabic words which constitute the majority of the items used in the experiment, than for the monosyllabic words analysed by Baayen et al. (2006). The consistency measures are nonetheless brought up here, for two reasons: Firstly, the consistency effects are evidence of the effect of lexical neighbourhoods on word recognition. Though the effect of raw neighbourhood density is perhaps questionable, as discussed above, these more refined measures do affect recognition time.

Secondly, the effects of these consistency measures in the analyses of the English Lexicon Project data by Balota et al. (2004) and Baayen et al. (2006) show the influence of phonology on visual word recognition (along with for instance Rubenstein, Lewis and Rubenstein, 1971; Stone, Vanhoy and Van Orden, 1997; Yates, Friend and Ploetz, 2008; for a review see Frost, 1998). In contrast, Ziegler, Petrova and Ferrand (2008) found no effect of sound-to-spelling consistency on visual word recognition. Conversely, there are also effects of orthography on auditory word recognition (see Ziegler, Muneaux and Grainger, 2003; Chéreau, Gaskell and Dumay, 2007; Taft, Castles, Davis, Lazendic and Nguyen-Hoan, 2008 and others). The issues of the influence of phonology on visual word recognition and orthography on auditory word recognition are not central to the main questions concerning morphological processing. They are, however, relevant in the experiments reported in chapters 5 to 7 below, since orthographic measures were investigated for the auditory Experiments 1a, 2, 3a and 3b, and phoneme-based measures in the visual Experiment 1b.

Embedded words

The standard measures of lexical neighbourhoods are most informative for shorter words (Baayen et al., 2006), because the neighbourhood density quickly drops off as words become longer and frequently has a value of 0 for words that are polysyllabic and thus relatively long. In the large dataset of Balota et al. (2004), there is a strong negative correlation between word length and neighbourhood density, such that longer words have smaller neighbourhoods. The words in the Balota et al. study are all monosyllabic with a mean length of 4.9 letters and still show this correlation; it follows that for the polysyllabic morphologically complex words in the three experiments reported below (which have mean lengths of 7.8, 7.2 and 7.8 letters), the neighbourhood density, when defined as Coltheart's N, is relatively uninformative. However, the fact that the standard measure of neighbourhood density is uninformative for a set of words does not mean that lexical neighbourhoods do not affect the processing of these words, but indicates that alternative measures may be required.

Such alternative measures have been developed by Bowers et al. (2005) and Baayen et al. (2007). Bowers et al. investigated the semantic activation of words that were substrings of a target. In a semantic categorisation task, Bowers and colleagues found that the presence of the string *hat* in the word *hatch* made it more difficult to determine that *hatch* does not belong to the category 'article of clothing', relative to matched controls where no substring of the target belonged to the category in question. This effect was found irrespectively of where in the target word the substring occurred, but only when the substring was of higher frequency than the target. Similarly, Bowers et al. found that words that were superstrings of the target were also semantically activated, such that the superstring *hatch* could interfere with the semantic categorisation of the target *hat*. Based on the results of Bowers et al., Baayen et al. (2007) extracted all words that were contiguously embedded in their target words or in which their target words were embedded with or without intervening letters, excepting morphologically related words. Baayen et al. calculated an entropy measure based on the embedded words and an entropy measure based on the embedding words.

In information theory, high entropy is associated with high uncertainty of the outcome of an event. The more uncertain an outcome is, the more bits of information are required to represent each possible outcome, and thus entropy is a measure of informational complexity. The entropy of an inflectional paradigm (Moscoso del Prado Martín et al., 2004b) or another set of words, such as the embedded forms, is higher if there are more words in the set, i.e. if the set has a higher type frequency, or if the words are more equiprobable, i.e. have similar token frequencies. Entropy measures can therefore be understood as token-weighted type counts. If there are more members in a set, each individual member is less probable, making more bits necessary to represent each member and increasing the entropy. To understand why more equiprobable members in a set make the entropy higher, one can simply compare the situation where all members are equally probable to the situation where one member is much more probable than the others; it is intuitively clear that the uncertainty is higher in the former situation.

Both the embedded and embedding entropies of Baayen et al. (2007) showed significant inhibitory effects on visual lexical decision time, indicating that the sub- and superstrings are semantically activated and that they interfere with the recognition of the target word because they are semantically unrelated to it. The embedded and embedding entropy measures were also calculated for the items in the visual lexical decision experiment reported in chapter 5, but no effects were found. These measures are relatively specific to visual word recognition, where the entire word is available at once, whereas they cannot be expected to affect auditory word recognition, at least not with the letter-based definition applied by Bowers et al. (2005) and Baayen et al. (2007). The results of Bowers et al. and Baayen et al. remain relevant because they show the inhibitory effect of non-morphological embeddings and the very fast activation of semantic information (see also Reimer, Lorsbach and Bleakney, 2008).

Conclusion

Even when broadly defined, the various neighbourhood effects discussed in this section are not of central relevance to the study of morphological processing. However, effects of the formal similarity between target words and their lexical neighbourhoods constitute an important background to the understanding of morphological similarity effects. Neighbourhood effects, which involve formal similarity only, tend to be inhibitory on the lexical level, while morphological similarity, which involves both formal and semantic similarity, tends to have a facilitatory effect. The activation of lexical neighbours is caused by formal similarity to the target, but it is the semantic dissimilarity between the neighbours and the target that result in inhibitory effects on lexical processing. In the experiments reported below, neighbourhood density is included as a control predictor. The uniqueness point measures discussed next are of more central importance here, because they take into account the temporal unfolding of the signal in the auditory modality. More specifically, one of the main objectives of the present study is to investigate the composition of the competition cohort with respect to the distinction between morphologically related and unrelated words.

2.3.2 Competition cohorts and uniqueness points

Another way of indexing the lexical similarity neighbourhood of a word is through the cohort-concept, which is mainly relevant for the auditory modality where the signal unfolds gradually over time. At a given point in a word, the cohort is defined as the group of words that are compatible with the input up to that point. In other words, the cohort is a neighbourhood measure which is based on the similarity of onsets in contrast to the more global similarity measured by standard neighbourhood density (Magnuson et al., 2007). The cohort changes over time: Early in the auditory signal, many candidates are compatible with the input and the cohort of activated lexical candidates is large. As more of the word is heard, the cohort becomes gradually smaller until the uniqueness point (UP) is reached. The UP is the phoneme at which the target word becomes uniquely distinguishable from all other words in the language, i.e. where the cohort is reduced to one. A Danish word like *accept* ([ɑgˈsɛbt][1], 'acceptance') activates a large and diverse word-initial cohort including all words starting with the vowel [ɑ], e.g. *araber* ('Arab') and *aksiom* ('axiom'). As more of the word is heard, more and more of the competitors become incompatible with the input and disappear from the cohort; for instance, at the [g] of [ɑgˈsɛbt], *araber* is no longer compatible and drops from the cohort. At [ɛ] (the 'e' of the spelling), only the stem *accept* and suffixed and compound continuations of it remain in the cohort, and [ɛ] is identified as the UP. The UP is theoretically the earliest possible point at which word recognition can be achieved; the later in a word the UP occurs, the longer it takes to recognise that word (e.g Marslen-Wilson, 1984, 1990). The fact that onsets are available first in auditory processing makes the onset-alignment of the cohort competitors more important in auditory than in visual processing, where the whole word is in principle available immediately (though longer words may require more than one fixation); the following focuses on the auditory modality.

In its original form (Marslen-Wilson, 1984), the UP is a-morphological: All compounds and suffixed forms are excluded from the cohorts on which UP-calculations are based. If suffixed words and compounds were included in the competition cohort, the majority of words in languages like English and Danish would have UP after word offset, rendering the construct virtually meaningless. Even with this definition, the frequency-weighted probability of English words becoming unique before word offset is .39 (Luce, 1986), while for German, the proportion of words that become unique before word offset is higher (Bölte and Uhe, 2004). However, there are two related problems with this exclusion of suffixed and compound forms: Firstly, it means that the UP disregards the morphological structure of complex words, which make up a substantial part of the words in many languages, and becomes less useful as a diagnostic of morphological processing. Secondly, the UP provides no way of measuring or understanding the competition that may be going on between different suffixed and compound continuations of the same stem. Various ways that have been suggested as remedies for the first problem, particularly with reference to prefixed words, are presented in this section, before two new UPs are introduced which take into account the morphological structure of all types of complex words.

[1]The transcriptions are broad phonetic transcriptions based on the Dania-transcriptions in Becker-Christensen, Appel, Katlev, Rasmussen and Troelsgaard (2005), though the vowels are adapted to IPA-conventions based on Grønnum (2005).

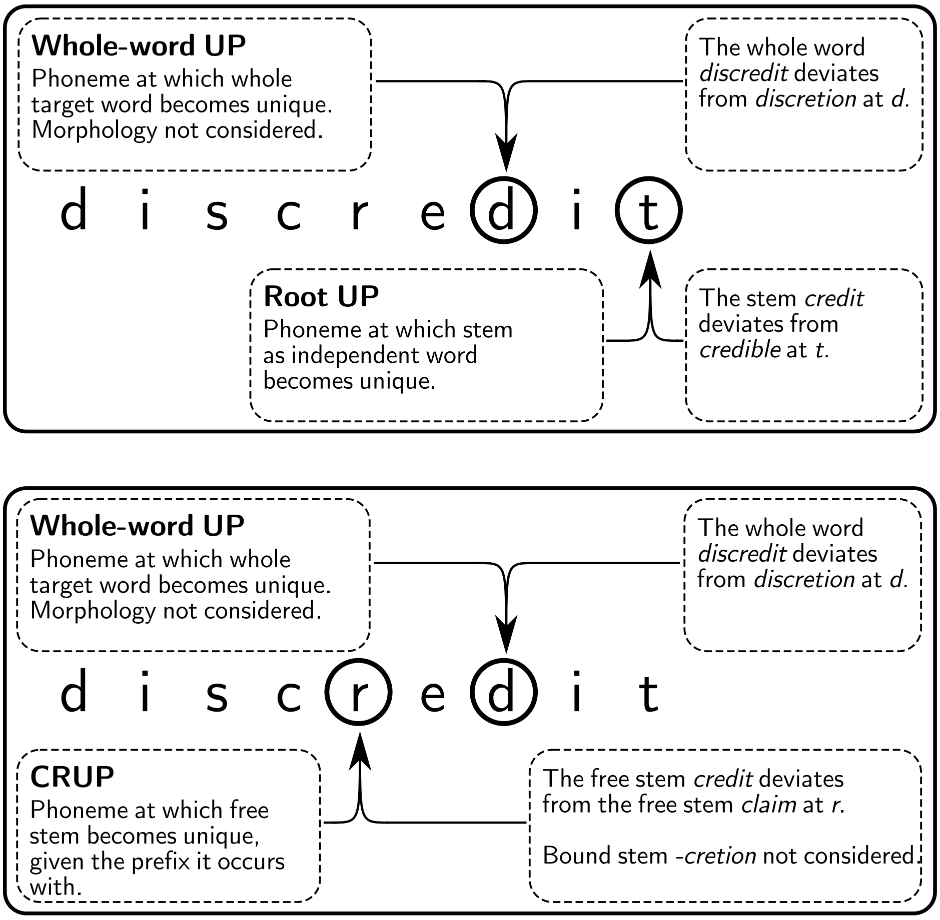

Figure 2.1: Definitions and examples of whole-word and stem UP in the top panel, the effects of which were tested by Tyler et al. (1988), Schriefers et al. (1991) and Wurm (1997), and whole-word UP and CRUP as established by Wurm (1997) in the bottom panel. The example and the UPs are taken from Wurm (1997: 454).

Stem UP and CRUP for prefixed words

For prefixed words, early studies (Tyler et al., 1988; Schriefers et al., 1991; Wurm, 1997) compared the effects of the UP of the whole word with the effects of the UP of the stem considered as an independent word, i.e. for a prefixed word like *dis-credit*, comparing the effect of the UP of the whole word *discredit* and the UP of the stem *credit* as an independent word. The whole-word and stem UPs are shown for the complex word *discredit* in the top panel of fig. 2.1. Using several different tasks, all three studies found reliable effects of the whole-word but not of the stem UP. Particularly the experiments of Tyler et al. and Schriefers et al. were designed to test the predictions of the prefix-stripping model (Taft and Forster, 1975), according to which prefixed words are recognised on the basis of the stem. The absence of a stem UP effect is problematic to this model. However, in terms of the role of morphological structure considered more broadly, the UP of the stem as an independent word is arguably an inappropriate measure of lexical competition in prefixed words, because the prefix has already been encountered before the stem is processed. Therefore, Wurm (1997) defined a context-sensitive stem UP: the UP of the root (stem) given the prefix it occurs with. In *discredit*, the conditional root uniqueness point (CRUP) is the point where the stem *credit* becomes uniquely distinguishable from all other free stems that can combine with the prefix *dis-*, i.e. including free stems like *claim*, but excluding bound stems like *-cretion*. The distinction between whole-word UP and CRUP is shown in the bottom panel of fig. 2.1. Wurm (1997), Wurm and Ross (2001) and Wurm, Ernestus, Schreuder and Baayen (2006) found that words where the CRUP preceded the standard, whole-word-based UP (so-called CRUP-words, e.g. *discredit*) were processed significantly faster than those where they coincided. The CRUP takes into account the morphological structure of the target word, but does not assume that the stem is processed independently of the prefix that it occurs with. Instead, the CRUP allows for the possibility that the processing of the stem is affected by the fact that the prefix has been processed; the stem-processing is conditional on the prefix having been heard. The advantage for CRUP-words, where the CRUP precedes the UP, is substantial, but the CRUP only precedes the UP in a minority of prefixed words: approximately one eighth of prefixed words with free stems in English are CRUP-words (Wurm and Ross, 2001) and a smaller proportion in Dutch (Wurm et al., 2006).

Continuations

Another cohort-based measure considers the morphologically related continuations that are compatible with the target at word offset, i.e. the cohort at the final segment. The continuation forms of a given word encompass both inflected, derived and compound forms which are continuations of that word; for instance, the set of continuation forms for the Danish word *accept* ('acceptance') extracted from a Danish corpus (see chapter 3) included inflectional forms like *accept-en* ('the acceptance', literally 'accept-the'), derivational forms like *accept-abel* ('accept-able') and compounds like *accept-tale* ('acceptance speech', literally 'accept speech'). Such continuation forms are excluded from the traditional UP-calculations. The entropy was calculated across the continuation forms for prefixed words by Wurm et al. (2006) and for simple words by Kemps, Wurm, Ernestus, Schreuder and Baayen (2005a). This measure is termed late entropy by Wurm et al. and cohort entropy by Kemps et al. Since the cohort and the corresponding entropy can be established for different points in the word, and late entropy is a rather vague term, the term continuation entropy is used here instead. The continuation entropy can be understood as a token-weighted type-count of the words in the cohort at the final segment of the word.

Wurm et al. (2006) observed a facilitatory effect of this measure: The higher the continuation entropy (the more continuation forms or the more equiprobable continuations), the faster the reaction time to the prefixed items presented. The existence of many continuations supports the lexicality of the target word, as well as the semantic interpretation of it, and makes recognition faster. Wurm et al. frame the discussion of the continuation effect within the UP research programme. However, the continuation effect can also be understood as an effect of morphological family: In contrast to the cohorts at different points preceding the UP, the cohort at the final segment generally only contains words that are morphologically related. These morphologically related continuation forms include derivations and compounds, which are also included in the morphological family, as well as inflectional forms which are not (see section 2.5). The continuations differ from the standard family measures by only including words which are onset-aligned with the target, a factor which may be particularly salient in the auditory modality and plays a role in the interpretation of the experiments reported below, see discussion in section 8.2.

For the prefixed word in the study of Wurm et al. (2006), the cohort at the final segment included only words which were morphologically related. For simple words, the morphologically related words are also dominant in the cohort at the final phoneme, but it may also contain non-related words (Kemps et al., 2005a). This is the likely reason why the direction of the continuation entropy effect differs between Wurm et al. and Kemps et al. Kemps et al. observed an inhibitory effect of continuation entropy, such that higher continuation entropy (i.e. more continuations or more equiprobable continuations) resulted in longer RTs. The continuation entropy for the monomorphemic words of Kemps et al. reflects the uncertainty about which particular continuation is the target. This is also what the measure reflects for the prefixed words in the study of Wurm et al., but the uncertainty as to which specific morphological continuation is the target seems to be less important for a correct lexical decision to be made than the fact that many continuations support the lexicality of the target, resulting in faster RTs for words with higher continuation entropy.

In sum, three existing measures of lexical competition between cohort members have been found to predict auditory word recognition latencies for morphologically complex words: the UP, the continuations and the CRUP. However, the three measures only provide an incomplete picture of the role of morphological structure. The traditional UP indexes competition between words, but it was constructed as an a-morphological measure and excludes morphologically related forms from the cohort. The continuations include some of these morphologically related forms, but only those that are compatible with the target at word offset. The CRUP takes into account the morphological structure of certain prefixed words, by indexing competition between free stems given the prefix that has occurred. The relevance of the CRUP is limited to a minority of prefixed words, but the central insight of the CRUP-construct — that processing of the second constituent is conditioned by the processing of the first constituent — may be more broadly applicable. The two new UPs introduced next can be used to investigate whether such conditional processing may occur also in other types of complex words.

Two new UPs

For suffixed words, the existing measures do not account for competition between different suffixed and compound forms of the stem: the traditional UP is a measure of the competition between the stem and unrelated words, and the continuation entropy is a measure of the continuations of the whole suffixed word, but neither measure considers alternative suffixed and compound forms of the stem. For instance, in a suffixed derived word like *accept-ere* ([ɑgsɛbˈteʔʌ], 'to accept', the simple noun 'accept' plus a verbalising suffix), competition from morphologically unrelated words is resolved at the traditional UP, the [ɛ], but a number of suffixed and compound forms are still compatible with the input, including both the target and competing words like *accept-abel* ([ɑgsɛbˈtaʔbəl], 'accept-able'). If suffixed and compound competitors are not considered, the suffixed word *accept-ere* has the same UP as its stem *accept* as an independent word. This illustrates the aspects of the competition process that the traditional UP was not designed to account for.

In order to account for possible competition between different continuations of the stem, a new UP-measure is introduced here for complex words: the Complex UP (CUP). This is most easily illustrated for suffixed words: The CUP of a suffixed word is the point at which it becomes uniquely distinguishable from alternative suffixed and compound forms of the target stem; for the example *accept-ere* ([ɑgsɛbˈteʔʌ]), the CUP occurs at [e] where *accept-ere* deviates from other continuations of the stem *accept*, e.g. from forms like *accept-abel* ([ɑgsɛbˈtaʔbəl], 'acceptable'). In this way, the CUP measures the competition between words that are morphologically related to the suffixed targets through their stems, while the traditional UP measures competition from unrelated forms. The two different UPs are illustrated for the suffixed word *accept-ere* in the top panel of fig. 2.2 (where the traditional UP is termed UP1, see further below). Both the UP and the CUP exclude words that are continuations of the whole word; these only drop out of the cohort at word offset and are measured as the set of continuation forms. Together, the three measures can provide a precise picture of the lexical

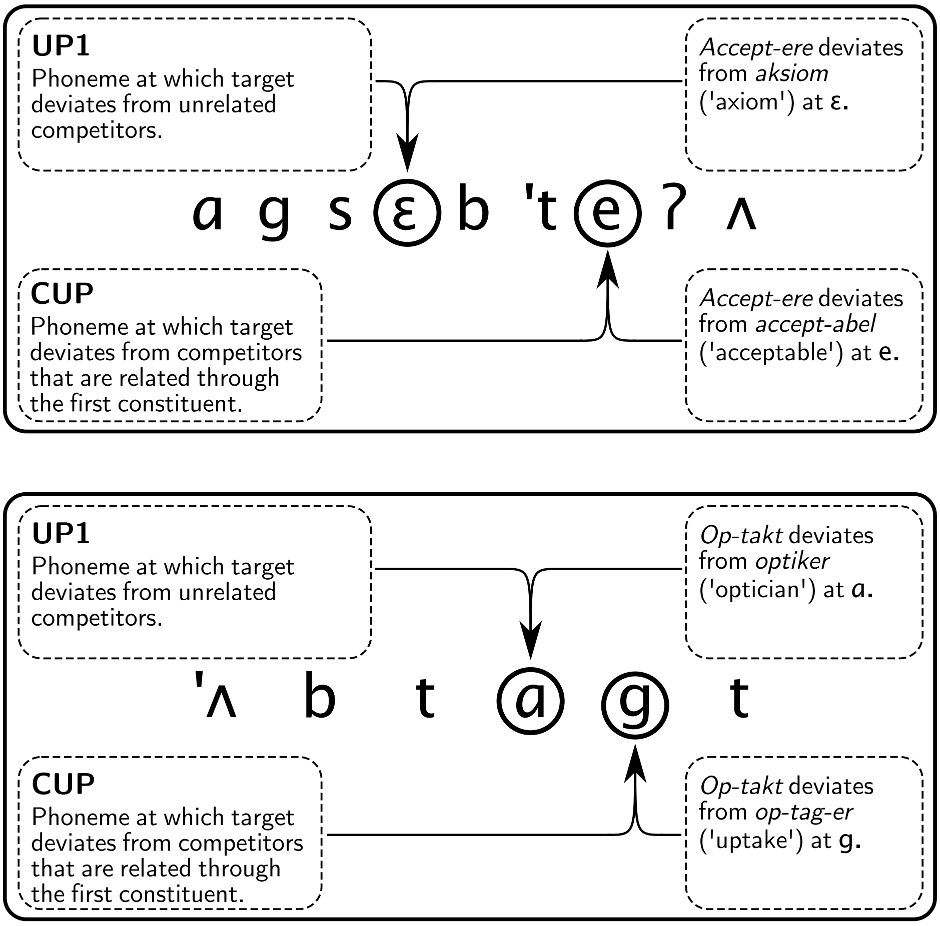

Figure 2.2: Definitions and examples of UP1 and CUP for the suffixed word *accept-ere* ('to accept') in the top panel and the particle prefixed word *op-takt* ('upbeat') in the bottom panel. The translations in the figure are short versions, full translations are given in the text.

competition process for suffixed words. The effect of the new CUP-measure is experimentally investigated in the Experiment 2 (see chapter 6), together with the traditional UP and the continuation forms.

Since both the traditional UP and the CUP turned out to be significant predictors of recognition latency for the suffixed items in Experiment 2, it was desirable to include both UPs in the analyses of Experiments 1 and 3, applying them also for compounds and prefixed words. For the compounds in Experiment 3, the logic is straightforward, although I know of no existing studies of UP-effects for compounds: The traditional UP is parallel to that of the suffixed words, excluding continuation forms of the first constituent from the competition cohort. The UP is defined as the point where the target compound deviates from other words in the language, except continuation forms of the first constituent. For example, the UP of the compound *fod-bold* ([ˈfoðbʌld], ’foot-ball’) is the [ð] where the first constituent *fod* (’foot’) deviates from other words in the language, including words like *foto* ([ˈfoto], ’photo’) but excepting continuations of *fod* itself. The CUP of the compound is then the point where the competition between these morphologically related continuation forms of the first constituent is resolved. For *fod-bold*, the CUP occurs when all continuations of *fod* which are not also continuations of the whole word *fod-bold* have dropped from the cohort. The CUP of *fod-bold* occurs at [ʌ] where *fod-bold* deviates from related words like *fod-bad* ([ˈfoðbað], ’foot-bath’).

For prefixed words, the picture is more complicated: Whereas the traditional UP for suffixed words clearly indexes competition from non-related words (e.g. the example *accept-ere* in fig. 2.2), this is not the case for prefixed words. Instead, the competition that is resolved at the traditional UP of prefixed words is in most cases competition between the prefixed target and words that share the same prefix. Words that partly overlap with the target, but do not contain the prefix in question, tend to fall out of the competition earlier than words that do contain the given prefix. For instance, the traditional UP of the particle prefixed word *op-takt* ([ˈʌbtɑgt], ’upbeat’) is at the phoneme [g] where a word like *op-tag-er* ([ˈʌbtɑʔ], one possible pronunciation of the present tense of the verb ’to take up’ or ’to record’), which carries the same prefix as the target, becomes incompatible with the input. However, the last unrelated words, e.g. *optiker* ([ˈʌbtigʌ], ’optician’), disappear from the competition cohort earlier, namely at the [ɑ]. There are cases where the last UP-competitor for a prefixed word is an unrelated word, but it is the exception rather than the rule; this is the case for none of the 100 prefixed words in Experiment 3 and for only two of the 75 prefixed words in Experiment 1.

Therefore, the traditional UP for prefixed words is not an appropriate parallel to the UP for the suffixed words, if the UP is to be understood as a measure of competition between unrelated words. What we need as a parallel to the UP for the suffixed words is the point in prefixed words where all non-related competitors cease to be compatible with the input. For *op-takt*, this would be the [ɑ] where *op-takt* deviates from all non-related words, e.g. *optiker*, while related words like *op-tag-er* are still compatible. The point where competition from non-related words is resolved is termed UP1 in the analyses below; this corresponds to the traditional UP for the suffixed words, but not for the prefixed. The term UP1 is used in order to emphasise the fact that it differs from the traditional a-morphological UP for the prefixed words, and the fact that it relates specifically to the uniqueness of the first constituent.

What remains is the definition of the Complex UP or CUP for the prefixed words. For the suffixed words, the CUP is defined as the point where the target deviates from the different suffixed and compounding forms of the same stem, i.e. for *accept-ere*, the [e] (the second 'e' of the spelling) where *accept-ere* deviates from *accept-abel*. For the prefixed words, the CUP is the point where the prefixed target deviates from the group of words that share the same particle prefix, i.e. for *op-takt*, the [g] (the 'k' of the spelling) where *op-takt* deviates from *op-tag-er*. UP1 and CUP for prefixed words are illustrated for the example word *op-takt* in the bottom panel of fig. 2.2; *op-takt* is a particle prefixed word, but the same logic applies for standard prefixed words. In all cases, continuation forms of the whole word itself are excluded from the UP-calculations. The logic of the CUP is the same for all types of complex words, they differ only in whether they are related through the stem or through the prefix.

In practice, the CUP corresponds to the traditional, a-morphological UP for prefixed words. For suffixed words, in contrast, it is UP1 that corresponds to the traditional UP. This means that in the analyses of Experiments 1 and 3, the traditional UP is not a single entity, because it corresponds to UP1 for the suffixed and compound words and to CUP for the prefixed words. An overview of the different UPs, including stem UP and CRUP, is provided in table 2.1 for suffixed, prefixed and prefixed CRUP-words. There are two advantages of operating with the two new UPs instead of the single traditional one: Firstly, the understanding of the lexical competition process for complex words is enhanced by dividing lexical competition into the competition from related and the competition from non-related words. The morphological structure of the target words is taken into account as are the morphologically complex competitors which are excluded from the calculation of the traditional UP (though still excepting continuation forms of the whole complex words). Secondly, this account in terms of non-related vs. related competition provides a uniform account of competition and conditional processing across different types of morphologically complex words, being demonstrated in the experiments reported below for prefixed, suffixed and compound words.

The distinction between UP1 and CUP builds on the insights implemented in Wurm's distinction between UP and CRUP. In both cases, the competition usually measured by one UP is split between two different UPs for the prefixed words: In the present experiments, UP1 indexes the competition between non-related forms and CUP the competition between related forms. Wurm operates with the traditional UP, which measures competition for prefixed words irrespective of whether the competitors are prefixed or not, while the CRUP specifically measures the competition between stems given that the prefix has occurred, i.e. competition between a subset of morphologically

Table 2.1: Overview over different UPs, for the suffixed word *accept-ere* ([ɑgsɛb'teʔʌ], noun 'accept' plus verbalising suffix), the prefixed word *op-takt* (['ʌbtɑgt], 'upbeat') and the prefixed CRUP-word *van-skæbne* (['van-skɛbnə], 'misfate'). The letters corresponding to the UP-phonemes are capital letters. For *op-takt*, the CRUP coincides with the whole-word UP, so *op-takt* is not defined as a CRUP-word, while in *van-skæbne*, CRUP precedes whole-word UP, making *van-skæbne* a CRUP-word. The continuations are based on spelling rather than pronunciation.

	Suffixed	Prefixed	Prefixed CRUP-word
Whole-word UP	a c c **E** p t - e r e	o p - t a **K** t	v a n - s k æ **B** n e
Stem UP	a c c **E** p t - e r e	o p - t a k **T**	v a n - s k æ **B** n e
CRUP	N.A.	o p - t a **K** t	v a n - s k **Æ** b n e
UP1	a c c **E** p t - e r e	o p - t **A** k t	v a n - s k æ **B** n e
Complex UP	a c c e p t - **E** r e	o p - t a **K** t	v a n - s k **Æ** b n e
Continuations	*accepter-ende*	*op-takt-en*	*van-skæbne-n*
	accepter-et	*op-takt-s-fase*	*van-skæbne-r*
	etc.	etc.	etc.

related words. Wurm's basic idea of conditioning the competition between morphologically related words on the presence of a prefix is applied to other types of complex words in the form of the Complex UP.

Wurm excludes bound stems from the calculation of the CRUP, which means that the CRUP can sometimes occur before the UP, though this is relatively rare. In contrast, the CUP-cohort does include prefixed words with bound stems, because these words are morphologically related to the target and should activate the semantics of the prefix, and because other bound constituents, namely the suffixes, must be considered in order for the CUP to make sense for suffixed words. The advantage for CRUP-words indicates that the free stems are stronger competitors, because the CRUP-advantage is based on earlier uniqueness from competitors with free stems than from competitors with bound stems. The effects of CUP indicate that bound constituents, primarily suffixes, also enter into the competition.

Differences between CRUP and UP arise when uniqueness from competitors with free stems precedes uniqueness from competitors with bounds stems (or, in principle, unrelated competitors), but this happens only for a minority of prefixed words. In contrast, the distinction between UP1 and CUP relies on differences between unrelated and related competitors and applies to all types of morphologically complex words. Both CRUP and CUP demonstrate the role of conditional processing in the recognition of morphologically complex words.

Problems with the UP-construct

Both the traditional UP and the new UPs introduced here are based on phonemes (though the new UPs take some broad phonetic differences into account, see details of the UP-calculations in section 3.4.4). This means that the UP-measures are relatively coarse, in view of the substantial evidence of listeners' sensitivity to subphonemic detail in processing (Marslen-Wilson and Warren, 1994; Davis, Marslen-Wilson and Gaskell, 2002; Salverda, Dahan and McQueen, 2003; Kemps et al., 2005a; Kemps, Ernestus, Schreuder and Baayen, 2005b; Salverda, Dahan, Tanenhaus, Crosswhite, Masharov and McDonough, 2007). For instance, Davis et al. found that subphonemic acoustic cues helped listeners distinguish between the same phoneme sequence functioning as an independent word and as the onset of a longer word. Kemps et al. (2005b) demonstrated that this sensitivity extends to the difference between noun stems occurring as free words vs. the same stems occurring as the stems of plural nouns. This sensitivity to subphonemic detail implies that the phoneme-based cohorts, from which the UPs are derived, are not to be understood as absolute measures of the lexical candidates activated due to onset overlap, rather the phoneme-based competition cohorts and the UP measures are approximations of the competition at various points in the auditory signal. In reality, subphonemic differences are likely to make some competitors weaker candidates than others, although the phoneme-based calculations allow no such distinctions.

At the same time, fluent restorations of mispronunciations in the shadowing task (where listeners repeat a text they hear and sometimes restore mispronunciations, Marslen-Wilson and Welsh, 1978) indicate that processing remains possible in spite of phonemic mispronunciations, although the studies mentioned above show that listeners are sensitive to a wealth of subphonemic detail. This is in a sense the opposite problem to that of the phoneme-based measures being too coarse: the phoneme-based cohorts on which the UPs are based are not inclusive enough to allow mispronounced words into the cohort (Marslen-Wilson, 1987). This is an important concern in the formulation of a cohort-based model of speech recognition, but is not a major concern in the study of morphological processing.

In spite of both these problems, the phoneme-based UPs show substantial effects on auditory word recognition and remain good indices of the lexical competition process.

Another approach to cohorts is that of Magnuson et al. (2007) who avoid determining the UP, but simply count the number of lexical candidates that overlap for the first two phonemes and compute a frequency-weighted cohort density measure that has significant effects in the visual world paradigm. The possibility of frequency-weighting the cohort members is useful, but the simple counting of initially activated cohort members does not take into account the changes to the cohort over time.

Radeau, Morais, Mousty and Bertelson (2000) found that UP-effects in gender decision and shadowing were restricted to slower speech rates. This suggests that UP-effects are less pervasive in natural language use than the large laboratory effects would suggest. Possibly, this is due to faster speech rates being characterised by more acoustic reduction. However, the wealth of evidence of UP-effects for laboratory speech that contrasts with the findings of Radeau et al. indicates that the UP-construct is viable, though arguably less strong in everyday language use than in psycholinguistic experiments.

Summary

This section has presented several existing UP-measures. The standard UP has significant effects on recognition time for prefixed words, but it is a-morphological and does not take the morphological structure into account. The UP of the stem as an independent word has been suggested as one way of taking the morphological structure into account, but this stem UP has no significant effects on various recognition measures. The CRUP of Wurm (1997) demonstrates that the UP-construct can account for the role of the morphological structure of the target and its cohort competitors, for a subset of prefixed words. In order to extend the applicability of the UP-construct to other types of complex words, a new distinction between two UPs is investigated in the experiments reported below: UP1 indexes competition from words that are not related to the complex target. The Complex UP or CUP measures the duration of competition from words that are morphologically related to the complex words through sharing the same first constituent, be this the stem of a suffixed word, the prefix of a prefixed word, or the first constituent of a compound. In all cases, the processing of the second constituent indexed by the CUP is conditional on the first constituent having been processed, as is the case for the CRUP. These new UP-measures receive strong experimental support from the three experiments reported below, as discussed in chapters 5 to 7 and in section 8.4.

2.3.3 Transparency and relatedness in the priming literature

Morphologically related words are generally also related in meaning and form, though to varying extents. This fact motivates the question whether there are effects of morphological relatedness in priming over and above any effects of semantic and formal relatedness, which in turn has been identified with the question of whether morphological effects in lexical processing more generally are real or epiphenomenal (Seidenberg and Gonnerman, 2000; Gonnerman et al., 2007). Both questions have been extensively researched and a review of this research is presented in this section; it is kept relatively general, because the effects from priming do not necessarily generalise to non-primed word recognition (cf. section 2.2.1). The studies of the effects of semantic relatedness and transparency (and thus also the review in the present section) tend to focus on derived words because, in contrast to inflected words, they vary considerably in semantic transparency, while they are more regular than compounds.

There are two connected ways of approaching the issue of how morphology, semantics and orthography/phonology are related in the mental lexicon: Firstly, morphological priming effects can be compared to purely formal or purely semantic priming effects; secondly, it can be investigated whether morphological priming effects differ as a function of the formal or semantic transparency of the morphologically related prime-target pairs.

Morphological, formal and semantic priming

The comparison of prime-target pairs that are morphologically, formally or semantically related is arguably most informative if the time-course of the different effects is considered by investigating priming effects at different stimulus onset asynchronies (SOAs): Prime-target pairs which are transparently morphologically related tend to show facilitatory effects at all SOAs: in delayed priming with 7 to 13 items intervening between prime and target (Feldman, 2000; Rueckl and Aicher, 2008), in immediate priming where the prime is presented long enough to be consciously processed (Marslen-Wilson et al., 1994; Marslen-Wilson and Zhou, 1999; Rastle, Davis, Marslen-Wilson and Tyler, 2000; Feldman, 2000; Longtin, Segui and Hallé, 2003; Gonnerman et al., 2007) and in subliminal (typically masked) priming tasks where the prime is not consciously processed (Rastle et al., 2000; Longtin et al., 2003; Diependaele, Sandra and Grainger, 2005; Marslen-Wilson, Bozic and Randall, 2008).

Facilitation from a purely semantically related prime (e.g. in pairs such as *idea–notion*) is typically found in immediate priming (Bentin and Feldman, 1990; Marslen-Wilson et al., 1994; Feldman, 2000; Rastle et al., 2000; Meunier and Segui, 2002; Gonnerman et al., 2007), but not systematically when the prime is masked (Frost, Forster and Deutsch, 1997; Rastle et al., 2000), and not in delayed priming (Bentin and Feldman, 1990; Feldman, 2000).

Formal priming effects — i.e. priming between prime-target pairs that are orthographically and/or phonologically related, such as *tinsel–tin* — are even more elusive, with no effect in delayed priming (Drews and Zwitserlood, 1995; Feldman, 2000), and no effect (Marslen-Wilson and Zhou, 1999; Longtin et al., 2003; Gonnerman et al., 2007; Marslen-Wilson et al., 2008) or inhibition (Grainger, 1990; Drews and Zwitserlood, 1995; Feldman, 2000; Rastle et al., 2000; Longtin et al., 2003) in immediate and masked priming. Although they are not entirely systematic, the inhibitory effects of formal overlap are related to the inhibitory effects of neighbourhood and cohort density: the semantics of the formally overlapping prime is presumably activated and competes with the semantics of the target.

The comparison of morphological, semantic and formal priming effects cannot conclusively answer the core question of whether morphological relatedness is epiphenomenal of semantic and formal relatedness: The fact that the morphological priming effects are generally larger than the added effects of formal and semantic priming (Marslen-Wilson et al., 1994) and the fact that they differ in their time courses suggest that the morphological relatedness is qualitatively different from semantic or formal relatedness. However, this does not rule out the possibility that morphological effects are caused by the non-additive or interacting effects of formal and semantic effects (Seidenberg and Gonnerman, 2000), but such non-additivity is impossible to demonstrate experimentally. Moreover, semantic and formal relatedness cannot be manipulated in a set of prime-target pairs without these pairs also being morphologically related, with the possible exception of phonaesthemes, e.g. the sets of words beginning with *gl-* in English or Danish that are similar in meaning. Priming effects between phonaesthemes are documented by Bergen (2004). On a morpheme-based view of morphology, such effects are problematic, but in word- and paradigm-based morphology, the shared form-segments of phonaesthemes are not that different from bound morphemes, except that a complete morphological parse of the words with phonaesthemes is meaningless.

Transparency

The second way of approaching the role of semantic and formal overlap on morphological effects is to investigate the semantic or formal transparency of morphologically complex words. Complex words are defined as semantically transparent if the meaning of the complex word is closely related to the meaning of the stem (e.g. *punishment–punish*), while they are semantically opaque if the meaning of the complex word is not related to the meaning of the stem (e.g. *casualty–casual*). Examples of more or less semantically transparent words in Danish are provided in chapter 6 (and in Appendix B), since the experiment reported in that chapter includes semantic transparency as a co-variate; the English examples are drawn from the seminal study of Marslen-Wilson et al. (1994).

The definition of formal transparency is mainly based on pronunciation (though spelling often varies with the pronunciation): words are considered formally transparent if the stem is pronounced the same way (at least on a phoneme-level) inside the complex word as it is when it occurs as an independent word. Derived words with low formal transparency are defined as those in which pronunciation differs between the derived word in question and its stem as an independent word. For example, *vanity* has low formal transparency because the pronunciation of the vowel of the stem differs from the free form of that stem (*vain*); *elusive* is defined as formally non-transparent because it differs in the stem-final consonant from the free form *elude*. Semantic transparency is usually assessed in rating studies where a group of participants rate the degree of relatedness between the meaning of the derived words of interest and their stems. The semantic transparency of a complex word is thus defined as the degree of semantic relatedness between the complex word and its stem.

In contrast to the comparison of morphological, semantic and formal priming effects manipulated individually, the transparency approach may be a more productive way of getting at the core questions: If the degrees of semantic or formal transparency do not affect morphological priming effects, it is fairly strong evidence that morphological effects are not epiphenomenal, whereas the opposite pattern — effects of semantic or formal transparency on morphological priming — does not provide conclusive evidence that morphology is an epiphenomenon, though it is congruent with that hypothesis.

Formal transparency generally has no impact on morphological priming effects in immediate (Marslen-Wilson et al., 1994; Marslen-Wilson and Zhou, 1999; the suffixed items of Meunier and Segui, 2002) or delayed priming (Fowler, Napps and Feldman, 1985): the morphological priming effects were equal for prime-target pairs that varied (e.g. *elusive–elude* and *vanity–vain*) and pairs that did not vary. In a cross-modal immediate priming task similar to that of Marslen-Wilson et al. (1994), Gonnerman et al. (2007) found priming effects for both formally transparent and differing degrees of formally opaque prime-target pairs, but they also observed some decrease in priming from the least opaque (pairs where only a consonant changed) to the most opaque pairs (where both the vowel and a consonant of the stem changed), indicating that the formal transparency may affect priming effects for semantically very transparent prime-target pairs. Regular and irregular past tense forms also vary in formal transparency, so different priming effects for regular prime-target pairs than for irregular pairs can be explained as effects of formal transparency (see section 2.6 below).

The effects of semantic transparency on morphological priming effects vary systematically with SOA between prime and target: Using immediate cross-modal priming and conscious prime presentation, Marslen-Wilson et al. (1994) reported clear effects of semantic transparency on morphological priming: opaque derived words did not prime their stems and were not primed by their stems, while transparent derived words produced clear facilitatory priming effects in both directions. Marslen-Wilson et al. measured semantic transparency in a rating test; however, for the factorial design and analyses, Marslen-Wilson et al. dichotomised the graded ratings into high vs. low semantic transparency. The effect of high vs. low transparency was then interpreted in terms of transparent words being morphologically represented in the mental lexicon, while opaque words are represented as whole words with no morphological structure. In addition to the statistical problems associated with dichotomising of continuous variables (see section 4.3), the dichotomisation results in a misleading interpretation, as pointed out by

Wurm (1997): If the graded variable semantic transparency has a graded effect on word recognition, the variable cannot be mapped onto a binary distinction between morphemic and whole-word representations in the mental lexicon, but the use of a binary transparency variable means that such gradedness cannot be detected. A number of other studies report similar results, but with similar binary distinctions between semantically transparent and opaque words which make them vulnerable to the same criticism; these include Longtin et al. (2003), Marslen-Wilson and Zhou (1999), Rastle et al. (2000) and Feldman, Soltano, Pastizzo and Francis (2004).

Gonnerman et al. (2007) also operate with conditions defined according to semantic transparency, as is necessary for the purposes of analysis of variance, but they choose prime-target pairs in three conditions — low, moderate and high transparency — which allows for a more detailed picture. Gonnerman et al. observed no significant priming effects for the low transparency condition, and significant facilitation in both the moderate and the high conditions. The facilitation in the moderate transparency condition was numerically smaller than that in the high transparency condition, but this difference was only marginally significant. More interestingly, Gonnerman et al.'s post hoc regression analysis showed that the graded semantic transparency ratings constituted a significant predictor of the amount of priming; unfortunately, this regression analysis is only reported as a post-hoc analysis and hardly any detail is provided.

In studies that use subliminal (typically masked) priming, morphological priming effects are generally found irrespective of the semantic transparency of the morphologically complex words (Rastle et al., 2000; Longtin et al., 2003; Feldman et al., 2004; Marslen-Wilson et al., 2008; Kazanina, Dukova-Zheleva, Geber, Kharlamov and Tonciulescu, 2008). The priming from semantically transparent prime-target pairs persists also at longer SOAs, while the effects for opaque prime-target pairs disappear as the prime becomes consciously available to the participants. For instance, the morphological priming effects produced by opaque prime-target pairs in the study of Rastle et al. at an SOA of 43 ms disappeared at an SOA of 72 ms, while 72 ms did not seem to be enough for conscious prime processing, and thus for effects of semantic transparency on morphological priming, in the study of Marslen-Wilson et al. (2008). Masked priming is only an option with visual primes and usually targets are also visual. Diependaele et al. (2005) use auditory targets in their cross-modal masked priming task; their results are broadly similar to those of the visual-visual priming tasks.

Longtin and Meunier (2005) and Meunier and Longtin (2007) used an interesting novel approach to the connection between morphology, semantics and form: Instead of comparing effects of the different types of relatedness or effects of degree of transparency, Longtin and Meunier investigated the priming effects of different types of pseudowords on “related” word targets. Longtin and Meunier (2005) employed masked priming and observed priming for both pseudowords that were uninterpretable (e.g. *sport-ation*) and interpretable (e.g. *quick-ify*), while the cross-modal immediate priming task (where the visual target was presented at the offset of the auditory prime) of Meunier and Longtin (2007) showed that interpretability is necessary for priming effects to arise in immediate priming. These results pose a challenge to the Supralexical Model of Giraudo and Grainger (2001) which operates with supralexical morphemic representations that can only be activated by the activation of word representations.

The difference between brief and longer prime presentation in terms of the effect of semantic transparency (e.g. Rastle et al., 2000; Marslen-Wilson et al., 2008) and of pseudoword priming (Longtin and Meunier, 2005; Meunier and Longtin, 2007) suggests that there are two stages of morphological processing (Rastle, Davis and New, 2004): The first stage is an early obligatory morphological decomposition stage in visual processing, which is blind to the semantics of the morpheme combination, as evidenced by the non-effect of semantic transparency in masked priming. Longtin and Meunier’s (2005) study further shows that this process is lexically blind, triggered both by morphologically complex words and nonwords. However, as shown by Rastle et al. (2004), the decomposition process is not morphologically blind: it requires that the prime is made up of real morphemes, e.g. *corn-er* primes *corn*, but *broth-el* does not prime *broth*. The second stage is a semantic integration of the morphemes where a morphologically related prime only aids the recognition of the target if the morphological relation is semantically transparent. This account is based on masked priming and therefore limited in that it applies only to the visual modality. More importantly, it can be questioned to which extent the results of masked priming, with the quickly disappearing prime and consequently interrupted lexical processing, generalise to natural reading processes. In the non-primed gender decision task of Meunier, Seigneuric and Spinelli (2008), a root or pseudo-root with the opposite gender of the whole complex or pseudo-complex word slowed down the gender decision to the whole word, provided that a complete morphological parse of the target was possible. This indicates that effects of non-transparent, (pseudo-)morphological structure are not limited to very early stages of word recognition.

Effects of semantic transparency are also attested in experiments that do not employ priming: Wurm (1997) found that semantically transparent words were processed faster than semantically opaque words in auditory lexical decision and gating; Baayen et al. (2007) showed similar results in auditory and visual lexical decision. These results are arguably more indicative of natural language processes than the effects of semantic transparency in priming studies because in non-primed word recognition, semantic transparency is an intrinsic property of the complex experimental items. Semantic relatedness also affects word recognition in non-primed experiments: Rodd, Gaskell and Marslen-Wilson (2002) found that polysemy — several related senses of the same word — speeded up lexical decision latencies, while homonymy — several unrelated meanings — resulted in competition and longer lexicon decision time (see also Klepousniotou and Baum, 2007).

Semantic transparency is also involved in the distinction between free and bound stems. Free stems are more meaningful than bound stems, and complex words with free stems generally more transparent. Like other semantically opaque complex words, words with bound stems produce priming effects in brief masked priming (Forster and Azuma, 2000; Pastizzo and Feldman, 2004), but not in cross-modal immediate priming (Marslen-Wilson et al., 1994). In non-primed lexical decision, Wurm (2000) found longer rejection times for pseudowords with free than with bound stems; this could be an effect of the meaningfulness of free as opposed to bound stems. In the experiments reported below, bound stems were avoided in the morphologically complex items, and the distinction was not addressed.

Cross-linguistic differences

The review in the present section has outlined the broad tendencies of the literature on the relation between morphological and semantic priming effects, with little attention paid to the issue of differences between different languages. However, there seems to be at least one important difference in this respect between the concatenative morphology of Indoeuropean languages, exemplified by English, and the non-concatenative morphology of Semitic languages. In Semitic languages, roots are consonantal, and different words are formed by inserting different vowel patterns between the consonants of the roots. In languages like English and French, immediate priming tasks show no priming effects for semantically opaque morphologically related words, while both transparent and opaque forms prime their morphological relatives in Hebrew (Bentin and Feldman, 1990; Frost et al., 1997; Frost, Deutsch, Gilboa, Tannenbaum and Marslen-Wilson, 2000) and Arabic (Boudelaa and Marslen-Wilson, 2004b; Boudelaa and Marslen-Wilson, 2000 cited by Boudelaa and Marslen-Wilson, 2004a). Frost et al. (2000) explain this in terms of the fact that morphological structure is obligatory in Hebrew and Arabic, but not in for instance English; Plaut and Gonnerman (2000) explain a similar set of simulation results in a similar way.

Summary

Overall, the experiments that investigate semantic transparency indicate that morphological effects do to some extent depend on the degree of semantic transparency of complex words. As long as semantic transparency is dichotomised, as discussed above, this does not challenge the psychological reality of the morphological effects. In contrast, the more graded effects observed by Gonnerman et al. (2007) and, especially, Wurm (1997) and Baayen et al. (2007) are difficult to account for in terms of a theory of the mental lexicon according to which complex words are either represented as morpheme representations or as whole words with no internal morphological structure. The effects are compatible with a model that ascribes morphological effects to non-additive effects of overlap in semantics and form, like that proposed by Seidenberg and Gonnerman (2000) and further developed by Gonnerman et al., but graded effects can also be understood within a dual-route perspective, provided that the two routes are understood as interacting rather than racing. The masked priming studies which show morphological effects irrespectively of semantic transparency indicate that an additional process may be going on in visual word recognition, namely a very early, semantically blind morphological decomposition process. The results on formal transparency and formal priming effects are more mixed, allowing no firm conclusions to be drawn. However, the fact that several studies find no effects of formal transparency on morphological priming effects, as discussed above, calls into question the account of morphological effects as epiphenomenal of semantic and formal relatedness.

2.4 Frequency effects

Frequency effects are among the most basic in cognitive psychology generally (see Hasher and Zacks, 1984) and play an important role in the study of word recognition, both auditory and visual, where it has been found that more frequent words, morphemes, syllables and letter or phoneme strings are processed faster than less frequent ones, as reviewed in this section. The central role of frequency effects in the study of morphological processing relies on the fact that for morphologically complex words, the frequency of occurrence of the whole complex word can be compared to the frequencies of the constituents, thus gauging to what extent recognition of complex words is based on whole-word or constituent information or representations.

The terminology used to refer to the different frequencies for morphologically complex words varies: The corpus frequency of the whole complex word is termed whole-word frequency here, but is also known as surface or full-form frequency in the literature. The corpus frequency of the stem as an independent word can be called stem, root or base frequency; the term stem frequency is used here, to distinguish it from the base form, which can also refer to the citation form of a word, whether it is simple or complex. Base form frequency in this sense contrasts with the frequency of the lemma, which includes not only the citation form but also all other inflectional forms of the word in question. The present section focuses on token frequencies of words and morphemes while morphological family size, which is discussed in the next section, is a type count of the derived words and compounds that contain a given morpheme. Morphological family frequency is the token count corresponding to the morphological family size. All the frequencies index language users' familiarity with a given form; the frequencies used are typically based on written corpora, though spoken language frequencies may be preferable (Baayen et al., 2006). However, searchable spoken language corpora of sufficient size are unavailable for most languages, including Danish.

The basic understanding of frequency effects for morphologically complex words, at least since the seminal work of Taft (1979), is that whole-word frequency effects are the hallmark of whole-word representations of morphologically complex words, while stem frequency effects are the primary indication of morphemic representation of complex words. Thus, in the simplest case, full-storage models (see section 1.3) would predict whole-word frequency effects only, while full-parsing models would predict stem frequency effects only. When both whole-word and stem frequency effects are observed, it is taken as evidence of some kind of dual model. Such results may also be compatible with a connectionist model in which whole-word frequency effects originate with distributed whole-word representations and stem frequency effects are the product of generalisations across the representations that share the stem in question (Joanisse and Seidenberg, 1999; Seidenberg and Gonnerman, 2000). Whole-word frequency also affects the processing of monomorphemic words (such as those included in the experiments reported in chapters 5 to 7), but for polymorphemic words, the contrast between stem and whole-word frequency is central, as discussed in the following.

Whole-word frequency effects are found for a wide variety of affixed words, in a wide range of languages including English (Taft, 1979; Sereno and Jongman, 1997; Wurm, 1997; New, Brysbaert, Segui, Ferrand and Rastle, 2004; Wurm et al., 2006; Baayen et al., 2007), Dutch (Baayen, Dijkstra and Schreuder, 1997; Bertram, Schreuder and Baayen, 2000b; Tabak, Schreuder and Baayen, 2005), Swedish (Lehtonen, Niska, Wande, Niemi and Laine, 2006), French (Meunier and Segui, 1999a,b; New et al., 2004), Italian (Burani and Thornton, 2003), Spanish (Alvarez, Carreiras and Taft, 2001), Finnish (Vannest, Bertram, Järvikivi and Niemi, 2002; Moscoso del Prado Martín, Bertram, Häikiö, Schreuder and Baayen, 2004a; Soveri, Lehtonen and Laine, 2007) and Hebrew (Moscoso del Prado Martín, Deutsch, Frost, Schreuder, De Jong and Baayen, 2005). Stem frequency effects are less consistently found: many of the same studies (e.g. Taft, 1979; Baayen et al., 1997; Vannest et al., 2002; New et al., 2004) do report effects of stem frequency, but others do not (Sereno and Jongman, 1997; Wurm et al., 2006, the lexical decision experiments of Baayen et al., 2007). Most of these experiments employ lexical decision, though strong facilitatory whole-word frequency effects and more inconsistent stem frequency effects are also observed in gating by Wurm (1997) and in auditory and visual naming by Baayen et al. (2007).

The distinction between stem and whole-word frequency focuses on affixed words, but a similar distinction is relevant for compounds, between the frequency of the whole compound and the frequencies of the constituents, particularly the first constituent. As for affixed words, whole-word frequency effects for compounds (Hyönä and Olson, 1995; De Jong, Feldman, Schreuder, Pastizzo and Baayen, 2002; Bertram and Hyönä, 2003) are interpreted as evidence of whole-word representations of compounds, while compound constituent frequency effects (e.g. Bertram and Hyönä, 2003) are understood as evidence of decompositional processing. Since the experiments reported below focus on affixed words (though compounds are included in Experiment 3), this section is mainly concerned with frequency effects in affixed words, but similar considerations apply for compounds and are referenced where relevant.

Different properties of complex words have been shown to trigger either whole-word or stem frequency effects and have been invested with considerable theoretical importance. In general, specific types of complex words are predicted to show stem frequency effects, while whole-word frequency effects are expected elsewhere. One parameter is the regularity of morphologically complex words: regularity is a precondition for morphological processing in most models that operate with morphemic representations, but it is particularly important to the Declarative/Procedural (DP) Model (Ullman, 2001, 2004). The basic claim of the DP-Model is that regular complex words are produced and recognised through the application of a rule-mechanism, while irregular complex words must be recognised through whole-word representations in declarative memory. With respect to frequency effects, the strong original claim (Pinker, 1991) is dichotomous: Regular inflected words should show stem frequency effects only, as evidence of morphemic, rule-based processing, while irregular complex words should result in whole-word frequency effects only. However, the evidence is not consistent with this strong claim; for instance, Baayen et al. (1997) found whole-word frequency effects for regular Dutch plural forms. This leads to a modified claim, namely that regular complex forms may be, but do not have to be, stored as whole-word representations in declarative memory, in contrast to irregular forms which must be stored as whole words (Pinker and Ullman, 2002). One criterion that has been suggested for whether regular complex forms are stored is the frequency of the whole word (Alegre and Gordon, 1999): If a complex word is frequent — according to Alegre and Gordon, above a corpus frequency threshold of six per million — it may be economical to store it as a whole in the mental lexicon, with the result that high-frequency regular words are likely to show whole-word frequency effects. Below this threshold, complex words should be processed morphemically by the procedural system, resulting in stem frequency effects. Vannest et al. (2002) argue that the threshold is specific to individual derivational affixes, while Hay (2001) argues that it is the relative frequencies of the stem and the whole word which determine whether complex words are represented as morphemes (resulting in stem frequency effects) or as whole words (resulting in whole-word frequency effects).

Despite the modification of the strong original prediction, problems remain for the DP-Model in the pattern of frequency effects. Three aspects are problematic: Firstly, stem frequency effects are elusive even for regular words (Sereno and Jongman, 1997; Bertram et al., 2000b; Baayen et al., 2007). Secondly, stem frequency effects are occasionally found for irregular words (Baayen et al., 2007), though these may be understood within the DP-Model as generalisations across whole-word representations in associative memory, as they are in the Triangle model. In other words, regularity is neither a sufficient nor a necessary criterion for stem frequency effects. Thirdly, whole-word frequency effects are found even for low-frequency words by for instance Baayen et al. (1997), Baayen, McQueen, Dijkstra and Schreuder (2003) and Baayen et al. (2007), against the specific predictions of Alegre and Gordon (1999).

The pervasiveness of whole-word frequency effects in word recognition experiments also questions the proposed rule-mechanism of the DP-Model. A partial explanation of this pervasiveness of whole-word frequency effects within the framework of the DP-Model could be that there are differences between men and women in this respect, and that the majority of participants in word recognition experiments tend to be women. Ullman et al. (2002) report that females showed whole-word frequency effects for both regulars and irregulars in past-tense production studies in English and Spanish, while males only showed whole-word frequency effects for irregulars. This is presumably a consequence of the superior verbal memory of females (Kimura, 1999). In word recognition, the results are mixed: Tabak et al. (2005) found no significant effects of sex in visual lexical decision to Dutch inflected words, while Kuperman et al. (2008b) found stronger evidence of morphological decomposition for males than for females in an eye-tracking study. Many word recognition studies do not take the sex of the participants into consideration, and the results may be skewed as a consequence. The sex of the participants is included in the analyses reported below, but the pattern observed is only partly compatible with the predictions of the DP-Model.

The DP-Model mainly contrasts with single-mechanism models such as the Triangle Model of Seidenberg and McClelland (1989) and Joanisse and Seidenberg (1999). The Triangle Model is based on connectionist networks in which word representations are distributed over banks of orthographic, phonological and semantic units. Representational units which are recurrently activated together develop stronger interconnections. These stronger connections result in whole-word frequency effects for words, and may also result in frequency effects for morphemes. The central point for the single-mechanism position is that both types of frequency effects arise in the same system, with no categorical difference between regulars and irregulars, once variations in frequency and formal and semantic overlap are taken into account. Most single-mechanism models focus, explicitly or implicitly, on the visual modality. Models such as TRACE (McClelland and Elman, 1986) and Shortlist (Norris, 1994) model auditory word recognition, but they are not concerned with complex words. This means that it is unclear how connectionist models would account for the auditory recognition of morphologically complex words.

The pervasiveness of whole-word frequency effects may also be understood as evidence of exemplar-storage of whole words, without a specific commitment to a connectionist model architecture (cf. section 1.3). In an exemplar-based model, morphemic effects could arise because of morphological generalisations across exemplars (Hay and Baayen, 2005).

Another alternative to the DP-Model is the class of dual-route models. Dual-route models assume that complex words can be recognised via a whole-word route or via a morphemic route or both. Whole-word frequency effects indicate recognition via the whole-word route, while stem frequency effects are diagnostic of recognition via the morphemic route; if the routes are allowed to interact, effects of both frequencies may arise for the same words. Among the properties that determine the balance between the routes are semantic transparency (Schreuder and Baayen, 1995), with opacity favouring recognition by the whole-word route and thus triggering whole-word frequency effects. The logic is the same as for priming effects: Morpheme-related effects — whether priming effects or morpheme frequency effects — arise if the complex word is transparent. Affix homonymy is another parameter for the balance of stem and whole-word processing and thus for stem and whole-word frequency effects: Bertram et al. (2000b) found that complex words showed stem frequency effects only if the affix they carried was not homonymous with other affixes in the language. Affix homonymy is also taken into account in the experiments on Danish reported below. Bertram and colleagues further showed that more productive affixes and more meaning-invariant morphological processes (see section 2.6.2) also favour decompositional processing. Additionally, Järvikivi, Bertram and Niemi (2006) demonstrate for Finnish that stem frequency effects can only be obtained for complex words if their suffixes do not vary allomorphically.

Another dual model which partly relies on frequency effects is the Supralexical Model of Giraudo and Grainger. Giraudo and Grainger (2000) found that high-frequency but not low-frequency derived primes facilitate their free stem targets in a masked priming task. Giraudo and Grainger interpret this as evidence of morphemic representations being supralexical, i.e. occurring at a later stage in the word recognition process and at a higher level of the model: At the very short prime durations of the masked priming task, only high-frequency derived primes are strong enough to help activate supralexical morpheme representations. Moreover, the high-frequency primes are more resistant to orthographic inhibition from morphological relatives at the lowest word representation level than the low-frequency primes. These results may be problematic to a cascaded dual-route model (such as the AAM of Caramazza et al., 1988) or a dual-route race model (e.g. Frauenfelder and Schreuder, 1992), but in a more flexible dual-route model where the recognition routes interact (as proposed by Wurm, 1997 and Pollatsek, Reichle and Rayner, 2003), it is not problematic that a more frequent derived word is a stronger prime than a less frequent one.

Although Taft is often associated with a full-parsing position, as the main proponent of the prefix-stripping model, the prefix-stripping model does in fact operate with whole-word representations on a central level of the mental lexicon, while morphological decomposition is obligatory in the earlier access stages of word recognition. To some extent, then, the model of Taft (1994) is also a dual model, but not a dual-route model since the morphemic representations are activated first and the whole-word representations later. This model architecture provides a way of accounting for the existence of both stem and whole-word frequency effects: facilitatory stem frequency effects arise because the complex words are decomposed into stems and affixes, and more frequent stems are processed faster. Whole-word frequency effects arise at a later stage, where the morphemes that have been activated in the decomposition stage must be recombined; this recombination is easier for complex words that are more frequent. Taft (2004) also explains the absence of stem frequency effects in some studies by positing that effects at the combination stage may obscure or cancel out stem frequency effects: if two words are matched for whole-word frequency, but differ in stem frequency, then the word with the higher stem frequency may be at a disadvantage at the combination stage because the other inflectional forms that count towards its stem frequency are strong relative to the whole word actually encountered. When the combination stage is emphasised in experiments, in

Taft (2004) by the nonwords in a lexical decision task being non-existing combinations of real morphemes, base frequency may become inhibitory, but this pattern is significant only for correctness in Taft's study. Taft argues that similar processes could explain the absence of base frequency effects in other experiments. A major challenge to this interpretation comes from eye-movement studies which document whole-word frequency effects that occur before the whole word has been read (Kuperman et al., 2008a,b, see further below).

Problems with the traditional interpretations of frequency effects

In spite of their differences, the various dual models share the basic understanding of stem frequency effects as indicative of morpheme representations and whole-word frequency effects as evidence of whole-word representations of complex words in the mental lexicon. Even a single-mechanism model like the Triangle Model of Joanisse and Seidenberg (1999) understands whole-word frequency as item-based and stem frequency as generalisations across items, even though no discrete word or morpheme representations are posited. Taft (2004) argues that the absence of stem frequency effects does not necessarily mean the absence of morpheme representations. Otherwise, the associations between stem frequency effects and morpheme representations and between whole-word frequency effects and whole-word representations are so strong that absence of the relevant frequency effect is generally interpreted as the absence of the corresponding representations. There are a number of problems with this understanding:

One main problem is that, as described, stem frequency effects are elusive, while morphological family size effects are found even in the absence of stem frequency effects (Bertram et al., 2000a; De Jong et al., 2002) and provide evidence of some degree of morphemic processing (see section 2.5). There are two ways of interpreting cases of absent stem frequency effects: Firstly, the absent stem frequency effects can be explained in the various ways outlined above — frequency thresholds, relative frequency, word formation type or affix homonymy, allomorphy and productivity. These explanations are based on the idea that if stem frequency effects are absent, the complex words in question cannot be represented morphemically, but must be stored as whole words. The second way of understanding the absent stem frequency effects is in terms of the stem frequency effects being too weak to be a reliable index of morphemic processing, but that morphological structure may be relevant all the same. This view is to some extent implicit in Taft's (2004) argument that early beneficial effects of stem frequency are cancelled out by later inhibition at a combination stage, but that explanation is challenged by the eye-movement studies of Kuperman et al. (2008a,b). A more convincing demonstration of the weakness of stem frequency effects is presented by

Baayen et al. (2007): Sampling randomly from a large dataset from the English Lexicon Project (Balota et al., 2002), Baayen et al. found that at least 1000 inflected and derived items were necessary to detect stem frequency effects in more than 50% of samples.

In addition to the more practical problems with stem frequency as the prototypical diagnostic of morphemic processing, there are at least two conceptual problems: Firstly, there is evidence from auditory word recognition that listeners are sensitive to subphonemic cues that differentiate a base in a complex word from the same base occurring as an independent word (Kemps et al., 2005a,b). Such subphonemic differences could detract from the relevance of stem frequency as a measure of the processing of complex words, because the stem frequency is based on the stem occurring as an independent word. Secondly, stem frequency measures the probability of the stem occurring, without considering the specific position of the stem or its morphological context in the given complex target (Baayen et al., 2007). The frequency of the stem (whether in its base form or including the entire inflectional paradigm) may be an inappropriate measure of the recognition of the stem occurring in a specific complex word. In itself, the stem frequency does not indicate how relevant the semantics of the stem is to the recognition of the complex word; the occurrence of the string *hat* in the word *that* (cf. section 2.3.1 above) is an extreme case of such irrelevance. Together, these considerations indicate that stem frequency may not be the reliable index of morpheme-based processing that it has previously been thought to be; this also helps explain why stem frequency effects are not systematically detected in standard non-primed word recognition studies.

If stem frequency effects do not constitute a reliable index of morphemic processing, alternative measures may be desirable. One obvious candidate is the converse constituent frequency, namely the frequency of the affix of inflected or derived words. Effects of affix properties, such as frequencies, have been found for Italian nonwords by Burani, Dovetto, Thornton and Laudanna (1997) and Burani and Thornton (2003), while for words, the evidence presented by Burani and Thornton (2003) is more mixed, with the most well-controlled experiment showing no effects of suffix frequency. Wurm (1997) found inhibitory effects of prefix frequency but explained these as the result of the denser lexical neighbourhoods of frequent prefixes. The effect of affix frequency is investigated in the experiments reported below, to assess the possibility of using affix frequency as an alternative index of morphological processing. Affix frequency can also be used for calculating measures of affix productivity (Baayen, 1992, 1993), but effects of these have only been found in one RT-based recognition study, namely a facilitatory effect in the lexical decision task of Baayen et al. (2007). The productivity measures are also investigated in the experiments reported below, for more detail on these measures see section 3.4.2. Affix productivity has also been understood as a prerequisite for morpheme representation and thus for stem frequency effects, along with absence of affix homonymy and allomorphy as described above.

Another alternative measure of morphemic processing may — contradictory as it may sound given the traditional interpretation — be whole-word frequency, as proposed by Baayen et al. (2007). Like the problems with stem frequency discussed above, this reinterpretation of whole-word frequency is based on the observation that it is more important for morphemic processing whether a morpheme is licensed as a morpheme by its context, than whether it occurs or not. The frequency of a bimorphemic word not only indexes listeners' familiarity with the whole string, which is the traditional interpretation, but also the probability of the two constituent morphemes co-occurring: if whole-word frequency is high, the combination of the morphemes is familiar and the interpretation of the combination is easier. A similar understanding of whole-word frequency is implicit in Taft's (2004) argument that whole-word frequency effects arise at a combination stage of word recognition, but Baayen et al.'s argument has the advantage of not relying on a specific ordering of whole-word and morphemic representations. This means that the very early whole-word frequency effects observed by Kuperman et al. (2008a) and Kuperman et al. (2008b) challenge the interpretation of Taft, but support the interpretation of Baayen et al.: when whole-word frequency affects reading measures at a stage where the whole word has not yet been read, it is evidence that the whole-word frequency effect is at least partially a combinatorial probability, since it is triggered by the reading of the first constituent, along with parafoveal information about the length of the whole word and, in the case of the sentence-embedded compounds of Kuperman et al. (2008a), the sentential context of the word.

This new interpretation of whole-word frequency is radically different from the traditional one in that it posits that whole-word frequency can be understood as a measure of morphemic processing. At the same time, the new interpretation co-exists with the traditional interpretation of whole-word frequency: Within an information-theoretical framework (see Baayen et al., 2007; Kuperman et al., 2008a), whole-word frequency can both be understood as the string probability of the whole-word and the combinatorial probability of the constituents. As a string probability, whole-word frequency measures the a priori, long-term probability of the word relative to all other words in the mental lexicon of the language user, indexed by its frequency in a representative corpus. Stem frequency may also be interpreted as the long-term, non-conditional probability of the stem in question. As a combinatorial probability, whole-word frequency represents the processing of one constituent given that the other has been processed. The probability of a word or constituent in isolation is small, because it is the probability relative to a large corpus (in language use, the words processed by the language user, in experimental work, the relevant corpus of written language). The combinatorial probability of the second constituent given the first constituent is larger because it is the probability of the target constituent combination relative to the restricted set of words that contain the first constituent, rather than relative to the whole mental lexicon of the language user. The fact that whole-word frequency involves both the long-term and the conditional probabilities may be one reason why whole-word frequency effects are generally stronger than stem frequency effects, and also much more systematically found. This new interpretation is the basis of the interpretation of the frequency effects and the interactions between whole-word and constituent frequencies in Experiments 1a and 2, see section 6.3.3 and chapter 8.

A central point of the information-theoretical perspective is that the different frequencies are not independent of each other. For any given set of words, the different frequencies are assigned specific weights which determine whether they interact and what the direction of interaction will be; this explains the different balances between stem and whole-word frequency effects observed in different studies. In terms of a route-metaphor, this translates to the different processing routes not being independent. This is an important advantage of the framework, because a number of effects in the literature show interactions between whole-word and morpheme-based processes, including the graded effects of semantic transparency (see section 2.3.3) and the frequency and productivity interactions observed by Baayen et al. (2007), Plag and Baayen (2008), Kuperman et al. (2008a) and Kuperman et al. (2008b). In the present experiments, particularly the interactions between whole-word and constituent frequencies are evidence of such interaction.

Frequency and age of acquisition

It has been argued that frequency effects may at least partially be artefacts of the age at which the relevant words were acquired (e.g. Morrison and Ellis, 1995; Stadthagen-Gonzalez, Bowers and Damian, 2004). Age of acquisition is also a possible factor in the frequency effects observed in the studies summarised in the present section and in the experiments reported below. However, the effect of age of acquisition and its relation to frequency is not undisputed, cf. for example Zevin and Seidenberg (2002); although age of acquisition probably has some effect, the evidence seems to indicate that frequency also plays a role over and above this effect. Moreover, as argued by Baayen et al. (2007), age of acquisition is likely to play a smaller role for low-frequency complex words which are likely to be acquired relatively late. Particularly the derived words, which make up the majority of the complex items in the experiments reported below, are likely to be acquired relatively late and thus be less susceptible to the influence of age of acquisition. The role of age of acquisition cannot be assessed in the present experiments because age of acquisition norms are only available for the earliest words in Danish (Wehberg, Vach, Bleses, Thomsen, Madsen and Basbøll, 2007). Age of acquisition effects themselves may be confounded with spoken frequency: if the spoken frequency of words is controlled, age of acquisition does not account for much additional variance (Baayen et al., 2006).

Other frequency effects

Other frequencies are relevant than the token frequencies of words and morphemes. Of these, the most important to the study of morphological processing are the morphological family effects discussed in the next section, while other frequencies are relevant to the present experiments mainly as control variables. The mean frequency of the letter pairs (bigrams) that make up written words and of the phoneme pairs (diphones) that make up spoken words may influence processing time; the evidence is mixed (see overview for instance in Westbury and Buchanan, 2002), but the variable should be controlled. Since a sufficiently large corpus is only available for written language in Danish, bigram frequency is used as the index of diphone frequency for the auditory experiments, although this is problematic for the deep orthography of Danish. In addition to the mean bigram frequency, the frequency of the bigram that occurs at the juncture between the constituent morphemes of a complex word may be relevant: The perception of word boundaries is eased by low-probability transitions between words; similarly, morpheme junctures are more salient if the frequency of the juncture is low and thus less likely to occur inside a word (Hay, 2002). The idea of so-called bigram troughs at syllable junctures was introduced by Seidenberg (1987) to account for experimental effects of syllable structure. More salient morpheme boundaries help listeners segment complex words, and may also ease recognition by making morpheme-related information more available. Juncture salience is indexed by juncture bigram frequency, in the absence of suitable spoken language corpora for the extraction of diphone frequencies. Finally, effects of the frequency of the syllable have been reported in the literature (see for instance Carreiras and Grainger, 2004), but the extraction of syllable frequencies requires intensive computational and manual work, and this variable was therefore not included in the analyses of the present experiments.

Summary

In sum, the evidence of stem frequency effects is mixed and the relevance of the frequency of the stem as an independent word is questionable when the stem is encountered in the context of one or more other morphemes as it is in morphologically complex words. This means that the traditional understanding of stem frequency effects as the primary diagnostic of morphemic processing becomes problematic. Stem frequency effects may occur, and they may be indicative of the long-term probability of the stem, but the presence or absence of stem frequency effects cannot be taken as an absolute measure of the presence or absence of morphemic processing. Whole-word frequency effects are much more systematically observed for a wide range of word types in different languages. Whole-word frequency is generally understood as indicating whole-word processing of complex words, but can also be interpreted as measuring the combinatorial probability of the constituents of a complex word, the likelihood of the constituents occurring together. This double function of the whole-word frequency effect can easily be interpreted within an information-theoretical framework. The information-theoretical perspective discussed above also provides a promising way of understanding complex interactions between different frequencies.

In the experiments reported below, stem frequency is investigated, but both affix and whole-word frequency are also included as potential alternative measures of morphemic processing. Both male and female participants were recruited for the experiments, and the effect of sex on frequency effects assessed. Interactions between the different frequencies, which favour an information-theoretical account, are also investigated in all three experiments and observed in two. Finally, the effects of juncture bigram frequency were investigated as an index of juncture salience.

2.5 Morphological family

A word's morphological family size is a type count of all derived forms and compounds that contain the stem of the target word, or contain the whole target word if it is simple. The summed token frequency of the family members is termed the morphological family frequency. For example, the morphological family of the Danish word *u-stabil* ('unstable') consists of all derived and compound words that contain the stem *stabil*; *u-stabil* has 119 family members (in the Danish corpus described in chapter 3), including words like *stabil-isere* ('to stabilise') and *varme-stabil* ('heat stable'). The summed token frequency of these 119 family members is the morphological family frequency of *u-stabil*.

Morphological family size has been found to facilitate processing of both simple (Schreuder and Baayen, 1997; Baayen et al., 2006) and complex words (Bertram et al., 2000a; De Jong, Schreuder and Baayen, 2000) in as different languages as Dutch, English, German, Finnish and Hebrew (Moscoso del Prado Martín et al., 2004a). Generally, words with larger morphological families are recognised faster in visual lexical decision, but recent regression studies (Tabak et al., 2005; Baayen et al., 2006) have shown inhibition in the higher ranges of family size. In bimorphemic compounds, both the first and the second constituent families are relevant, with stronger effects when the families of the first and second constituents are restricted to words in which the constituent in question occurs in the same position as in the target (De Jong et al., 2002). Most of the studies documenting effects of morphological family size are visual lexical decision experiments. For auditory lexical decision, Baayen, Tweedie and Schreuder (2002) report morphological family size effects for simple words and Wurm et al. (2006) for prefixed words. The experiments reported below test the effects of morphological family size and frequency in Danish for both simple and different kinds of morphologically complex words.

The effect of morphological family size (the type frequency) is generally stronger than the effect of morphological family frequency (the token frequency): Family size and frequency are usually highly correlated, making the effects difficult to tease apart, but Schreuder and Baayen (1997) and De Jong et al. (2000) found stronger correlations between family size and visual lexical decision latency than between family frequency and latency. Baayen et al. (2002) even show that family frequency can be inhibitory while family size is facilitatory. This family frequency effect is interpreted as a result of competition between family members, while the family size effect reflects the activation of shared semantics, as discussed below. The duality reflected in the recent regression studies, with inhibition in the high ranges, may reflect similar competition processes, see also the Danish results discussed in section 5.3.4. In contrast to this, De Jong et al. (2002) found that family frequency was a better predictor than family size, for the position-specific families for first and second constituents in bimorphemic compounds in Dutch and in English concatenated compounds (see similar results in Gagné and Spalding, 2008).

Derivational entropy (Moscoso del Prado Martín et al., 2004b) provides a way of including both family size and frequency. Derivational entropy is Shannon's entropy (briefly explained in section 2.3.1) calculated across the family members and can be understood as a token-weighted type count. Higher derivational entropy is associated with larger families and/or with more equiprobable family members, i.e. more uncertainty in the family. The higher the uncertainty, the more information is encoded in the family, and the easier it is to recognise the target in visual lexical decision (Moscoso del Prado Martín et al., 2004b; Baayen et al., 2006). Inflectional entropy based on inflectional paradigms (Moscoso del Prado Martín et al., 2004b) is parallel to derivational entropy which is based on morphological families. The inflectional entropy of a word is based on the number of members of the word's inflectional paradigm, weighted according to the token frequencies of the members; in a sense, then, inflectional entropy replaces the lemma frequency of a word (the cumulative frequency of all the inflectional forms of the word), but also considers the distribution of probabilities in the paradigm. Higher inflectional entropy is associated with shorter response latencies in visual lexical decision and naming (Baayen et al., 2006).

Another way of considering both family size and the frequency of the family members is applied by Meunier and Segui (1999a) who report that words are recognised faster when they have few family members with a higher frequency than the target, compared to words with more family members of higher frequency than the target. Meunier and Segui include only derived forms in the family, whereas the standard morphological family measure of Schreuder and Baayen (1997) includes both derived words and compounds, but the logic is similar. The results of Meunier and Segui show a sensitivity to the relative probabilities of the members of the morphological family that is similar to that observed by for instance Baayen et al. (2006) in the shape of a derivational entropy effect; the two approaches differ in that Meunier and Segui operate with the probability of family members relative to the target, while the derivational entropy involves the probability distribution of the family without reference to the specific target.

Similarly to Meunier and Segui (1999a), Colé, Segui and Taft (1997) found word frequency effects for complex words only when word frequency was higher than derived family frequency (like Meunier and Segui not including compounds in the family). Conversely, Colé et al. observed a derived family frequency effect only when this frequency was higher than the word frequency. This effect of relative frequencies is similar to the effects of relative stem and whole-word frequency observed by Hay (2001), but is based on a family measure (the frequency of all derived forms of the stem in question). Interestingly, the derived family frequency of Colé et al. was inhibitory, indicating that the derived forms of a simple word compete with the simple target word when the cumulative frequency is higher than the word frequency of the simple target word.

Morphological family size effects are the result of more systematic form-meaning mappings (as formulated by Feldman, Basnight-Brown and Pastizzo 2006: 60) when families are large than when they are small, which make words easier to recognise. The relations between form and meaning are less relevant in naming tasks which are less semantic in nature than lexical decision, and family effects are generally not found in naming (cf. for instance Baayen et al., 2006, 2007). The evidence of the semantic nature of morphological family size effects is substantial and varied: De Jong et al. (2000) find similar effects of morphological family size for regularly and irregularly inflected verbs, indicating that the morphological family size effect cannot be the result of low-level formal overlap. Several studies also show that the stronger the semantic coherence of the morphological families the stronger the facilitatory family effects (Bertram et al., 2000a; Moscoso del Prado Martín et al., 2004a), though Lüdeling and De Jong (2002) report stronger family size effects for opaque than transparent German particle verbs. Moscoso del Prado Martín et al. also report substantial cross-language predictivity of morphological family size effects; for instance, lexical decision latencies to a set of Dutch words correlated significantly with the family sizes of the Finnish translation equivalents of the Dutch words, while Dutch family size predicted decision latencies to Hebrew translation equivalents. This is strong evidence that the morphological family size is semantic or conceptual, since no formal relatedness exists between a Finnish word and the morphological family of its Dutch translation equivalent. No cross-linguistic family size effects were observed between Finnish and Hebrew; Moscoso del Prado Martín et al. argue that the differences in lexical organisation between Finnish and Hebrew are too large for such an effect to emerge. De Jong et al. (2000) found evidence that uninflected verbs activated only nominal family members, while inflected verbs also activated verbal families. Similarly, Bertram et al. (2000a) found that for a set of complex adjectives, the adjectival family members were not activated if they were incompatible with the specific affix on the targets. Both studies provide evidence of a genuine morphological component in the morphological family size, showing that the morphological family size effect is not only significant because it indexes the semantic connectivity of words in the mental lexicon.

The argument that morphological family effects are semantic in nature is also supported by the priming studies of Feldman and colleagues: Feldman et al. (2004) found that the difference in morphological priming effects between semantically transparent and semantically opaque pairs was larger when the family size of the prime-target pairs (which shared their stem and thus the family) was large. This effect of family size was found in both immediate and brief masked priming when family frequency and prime-target relatedness ratings were partialled out. Feldman, Basnight-Brown and Kanai (2007) also report an effect of morphological family size from a series of masked priming experiments. The fact that a semantic effect of morphological family size arises even at very brief prime exposures supports the very fast activation of semantics in word recognition, questioning a sequential model according to which form representations are activated and processed before semantics (e.g. Taft, 2004; Rastle et al., 2004).

The morphological family size effect is evidence of the role of morphology and morphological paradigms (in the widest sense) in word recognition. The effect as defined here originates in the dual-route framework of Schreuder and Baayen (1997) and can straightforwardly be accounted for within that model. Family effects do not provide evidence of decompositional processing: Morphological family effects size are reported also for words that do not exhibit constituent frequency effects (Bertram et al., 2000a; De Jong et al., 2002). The semantic nature of the morphological family size effects may be problematic for decompositional models that locate morphological effects on an access level of word recognition and whole-word effects at later, semantic stages (Taft, 2004). It is also questionable whether connectionist models can account for the morphological family size effect; the sensitivity of such models to token frequency makes the dominance of the type frequency of the morphological family problematic (Krott, Baayen and Schreuder, 2001). In the Supralexical Model, the whole-word representations at the lexical level which share a morpheme representation at the supralexical level are morphological family members; therefore, the morphological family effects follow naturally from this model.

The experiments reported below include morphological family size as a covariate; the surprising direction of this effect in Experiment 1a then motivated further investigation of this variable in Experiments 2 and, especially, 3a and 3b. The experiments add to our knowledge of morphological family size effects both by investigating a "new" language, Danish, which has relatively large morphological families, and by using the auditory lexical decision task, while most previous results come from visual lexical decision.

2.6 Types of words and types of morphology

This section deals with contrasts and similarities between words that are categorised as different types based on their morphological structure or lack of it. Four subsections discuss issues that are relevant to the experiments presented in chapters 5 to 7: First, section 2.6.1 outlines processing differences between morphologically simple and complex words. Section 2.6.2 then discusses differences between inflected, derived and compound words, based on morphological theory as well as psycholinguistic evidence. Section 2.6.3 focuses on the difference between regular and irregular inflection, which is less relevant to the experiments of this thesis, but crucial to the Declarative/Procedural Model. Finally, section 2.6.4 addresses the issue of differences between prefixed and suffixed words which plays a role in the understanding of the experimental results reported below.

2.6.1 Simple vs. complex words

The study of morphological processing often focuses on different properties of morphologically complex words, of which the most important were reviewed in the preceding sections. However, a more direct way of investigating the idea that morphology is functional may be to compare the processing of morphologically simple and complex words, as it is done in the experiments reported below. If morphological structure makes processing more efficient, morphologically complex words should have an advantage over simple words.

Several studies of Finnish (Bertram, Laine and Karvinen, 1999; Lehtonen and Laine, 2003; Lehtonen, Cunillera, Rodríguez-Fornells, Hultén, Tuomainen and Laine, 2007; Soveri et al., 2007) and one study of Swedish (Lehtonen et al., 2006) use the comparison between complex and simple words as a diagnostic of morphological processing: When complex words are recognised more slowly than simple words in visual lexical decision (e.g. Lehtonen et al., 2006; Experiment 1 of Bertram et al., 1999) or progressive demasking (Laine, Vainio and Hyönä, 1999), this is interpreted as evidence of the complex words being recognised only via the morphemic route of a dual-route model, requiring time-costly decomposition and recombination. ERP-measurements suggest that processing costs associated with morphological parsing occur at a late semantic-syntactic level (Lehtonen et al., 2007), i.e. at a recombination stage. In contrast, processing exclusively via the whole-word processing route is assumed when complex and simple words are recognised equally fast (e.g. Experiment 2 of Bertram et al., 1999). Finally, when complex words are recognised faster than simple words, as in Experiment 3 of Bertram et al. (1999), this advantage is explained in terms of the morphemic and the whole-word routes racing to achieve word recognition. The availability of two routes results in statistical facilitation (see sections 1.3 and 8.2) for complex words relative to simple words which can only be recognised in one way, via the whole-word route. A similar advantage for complex words is observed by Ji, Gagné and Kemps (2006) in lexical decision and by Inhoff, Briihl and Schwartz (1996) in naming for English compounds relative to simple words (though in the latter study, the control between the compound and the simple word conditions is arguably insufficient). Burani and Thornton (2003) report an advantage for derived words with high-frequency roots relative to simple words in Italian.

There are thus various outcomes of comparisons of simple and complex words in word recognition tasks. The explanation of these different outcomes that follows from the reasoning of Bertram, Lehtonen and colleagues is that some complex words are at a disadvantage because of the parsing route being relatively slow, while other complex words have an advantage because of the statistical facilitation that arises from the activation of two processing routes. This, however, is arguably based on a simplified understanding of the functionality of morphology, focusing exclusively on the storage efficiency of the parsing route. Moreover, this view of the processing routes is binary, recognition being achieved either via the whole-word route or via the decompositional route, also when both are available, but there is evidence that suggests that the two routes interact in the recognition of complex words, for instance in the interaction of different frequency measures observed by Baayen et al. (2007), Kuperman et al. (2008a) and Kuperman et al. (2008b), and in Experiments 1a and 2 below.

The comparison of morphologically simple and complex words is also interesting from a developmental perspective: Burani, Marcolini, De Luca and Zoccolotti (2008) report that dyslexic readers and unskilled young readers benefited from morphological structure in reading aloud of Italian words, while skilled readers did not. Similarly, Elbro and Arnbak (1996) showed for Danish that dyslexic readers benefited from transparent morphological structure while normally developing young readers did not. In adult word recognition, advantages for complex words are interpreted in terms of the availability of two processing routes. In contrast, the developmental results are interpreted as evidence that the morpheme provides a unit of an intermediate size which is manageable to dyslexic and unskilled readers, while grapheme-based reading leads to very slow analytic reading and words may be too long as reading units for these struggling readers.

The processing of simple and complex words is also compared in the experiments reported below. These new experiments are relevant because, in contrast to the studies reviewed in this section, regression designs are employed. Regression designs allow the comparison of typical examples of simple and complex words, as well as the inclusion of a number of relevant variables which are difficult to control between groups of simple and complex words. The prediction that, all things being equal, complex words are recognised faster than simple words, follows from the notion that morphology is functional (see section 1.4 above).

2.6.2 Inflection, derivation and compounding

In both morphological theory and psycholinguistics, the difference between inflection and derivation has been the subject of considerable attention, while the contrast between derivation and compounding has been more peripheral. One reason for this is that inflection can be understood as combinatorial and is therefore more relevant to what Jackendoff (2002) terms the dominant syntactocentric view of language. The distinction between inflection and derivation is basic in the Lexicalist Hypothesis (originating with Chomsky (1970) and to some extent psycholinguistically implemented in the dual-mechanism model of Pinker (1997b)) which argues that syntax and lexicon are distinct modules, with inflection being syntactic and derivation lexical.

There are a number of criteria for distinguishing between derivation and inflection (see for example Bauer, 1983 and Haspelmath, 2002): for instance, inflection tends to be obligatory, semantically regular and productive. The fact that some of the criteria proposed in the literature are graded and the fact that there are exceptions to several of the dichotomous criteria suggest that morphological processes are best understood as falling on a continuum (Haspelmath, 2002). On the other hand, the criteria do tend to partition morphological processes along the same lines (Stump, 1998), suggesting that there may be a categorical difference between inflection and derivation. Either way, the different criteria can be used to distinguish between prototypical inflection and prototypical derivation; in this thesis, the categories of inflected and derived words are mainly such prototypical cases. The difference between category and continuum views in theoretical morphology is aligned with psycholinguistic models: the DP-Model is based on a category view, while single-mechanism models predict continuum-effects. The different positions can be tested by comparing the processing of inflection and derivation.

The psycholinguistic evidence is mixed: In priming tasks, Raveh and Rueckl (2000) reported no categorical differences between inflected and derived forms, while Feldman (1994; Feldman et al., 2006) found stronger priming effects for inflected than for derived forms. In unprimed lexical decision and naming, both auditory and visual, Baayen et al. (2007) observed no effects of the categorical difference between inflection and derivation; the same holds in Experiment 2 of this thesis. In contrast to the categorical difference, effects of the graded differences between inflection and derivation are observed, notably effects of semantic transparency (see section 2.3.3 above) which tends to be high for inflectional forms and varying for derivational forms. It is problematic for the DP-Model when the categorical difference is non-significant and the graded differences are significant, because the DP-Model is based on a binary distinction which does not easily account for graded effects. A categorical difference between inflection and derivation does emerge in Baayen et al.'s (2007) analysis of a large dataset from the English Lexicon Project (Balota et al., 2002), in line with the predictions of the DP-Model, but the categorical difference may be due to differences in semantic transparency which were not controlled. In the same analysis, inflected and derived words showed equivalent stem frequency effects which again questions the predictions of the DP-Model.

Inflection and derivation differ in the extent to which they change the semantics of their stems, which is related but not identical to differences in semantic transparency of the whole complex words. Bertram et al. (1999) and Bertram et al. (2000b) operate with a continuum of morphological processes from meaning-invariant via meaning-adding to meaning-changing morphological processes, which is similar to Booij's (1993) distinction between contextual and inherent inflection. Many derivational forms are meaning-changing, while meaning-adding morphology includes both affixes usually categorised as derivational (e.g. diminutives) and affixes usually understood as inflectional (noun plurals). Inflectional agreement markers do not change the meaning of their stems. Bertram and colleagues argue that the more meaning-changing the process, the less likely the resulting complex word is to rely on morpheme-based processing. Morpheme vs. whole-word processing is also co-determined by productivity. Productivity is also argued by Blevins (1999) to be the primary difference between forms categorised as inflectional vs. derivational, though chronometric effects of quantitative productivity measures are scarce in the psycholinguistic literature (but found by Baayen et al., 2007).

Experiment 2 (see chapter 6) is concerned with the difference between inflected and derived words, investigating the categorical difference and two graded differences that can be implemented experimentally, namely semantic transparency and affix productivity.

The difference between derivation and compounding is also relevant to the present experiments, in two ways: Firstly, differences between derivation and compounding are investigated in Experiment 3 (see chapter 7). Secondly, the distinction between derivational forms and compounds is not always straightforward; since the experiments are centred around derivation, the definition of derivation matters. On a narrow definition, derivational affixes are recurrent morphemes which do not correspond to independently occurring words with a corresponding meaning (the criterion applied to Danish by Hansen (1967) and Diderichsen (1987)). A broader definition is given and followed by Arndt (1997) who argues that a morpheme is derivational when it is obviously and recurrently less semantically heavy than the other constituent in bimorphemic words. A similar argument is put forward by Beard (1998). This definition has the drawback of necessitating a definition of semantic heaviness which can be problematic. However, the narrow definition of Hansen (1967) and Diderichsen (1987) does not avoid problems with semantics as this definition hinges on what is considered meaning identity between a constituent of a complex word and a potentially corresponding independent word. In both cases, derivational affixes tend to be semantically more regular than compound constituents, and derived forms tend to be more semantically transparent, graded differences similar to the differences between inflection and derivation.

In Danish, one set of words that are categorised as derived rather than compounds according to the broad but not the narrow definition is complex words with non-separable particles, i.e. morphemes such as *op* ('up') and *efter* ('after') that can function as prepositions or particles but also as non-separable prefixes. These are termed particle prefixes and included in Experiments 1 and 3. The recurrence of these particle prefixes in many complex words and their relative semantic weakness makes them more similar to derivational morphemes than most compound constituents are. The particle prefixes are understood as derivational but also to some extent constitute an intermediate category between derivation and compounding. Particle prefixed words are included as a separate category in Experiments 1 and 3 to investigate possible similarities and differences to standard derived words and to compounds. Some English translation equivalents of the Danish particle prefixes, e.g. *over-*, are defined as derivational prefixes in CELEX (Baayen, Piepenbrock and Gulikers, 1995), and therefore included in the experiments of for instance Baayen et al. (2007). Schreuder, Grendel, Poulisse, Roelofs and van der Voort (1990) found morpheme priming effects for separable verbal particles in Dutch and German, but not for non-separable particles or inflectional affixes. However, the complex words with non-separable particle prefixes in Danish are likely to show morphological effects in non-priming experiments like those reported below (like the German particle verbs investigated by Lüdeling and De Jong, 2002), and it is an interesting question to what extent they are more similar to compounds or to derivations.

Factors such as semantic regularity and transparency vary both between inflection and derivation and between derivation and compounding. This favours a continuum approach of morphological processes, with derivation at the centre. The debate in theoretical morphology is complex and cannot be resolved in the present context; however, the corresponding psycholinguistic differences are tested in the experiments reported below.

2.6.3 Regular and irregular inflection

The regularity of morphological processes is important to models of morphological processing in general, and crucial to the DP-Model of Ullman (2001, 2004) and its predecessors, which operate with two distinct mechanisms, one for regular morphological forms and one for all other words, including both simple words and morphologically complex forms that are irregular in some way. The difference between inflection and derivation discussed in the previous section is partly an issue of regularity: the focus tends to be on regular inflected forms vs. irregular derived forms, such that the contrast becomes one of regularity. However, inflection may also be irregular, resulting in a similar contrast of regularity for inflectional forms. In terms of semantics, regular and irregular inflected forms are more similar than derived and inflected forms are, while formally regular derived and regular inflected words tend to be more similar in terms of phonological overlap between base and complex forms than regular and irregular inflected forms. In this way, the two comparisons may complement each other, though the primary focus in the debate between single- and dual-mechanism models has been on the distinction between regular and irregular inflection. In addition to semantic and phonological transparency, the regularity distinction is related to frequency because irregular inflectional forms generally have to be relatively high-frequent in order to survive as irregulars (Lieberman, Michel, Jackson, Tang and Nowak, 2007).

Within the context of the DP-Model, the regular default forms (such as the English *-ed* past tense) are argued to apply by rule independently of phonological similarity to existing exemplars. In contrast, words that are irregularly inflected form clusters in phonological space. An example of such a cluster of irregular words are English verbs whose present tense rhymes *ing* become *ang* in the past tense (*ring–rang* etc.). Generalisations of irregular inflection to novel forms are argued to be based on similarity to these clusters. The independence of regular forms from phonological similarity is crucial to the predictions of the DP-Model (see for instance Pinker and Ullman, 2002). In contrast, single-mechanism models assume that regular and irregular forms are processed in the same way and thus predict similar effects of similarity (and of frequencies as discussed above, section 2.4) for regular and irregular forms. Albright and Hayes (2003) found that naturalness ratings of novel past tense forms depend on the phonology of the stem both for irregulars as predicted by most models and, in opposition to the predictions of the DP-Model, also for regulars.

The phonological differences relative to their present tense forms are larger for irregular than for regular past tense forms in English, a factor which may explain processing differences between regulars and irregulars without a separate mechanism for processing regular forms (as is done for instance by Joanisse and Seidenberg (1999) in their modelling of acquired impairments in past-tense morphology) . In contrast, semantic differences are generally not assumed to exist between regular and irregular past tense forms in English. However, Baayen and Moscoso del Prado Martín (2005) show that there are also semantic differences between verbs with regular vs. irregular past tense forms: The irregular verbs tend to cluster in denser regions of semantic space than regulars, measured in a number of different ways. These semantic differences influence behavioural measures of word recognition and may also explain neuroimaging results that indicate differential processing of regulars and irregulars (e.g. Jaeger, Lockwood, Kemmerer, Van Valin, Murphy and Khalak, 1996). The results of Baayen and Moscoso del Prado Martín (and similar observations by Tabak et al., 2005) lead to two important conclusions: Firstly, such subtle but significant semantic differences may be confounding factors in existing experiments and should be controlled. Secondly, like the morphological family size effects discussed in section 2.5 above, the results demonstrate the sensitivity of language users to probabilistic differences. Influences of verb meaning on the choice of regular vs. irregular past tense forms are also attested by Ramscar (2002).

In the priming literature, categorical differences between regulars and irregulars are not reliable in Italian (Orsolini and Marslen-Wilson, 1997) or French (Meunier and Marslen-Wilson, 2004). In English, priming effects have been found for regular but not irregular forms (Stanners, Neisser, Hernon and Hall, 1979; Kempley and Morton, 1982), but Pastizzo and Feldman (2002) argue that this may be due to inappropriate baseline conditions. In a masked priming task, Pastizzo and Feldman found significant priming effects for base forms by their regular past tense forms and by irregular past tense forms with high formal overlap (such as *fall–fell*). Irregular past tense forms with low formal overlap did not significantly prime their base, e.g. *teach–taught*. This suggests that the degree of formal overlap, rather than the categorical distinction between regular and irregular, determines the amount of priming.

A final difference between regular and irregular inflectional forms is that irregular plurals may occur inside compounds while regular plurals generally do not, e.g. cases such as *mice-eater* occur and are accepted by children, cases such as *rats-eater* are not (Alegre and Gordon, 1996). Proponents of the single-mechanism approach explain this tendency in terms of semantic and phonological constraints (Haskell, MacDonald and Seidenberg, 2003), while proponents of the DP-Model see it as a level-ordering phenomenon: derivation and irregular inflection applies at an early level of word formation which may precede the formation of compounds, while regular inflection applies at the latest level, after compound formation (Kiparsky, 1982). The proximity to the stem is also a criterion for distinguishing between derivation and regular inflection, though there are exceptions to the rule that derivation is closer to the stem than inflection (Stump, 1998). Hay (2002), Hay and Plag (2004) and Plag and Baayen (2008) show that parsing restrictions provide a more precise account of affix ordering than the level-ordering hypothesis.

The differences between regular and irregular inflection are not investigated here, partly because the experiments are centred around derivation, but the contrast is important to models of morphological processing, particularly the DP-Model.

2.6.4 Prefixes and suffixes

The difference between suffixed and prefixed words is relevant to the results reported in chapters 5 to 7 below, though none of the experiments set out a priori to test this difference. The position relative to the stem is particularly salient in auditory processing, making the differences between prefixes and suffixes more important in the auditory modality. Since auditory processing is developmentally primary it may result in similar patterns in visual processing, although the visual processing of single words is more holistic than auditory processing. There are a number of consequences of the position of affixes relative to their stems.

Two linguistic observations can be understood as the result of the ordering of the prefix vs. the suffix relative to the stem: Firstly, the preference for suffixation across the languages of the world (Cutler, Hawkins and Gilligan, 1985) makes sense from a processing point of view: stems of complex words generally carry more information than affixes, hence affixed words in which the stem is encountered first are in many ways easier to process. Secondly, derivational prefixes tend to change the meaning of the stem, while derivational suffixes often only change the word class of the stem, i.e. a more semantic vs. a more grammatical function (cf. for instance Chalker and Weiner, 1994, pp. 309 and 384). Modification of the stem meaning may be easier to process when the semantic modifier precedes the stem to be modified. Relatedly, in Danish and many other Indoeuropean languages, inflection is exclusively suffixing, while derivational affixes can be either prefixes or suffixes.

In experimental studies, the stem is more salient in suffixed than in prefixed words. For instance, Colé, Beauvillain and Segui (1989) report cumulative root frequency for suffixed but not for prefixed derived words, supporting the idea that the stem is more prominent in suffixed words. This is in line with the argument of Marslen-Wilson et al. (1994) that in spoken word recognition, stem-based lexical access is costly for prefixed but not for suffixed words. Meunier and Segui (2002) found that priming effects for suffixed but not for prefixed words were affected by the allomorphy of the stem. Conversely, prefixes may be more salient than suffixes: Giraudo and Grainger (2003a) document affix priming effects for prefixes but not for suffixes. Laudanna, Burani and Cermele (1994), Laudanna and Burani (1995) and Wurm (1997) report effects of prefix likelihood (i.e. the percentage of words with a particular letter string as their onset in which that string functions as a real prefix), while effects of suffix likelihood are less likely and, to my knowledge, have never been reported.

Since most stems are less frequent than most affixes, the fact that the stem is encountered first may result in the activation of smaller competition cohorts for suffixed than for prefixed words, where frequent prefixes activate very large competitor sets. This difference may contribute to making suffixed words easier to recognise than prefixed words (Baayen et al., 2007). The earlier availability in suffixed words of the most informative constituent, namely the stem, may also be involved in this advantage for suffixed over prefixed words.

The difference between prefixed and suffixed words is central to the Stem-Affix Model of morphological processing proposed by Marslen-Wilson et al. (1994), based on a series of cross-modal priming experiments: Marslen-Wilson et al. found significant morphological priming effects for all transparent prime-target combinations of prefixed, suffixed and stem forms, except pairs of suffixed words. They argue that their cross-modal priming task taps representations in the central mental lexicon and hypothesise that for semantically transparent bimorphemic words, these representations are organised as affixes linked to a stem. Generally, repetitions of the stem result in priming effects because the stem has been preactivated when the target is encountered. When both prime and target are suffixed, the recognition of one suffixed word results in inhibition of other suffixes that are linked to the same stem; this inhibition lingers and cancels out the facilitatory effect of re-activation of the stem. A similar suffix-suffix interference is observed in other studies by Marslen-Wilson and colleagues in English (Marslen-Wilson and Zhou, 1999) and Polish (Reid and Marslen-Wilson, 2003), but not by other researchers (Giraudo and Grainger, 2001; Meunier and Segui, 2002; Gonnerman et al., 2007). Gonnerman et al. (2007) argue that the absence of priming between suffixed word pairs in the experiment of Marslen-Wilson et al. is due to lower semantic transparency of those pairs. Feldman and Larabee (2001) argue that the differences between prefixed and suffixed words reflect modality differences rather than central lexical representations. These results severely challenge the Stem-Affix Model of Marslen-Wilson et al. (1994). The model's binary distinction for morphologically complex words between those that are semantically transparent and represented as morphemes and those that are opaque and represented as whole-words is also problematic in the light of the graded effects of semantic transparency observed by Wurm (1997), Baayen et al. (2007) and Gonnerman et al. (2007), see section 2.3.3 above.

2.7 Conclusions

In a sense, effects of formal similarity and morphological relatedness are conceptually alike because both reflect the activation of words that are similar to a target word. However, the effects of the two types of similarity are fundamentally different: Effects of formal similarity are generally inhibitory in word recognition, because of the activation of the target-incongruent semantics of formally similar words. In contrast, morphological relatives support the recognition of the target, both in primed and non-primed tasks, through the activation of shared form-meaning mappings. The CRUP of Wurm (1997) involves both formal and morphological similarity, as does the Complex UP introduced in this thesis: uniqueness points are based on onset-aligned formal similarity but the competitors measured by the CRUP and the Complex UP are also morphologically related to the target.

The evidence that the morphological structure of target words and their morphological relatives in the mental lexicon affect word recognition is varied: Morphological priming effects have a different time-course than pure formal or pure semantic priming and are generally also stronger. Stem frequency effects are not systematic enough to merit their traditional role as the hallmark of morphemic processing, but they are often found and provide evidence that the morphological structure of complex words affect recognition. Various morphological family effects show the influence of morphological paradigms on processing: both simple and complex targets are recognised more easily if they have many morphological relatives. Differences between simple and complex words reflect a sensitivity to the morphological structure of complex words, and advantages for complex over simple words may demonstrate the functionality of morphology.

The morphemic effects on processing co-exist with effects of whole-word information, most clearly exemplified by frequencies: both constituent and whole-word frequency effects are observed, and whole-word frequency effects may be interpreted in terms of both whole-word string probability and combinatorial probability of the constituents. The interactions between different frequency effects observed in recent regression studies, particularly the eye-movement studies of Kuperman et al. (2008a) and Kuperman et al. (2008b), also provide clear evidence of the simultaneous activation of morphemic and whole-word information. The frequency interactions also indicate that different processing routes do not operate independently but interact. Such interaction also seems to be the only way that graded effects of semantic transparency can be accounted for within a dual-route framework: if the processing routes operate independently of each other, and if recognition is achieved either through the morphemic route or through the whole-word route, it is difficult to see how graded effects can emerge. If, on the other hand, the routes interact, the morphemic route may be more or less involved depending on for instance semantic transparency, and effects may become graded. Interaction between the routes is included in the dual-route models of Wurm (1997) and Pollatsek et al. (2003).

Both the simultaneity and the interaction of morphemic and whole-word information challenge models that operate with sequential processing of these different levels of structure. This holds both for decompositional models that operate with sublexical morphemic representations (i.e. Taft, 1991, 1994) and models that operate with supralexical representations (Giraudo and Grainger, 2000), as well as for the consecutive recognition routes in the AAM-model of Caramazza et al. (1988). The completely continuous, non-morphological processing of auditory stimuli posited by the Cohort Model (Marslen-Wilson, 1984, 1987) is also untenable in view of this evidence, but the basic insight of the Cohort Model, that lexical candidates are activated and processed in parallel, remains valid, as evidenced by the UP-effects.

Both the interactivity and gradedness of morphological effects make it difficult to account for morphological processing in terms of discrete morphemes: if a complex word is either recognised via the whole word or through its constituent morphemes, graded effects and interactions between morphemic and whole-word measures should not arise. Instead, these effects favour a probabilistic account of processing. The probabilistic nature of processing is also evident in the effects of derivational and inflectional entropy and the importance of a target word's relative frequency rank in its family, discussed in section 2.5. As argued in section 1.3, the problems with discrete morphemes do not rule out a role for morphology. Rather than effects of discrete morpheme representations that may or may not be activated, effects of morphological paradigms are observed, both in the broad sense of morphological families and in the more classic sense of inflectional paradigms. Specific morpheme representations may be established, but given the strong effects of morphological paradigms and of whole-word frequency, they are likely to depend on the whole-word exemplars in memory (Hay and Baayen, 2005). The term morphemic route may refer to specific morphemic representations but also to the use of morphemic information in processing.

The experimental phenomena summarised in this chapter support an interactive dual- or multi-route model. Kuperman and colleagues use information-theoretical concepts in their PROMISE-model of visual word recognition. The information-theoretical concepts are useful for understanding both the subtle probabilistic effects of paradigm structure and the interactions between different continuous predictors, such as different frequencies. The Convergence Account, according to which morphological effects are an epiphenomenon of overlapping form and semantics, can also account for the graded effects of morphological structure, but it is unclear how it accounts for the different interactions and the detailed paradigmatic effects. The fact that morphological structure is graded does not necessarily mean that it is epiphenomenal. As argued by Kuperman (2008), the time-course of activation of the different types of information may help to distinguish between the different accounts, but the time-course is not a concern here, partly because of the focus on auditory word recognition, in which different stages are difficult to measure (though see Magnuson et al., 2007).

The experiments reported below build on this understanding of morphological processing as graded and probabilistic. Crucially, the use of regression techniques allows the investigation of graded and interaction effects, of semantic transparency, affix productivity and word and constituent frequencies. The two new UPs introduced in section 2.3.2 measure the effects of both morphologically related and unrelated competitors on the recognition of morphologically complex words. The UPs are based on the notion that processing of one morpheme may be conditional on the presence of another, inspired by the CRUP-construct of Wurm (1997). The same principle of conditional processing is used to interpret interactions between different frequencies.

Chapter 3

Extraction of lexical variables

3.1 Outline

This chapter describes general characteristics of the stimulus items that were used in the experiments reported below. First, section 3.2 gives a brief description of Danish morphology and of other relevant aspects of the Danish language. The two next sections focus on corpus-based lexical variables that were extracted for the items in order to index different aspects of word recognition: Section 3.3 describes the corpus used and the substantial modifications to it that were necessary in order to extract the relevant variables. Section 3.4 describes the individual variables that were relevant to all experiments, including different type and token frequencies, uniqueness points and neighbourhood measures.

3.2 Danish morphology

Danish is a North Germanic language with a relatively simple inflectional system. Finite verbs are inflected for present or past tense; non-finite verb forms include the infinitive and present and past participles. Passive voice may be expressed through inflection (the *-es* used in Experiment 2) or with an auxiliary. Experiment 2 also includes both types of participles (present participles ending in *-ende* and past participles in *-et*) and the default past tense suffix *-ede*. Another past tense inflectional suffix also exists in Danish, as do various irregular inflections by ablaut. Danish verbs do not carry person-number agreement. Nouns inflect for number and definiteness. The definiteness marker on unmodified singular nouns differs between neuter (*-et*) and non-neuter (*-en*); when the same morpheme occurs as a separate article before the word, it marks indefiniteness. Adjectives are also inflected for number and gender, and for definiteness and syntactic position such that attributive uses in definite noun phrases take one inflectional form of the adjective and attributive uses in non-definite NPs and predicative uses take another form of the adjective. Comparative and superlative forms may be inflectional (the superlative suffix *-est* occurs in Experiment 2) or periphrastic. Danish has productive derivation, with word class preserving prefixes and with suffixes that often but not always change the word class. However, new lexical items are most often formed by compounding which is extremely productive in Danish, resulting in very large morphological families (see section 3.4.3).

As argued by Basbøll (2003, see also Basbøll, 2005, chapters 13 and 15), the roots of some Danish verbs do not have the canonical structure of prosodic words. Therefore, the citation form for verbs is the infinitive, which for consonant-final verbal roots consists of a root and a schwa, e.g. *at kigg-e* ('to look'), while for vowel-final verbal roots, the infinitive is the root, e.g. *slå* ('to hit'). Consonant-final verbal roots are only used as relatively rare imperative forms, e.g. *kig* ('look'). The schwa in the infinitive is prosodically motivated. The verb forms categorised as "simple" in the experiments are the infinitive citation forms; with one exception, *slå* ('to hit') in Experiment 2, they all carry this prosodically motivated schwa.

Two phonological factors may differ between complex words and their stems as independent words, namely stress and *stød*: Prefixes almost always assume primary stress in derived words, changing the stress of the stem relative to the independent word. In contrast, most suffixes are unstressed and the stems retain their usual stress in suffixed derivations and inflections. An exception is the suffix *-eri* ('-ery'), derived forms with this affix have primary stress on the last syllable. Danish, like other Germanic languages, is described as stress-timed (Grønnum, 2005), and unstressed syllables may be phonologically reduced.

Stød is usually translated as glottal stop, and it is transcribed with the glottal stop symbol (ʔ) in the Danish examples. *Stød* is more accurately described as creaky voice from approximately the middle of a syllable (Fischer-Jørgensen, 1989; Grønnum and Basbøll, 2001). It appears only in stressed, bimoraic syllables, i.e. syllables that contain either a short vowel followed by a sonorant consonant or a long vowel (Basbøll, 2003). *Stød* is a distinctive feature in Danish, and therefore understood as differentiating lexical candidates in the cohorts on which the uniqueness point calculations are based. *Stød* in simple words frequently disappears when inflectional or derivational suffixes are added. These variations are regular (Hansen, 1943) and complex words with such *stød*-variations are included in the experiments. There are also cases in which suffixed words acquire *stød*, while stems of prefixed words usually have the same *stød* as their stems as independent words. Basbøll (1998, 2001, 2003) argues that *stød* applies post-lexically to bimoraic syllables, whereas absence of *stød* on bimoraic syllables is lexically specified, either individually or in more general domains of word structure. Both *stød* and primary stress are considered in the calculations of uniqueness points.

3.3 Korpus90/2000

Two corpora of written Danish were used as the source for the distributional lexical variables: *Korpus 90* and *Korpus 2000* (Det Danske Sprog- og Litteraturselskab 1982-93, 1998-2002, henceforth *Korpus90/2000*). These corpora were chosen because they are relatively large and recent, and because they consist of texts from a wide variety of sources including newspapers, magazines, web pages and fiction as well as private texts such as letters and diaries. The corpora were collected mainly for lexicographic purposes and in order to track the development of the Danish language. The texts for

Korpus 90 were collected between 1983 and 1992 and those for *Korpus 2000* between 1998 and 2002. Because of the broad range of different texts in the corpora, the frequency counts should be more representative of the mental lexicons of Danish speakers than other Danish corpora, such as Berlingske's newspaper corpus (1999) or the transcripts of parliamentary debates (see http://visl.sdu.dk/corpus_linguistics.html).

The existing corpora of spoken Danish (e.g. *BySoc*, a sociolinguistic corpus) are relatively small and none of them are phonologically transcribed or tagged in a way that would make it possible to extract spoken frequencies, diphone frequencies or phonological neighbourhoods for the Danish stimuli. This is regrettable, but to be expected for a small language. Measures extracted from corpora of written Danish must serve as indices of the corresponding auditory measures, as described in section 3.4, although the relatively deep Danish orthography makes this problematic.

The downloadable versions of *Korpus 90* and *Korpus 2000* together make up 43.6 million words. They come automatically parsed and tagged by Eckhard Bick's DanPars-tagger with an accuracy [1] of 99% for part-of-speech-tagging and lemmatisation (E. Bick p.c. August 2006). The tagging includes tags for part of speech, grammatical function, some semantic markers and phenomena such as tense and definiteness which are expressed by inflection. Crucially, the morphological structure of derived forms and compounds is not marked in the corpora (except that pluses occasionally occur at morpheme boundaries in the lemmatisations, but only in a minority of relevant lemmas and in an apparently completely unsystematic way). The web interfaces available for querying the corpora did not allow efficient extraction of the necessary variables relating to the morphological structure. This holds both for the the interface at Det Danske Sprog- og Litteratur Selskab (the Danish Society for Language and Literature), the collectors and maintainers of both corpora, and Bick's interface at the University of Southern Denmark. Therefore, the corpora had to be downloaded and adjusted, so that constituent frequencies and morphological families could be determined. In the first place, it was attempted to search and extract information from the downloaded corpora in their raw form, using regular expression tools as implemented in the command line utilities in GNU/Linux. However, this required relatively complex regular expressions, and produced relatively unclear output. Instead, the corpora were parsed into a postgreSQL database which was accessed and queried using the database management software pgAdminIII (version 1.2.1, 2005). This solution allowed structuring of the data for searches using Structured Query Language (SQL).

[1]F-score accuracy which combines the accuracy of recall (the ratio of correct answers to possible correct answers) with precision accuracy (the ratio of correct answers to total answers).

To construct this database, the downloaded text-files were concatenated to a single file. An application using the programming language Perl was written in co-operation with Kristoffer Winther Balling to extract word forms and lemmas from the text file, to assemble sentences from the file so that the context of each word form could be checked when necessary, and to store the tags. Various non-alphanumeric characters were replaced because they function as word boundaries in SQL and therefore disrupt database queries: All punctuation marks (prefixed with '$' in the file) and word forms with pluses in them (just under 1300 in all) were removed. Equal-signs in both word forms and lemmas were replaced with spaces, since they indicate the spaces in entities that the DanPars-tagger categorises as multiword lemmas, but which must be separate lemmas to extract the variables described below. Discarded items were written to a separate file, to keep track of the modifications and removals. The modifications were either necessary for the automatic search or substantially eased automatic extraction and manual sorting. Manual sorting was required because of the absence of complete morphological parsing, and because many of the affixes and stems were homographic with letter strings with different functions and meanings. Within the database, different tables were constructed for different purposes.

A word form table was constructed which contained all the word forms of *Korpus90/2000*. In this table, identical tokens were collapsed such that each unique combination of word, lemma, tags and what seemed to be meaning identifiers (used to distinguish between the different senses of some homonymous words) occurred only once along with the frequency of this combination in the corpus. This collapsing meant that whenever the table was queried for a certain word form, the output would be only one line, instead of individual lines for each individual occurrence, provided that the tagging and lemmatisation was identical for all occurrences of that word form in the corpus. The word form table was used to extract stem, whole-word and bigram frequencies and to determine UP-phonemes as described below. For the word and constituent frequency counts, word forms with no lemmatisations were excluded.

A second table contained lemmatisations of content words, listing all unique noun, verb, adjective and adverb lemmas. This was used to obtain morphological family sizes and affix frequencies for which only content words are relevant. This table included separate entries for lemmas rather than for word forms, because the family and affix frequency measures are based on lemmas, i.e. counting all inflectional forms of a relevant word as one type. Again, instead of listing all tokens, identical lemmas were collapsed and their number of occurrences counted, substantially improving the speed of queries. Lemmas without word forms and word forms without lemmas were removed from this table, these constituted less than 0.04% of tokens and the overwhelming majority of them were symbols and abbreviations that were irrelevant for the lexical variables. Queries of this table were based on a column where plusses and hyphens, which occur in the lemmatisations in an unsystematic way, had been removed, because these function as word boundary operators in regular expression searches and thus disrupt queries that make use of word boundaries (primarily the affix searches, see section 3.4.2 below).

Finally, a third table was constructed which contained the word forms as in the word form table, along with their associated lemmatisation and tags and the sentence context in which they occurred. The purpose of this table was to allow checking of ambiguous lemmas and word forms in their contexts. Technically, this was not a separate table but an index combining the information of several tables, so that their respective kinds of information could be extracted in one search.

3.4 Lexical variables

3.4.1 Frequencies

This section first establishes the terminology used for different frequencies, then describes how these were extracted from *Korpus90/2000*.

For the constituent and word frequencies, two distinctions are relevant, though they overlap to some extent: The primary focus is on the relative roles of the frequency of the whole complex word, here termed whole-word frequency, and of the constituent frequencies, the stem and affix frequencies. Whole-word frequency is elsewhere referred to as full-form or surface frequency, but I prefer to emphasise the fact that this is the frequency of the whole complex word, by using the modifier 'whole-word'. Stem frequency is sometimes known as base frequency, but 'base' can also refer to the base form of any word, simple or complex, and is therefore not used for the frequency of the stem here.

The second distinction is between the base form frequency and the frequency of the lemma, i.e. the frequency of all inflectional forms. This distinction applies both to stem and whole-word frequencies. For instance, for the derived word *accept-ere* ('to accept', noun accept plus verbalising suffix), the stem frequency could be the frequency of the base form *accept* as a free form or the frequency of all inflectional forms of *accept*. Similarly, the whole-word frequency of *accept-ere* could be the frequency of the base form *accept-ere* or the frequency of all inflectional forms of *accept-ere*, the lemma frequency. This difference is relevant to the derived words and compounds only, since for inflected words, the lemma-frequency of the whole inflected word corresponds to the lemma frequency of the stem. Both measures were extracted, but in the analyses, the measure used was the whole-word frequency of the uninflected base form, because this is the form actually presented in the experiments, and because this measure assumes the least about the organisation of the mental lexicon. In contrast, stem frequency seems to be a stronger predictor when the lemma rather than the base word frequency is used, see section 5.3.3. The lemma frequencies included all words categorised as inflectional forms of the given lemma in *Korpus90/2000*.

For the affix frequencies, this distinction is not relevant, since the affixes generally cannot occur as independent words. The affix frequency instead distinguishes between type and token frequency, see section 3.4.2.

Whole-word and stem frequencies were extracted from the word form table in the same way: the table was queried for entries with the lemma of the whole word or the stem. The output was then manually sorted to exclude incorrect lemmatisations, and the frequencies of all inflectional forms were added to produce the lemma frequency for either stem or whole word. The base form frequency of the relevant whole word or stem could be directly obtained from the same output.

In addition to the whole-word and constituent frequencies, a more control-oriented frequency variable is the mean frequency of an item's bigrams, the letter pairs that make up the word. For example, for the simple word *lyng* ('heather'), the mean of the frequencies of the letter pairs 'ly', 'yn' and 'ng' in *Korpus90/2000* is the mean bigram frequency. Mean bigram or trigram frequency is often controlled in visual lexical decision studies (for instance Baayen et al., 1997 and Lehtonen et al., 2007), though it generally fails to show an effect (Balota et al., 2004). The mean frequency of the bigrams in each of the experimental items was determined here in order to make sure that this would not be a confounding factor. Moreover, the frequency of the bigram straddling the juncture between constituent morphemes of complex words could be a relevant predictor (Hay, 2002). In auditory processing, diphone frequencies would be preferable, both for the mean and for the juncture, but since no spoken corpus of the appropriate size and nature is available for Danish, the bigram frequencies were included as the best available index of diphone frequencies in the analyses of the auditory data. This is problematic given the depth of Danish orthography but there were no other options.

The bigram frequencies were extracted from the word form table using Perl-applications written in co-operation with Kristoffer Winther Balling. A first application produced a list of all 29 by 29 bigrams in Danish (there are 29 letters in the Danish alphabet), then counted the number of occurrences of each bigram in the words of the corpus, based on the word form table. The 841 bigram frequencies were then used as a table from which the frequency of all the bigrams in all the experimental items were extracted using a second Perl-script. Then, the juncture bigram frequency for each word was noted, and the mean of the bigrams in each item calculated.

3.4.2 Affix frequency and productivity

Two frequency-based measures for the affixes were investigated in the experiments reported below. One was affix frequency, regarded as a measure of morphemic processing, supplementary to or instead of the potentially problematic stem frequency. The other measure uses ratios between different frequencies of the same affix to estimate its productivity. Additionally, it was attempted to index productivity through a questionnaire, as also described in this section.

The possible affix frequency measures are the type frequency, which is comparable to the morphological family size, and the token frequency, which is more similar to stem frequency. Both were investigated in the analyses of the experiments.

The raw token frequency of an affix does not necessarily index the productivity of the affix, mainly because very frequent affixes can be frozen and relatively non-productive. Instead, Baayen (1992, 1993) suggests two lexical statistical measures which gauge the productivity of an affix: The first measure, P*, is based on the number of hapax legomena (or hapaxes, words that occur only once in a corpus) that carry a given affix; P* is the ratio of hapaxes with a given affix to hapaxes in the corpus in general. It quantifies the contribution of the given affix to the growth of the vocabulary. In *Korpus90/2000*, the total number of hapaxes (732,105) is likely to be somewhat overestimated, since spelling errors, misparsings and wrong lemmatisations do exist and are likely to be hapaxes. However, this is not relevant for the comparison between different counts from the same corpus, only when comparing with other corpora. The second productivity measure, P, draws in both overall and hapax frequency of the affix, by dividing the hapax frequency of an affix with its token frequency. P estimates the probability that a newly sampled token with the affix represents an unseen type.

Both P and P* are based on the assumption that a substantial portion of the hapaxes in a corpus are new words, though some hapaxes in a corpus will be old words in the process of falling out of use or terms that are rare for other reasons. A more problematic assumption is that the contribution of new words to the total number of hapaxes for a given affix is on the same order for different affixes. Affix productivity is difficult to quantify (Bauer, 2001), but P and P* have the advantage of being corpus- rather than dictionary-based, which arguably makes them more representative of language use and at least makes them more parallel to the other corpus-based measures, primarily frequencies. As such, P and P* are probably the best quantitative measures of this elusive quality.

The productivity measures are based on cases where the affix has been added in the last derivational 'cycle', on the rationale that what is relevant to the productivity of an affix are cases where the affix is a constituent of the outermost layer of morphological structure, rather than cases where it is part of a stem which in turn has undergone further derivation or compounding. While the inflectional suffixes by definition occur in the outermost layer if they are word-final, word-final occurrence is not a sufficient criterion for the derivational suffixes, which were therefore manually sorted as described below. In order not to have to manually sort the words carrying the affixes both for the productivity measures and for the affix frequency measures, the affix frequencies were also based on words in which the affix occurred in the outermost layer of morphological structure. Aside from the fact that the alternative was impracticable, this also makes the measures more comparable across inflectional and derivational forms.

For the inflectional suffixes (used in Experiment 2, see chapter 6 and Appendices B and D), the word form table was queried for all words that contained the affix string in final position and were marked with tags corresponding to the function of the given affix. In some cases, a number of irrelevant tags meant that many types were counted more than once and new subtables were constructed which retained only relevant tags, listing each word form only once and thus providing reliable type counts. The large number of types carrying these inflectional endings (over 100,000 types for the most frequent affix) ruled out manual cleaning, but with an F-score accuracy of the automatic parsing of 99%, this is acceptable. Moreover, in contrast to derivational affixes, word-final inflectional suffixes by definition occur in the outermost layer of morphological structure which is relevant to the productivity calculations.

For the derivational affixes (used in all experiments, listed in Appendix D), the process of obtaining affix frequencies and corpus-based productivity measures was much more complex because derivational structure was not tagged in *Korpus90/2000*. The lack of derivational parsing meant that the searches for the affixes were simply queries for words containing the letter strings of the affixes. The outputs of these queries had to be sorted in various ways, making it a very time-consuming process to obtain the affix frequencies and productivity measures. Over 90,000 word types were manually sorted in the following way:

In the first place, relatively productive affixes were chosen for investigation, on the assumption that morphological effects would be clearer for items carrying such affixes. This assessment of productivity was mainly performed impressionistically but with the help of several grammars of Danish (Hansen, 1967; Allan, Holmes and Lundskær-Nielsen, 1995; Arndt, 1997) and a dictionary of new words in Danish (Jarvad, 1999).

The lemma table was queried for lemmas where the prefixes occurred immediately following a word-boundary and the suffixes immediately preceding one. The output of these queries were then manually sorted, removing spelling errors and foreign words (Danish texts occasionally contain words in languages that are assumed to be familiar to the readers), accidental string matches (non-affix uses of the same letter string) and cases where the affix did not occur in the outermost layer of morphological structure. Any cases of proper nouns and abbreviations in the words were also removed, but this principle was mostly relevant to the morphological families, see section 3.4.3. When necessary the context of the word was checked.

The distinction between affix and non-affix uses is not always straightforward. From a linguistic point of view, a diachronic criterion can be applied, but the psycholinguistic relevance of this is questionable (e.g. Marslen-Wilson et al., 1994). Here, words were included if the meaning of the affix or of the stem was recognisable.

Whether an affix has been added in the last derivational cycle is not a hard and fast judgement, but a qualified estimate based on various criteria. Words like *vind-hast-ig-hed* ('wind speed', literally 'wind-speed-y-ness') were not included for the productivity of the suffix *-hed* because there is no established word *vindhastig*, but there is an existing word *hastighed*, making the structure of this polymorphemic word likely to be [*vind* + [[*hast*+*ig*] + *hed*]]. Generally, words with more than two morphemes were not counted when the relation between the affix of interest and the rest of the morphemes in the word was less transparent than the relation between the other morphemes in the word. In doubtful cases, the words were included, on the rationale that the productivity measures should be more accurate if they are inclusive rather than exclusive.

Based on these sorted lists, the number of tokens, types and hapaxes carrying each affix was established, and P and P* were calculated.

A questionnaire approach to productivity

To further investigate the productivity of the affixes chosen for Experiment 1, the lexical statistical approach was supplemented with a questionnaire that obtained naturalness ratings for nonwords carrying the affixes used in Experiment 1. The stems of these nonwords were a mixture of phonotactically acceptable nonsense stems and relatively recently loaned real stems in Danish. The design of the productivity questionnaire was guided by considerations of expedience: it should both be filled in relatively fast by the participants, and it should be constructed relatively fast, as it constituted an exploratory perspective on only one variable out of several. Therefore, control and counterbalancing was not complete. The construction of this questionnaire and the information it yielded are described in this chapter, because the questionnaire-based measures showed no significant effects in the lexical decision experiments reported in chapter 5. The productivity questionnaire is reproduced as Appendix E.

To construct the productivity questionnaire, a number of monosyllabic nonwords that did not rhyme with any existing Danish words (to avoid strong associations with existing words and their derivational patterns) were taken from Balling (2002). Sentence contexts were constructed that approximately defined the meaning of each word in a monomorphemic form, these were then followed by a sentence where a derivation of the nonword occurred, carrying one of the 26 affixes that were candidates for Experiment 1 (the ones used and listed in Appendix D as well as the prefix *inter-*, which was later discarded due to scarcity of appropriate items). For example the nonstem *plaute* was first put into a sentence context in its base form *'Hun er dygtig til at* ***plaute'*** ('She is good at **plauting**'; in contrast to the English translation, the verb form required by the context is the infinitive base form), then used in the next sentence in an affixed form: *'Men derfor behøver hun vel ikke* ***overplaute*** *det hele'* ('But that doesn't mean she has to **overplaute** everything').

Additionally, derived forms of relatively recent loanwords in Danish (first occurrence after 1990, according to Jarvad, 1999), such as *mail* and *chat*, were constructed with the same affixes and inserted in a sentence context; in these cases, a defining monomorphemic use was unnecessary.

Each affix occurred once with a nonsense stem and once with a real stem. None of the derived forms were found on the Internet in a Google search. Both stems and nonstems were used because nonstems may be expected to vary less in credibility of the derivation than real stems, whereas the real words may simply be easier to accept for some participants. Both stems and nonstems occurred several times with different affixes, nonstems in a systematic way (once with a suffix, once with a prefix and once with a particle prefix plus two additional stems for the "extra" suffixes) and the stems in a non-systematic way since the construction of sentences with a derivation of an existing stem depended on the meaning and word class of that stem.

The participants were instructed to rate the derived forms, which were highlighted in the questionnaire as shown in the example above and in Appendix E, on seven-point Likert-scales, ranging from '1' indicating that the form was 'Completely unnatural' to '7' meaning 'Completely natural'.

The questionnaire was piloted with three participants, to check whether the whole rating scale was used and whether any of the words were consistently unacceptable. One item was changed after this, because pilot participants remarked on the oddity of this item and gave it much lower ratings than the word-stem derivation with the same affix.

46 volunteers completed the questionnaire; all were students at the University of Aarhus. The ratings of five participants were excluded because they had other first languages than Danish, and the ratings of one participant were discarded because the same rating ('1', unnatural) was given to all derivations, indicating that this person either had not understood the task or was reluctant to accept its premise, that nonwords vary in naturalness.

The ratings that each affix received, averaged across both real stem and non-stem uses, were included as a predictor in the analyses of lexical decision times in Experiment 1 (see chapter 5). However, neither the ratings nor the lexical statistical measures of productivity had an effect on the recognition of the complex words in this experiment. Since effects of the lexical statistical productivity measures have only been found in one lexical decision experiment (Baayen et al., 2007), this is perhaps not surprising. There are two possible interpretations of the absence of productivity effects: either affix productivity does not affect word recognition strongly enough to be a systematically significant predictor, or the existing measures — including the questionnaire-based measure introduced here — do not provide an accurate index of productivity. The fact that the affixes did not vary maximally in productivity may make the measures less likely to be significant.

Specifically with respect to the questionnaire, the effects of questionnaire-based measures in for instance Wurm (1997) and Experiment 2 (see 6.2.1) suggest that it is not the instrument itself that is flawed. However, the absence of effects of the ratings could be related to two methodological factors: Firstly, the affixes did not vary maximally in productivity. Secondly, the counter-balancing was minimal and particularly the different stems and nonstems on which the affixes occurred and their defining contexts may have affected the ratings in a way which has nothing to do with the productivity of the affixes. However, with respect to the latter point, it is unclear how this could have been controlled while keeping the questionnaire reasonably short.

In addition to investigating whether the productivity ratings had an effect on the lexical decision latencies in Experiment 1, the relation between the questionnaire ratings and the corpus-based productivity measures was explored, in order to further understand both types of measures and their relation to productivity. The productivity ratings were used as a dependent variable in a multiple regression analysis (see further section 4.4), with the corpus-based measures as independent variables. Additionally, the roles of the sex and age of the participants, the type of affix (suffixed, prefixed, particle prefixed) and the status of the stem (real vs. nonsense stem) were included.

This analysis showed that the token frequency of an affix did not significantly predict the naturalness ratings for that affix. In contrast, the number of hapaxes with the given affix did predict the ratings it received: nonwords carrying affixes with more hapaxes were rated as more natural than those with affixes that had fewer hapaxes in the corpus. The hapax frequency was highly correlated with type frequency ($R^2 = 0.81$), with which it could be replaced in the analysis, but the hapax frequency was the stronger predictor of the two. The same holds for the P measure of productivity: being based partly on hapax frequency, P is somewhat correlated with hapax frequency ($R^2 = 0.36$), it can also replace hapax frequency in the analysis, but is clearly a weaker predictor. Assuming that the ratings do at least to some extent reflect the productivity of the affix, this suggests that hapax and type frequency are better measures of productivity than token frequency is. The other derived measure of productivity, P*, is essentially identical to the hapax frequency, except that it relates it to the total number of hapaxes in the corpus, something which is mostly relevant for comparison with other corpora and languages.

An important factor to control was the possible differences between ratings of the words with real stems vs. the words with the non-stems. For particle prefixes, there was no effect of the difference between real and nonsense stems, while for suffixed words the items with real stems were rated as being more natural than those with nonsense stems and vice versa for the prefixed words. The advantage for real stems for suffixed words is probably a result of the stem being word-initial and hence more prominent in suffixed words than in both types of prefixed words. It is unclear why the prefixed items should be rated as more natural when they contain non-stems, while the absence of an effect for the particle prefixed words seems to suggest that the particle (which can also function as an independent word) is a more prominent constituent than either suffixes or prefixes. In all three cases, the result could be an artefact of the specific real stems used with the words, since this could not easily be counterbalanced, due to the semantic constraints on the real stems.

There was also a difference between male and female participants in terms of real vs. nonsense stems: males rated items with real stems significantly higher than those with non-stems, while the female participants showed no such difference. This seems to indicate that the females were more willing to accept the items with nonstems as possible words.

In sum, the explorations of productivity, using both corpus- and questionnaire-based methods seem to suggest that affix hapax and type frequency counts are better indexes of productivity than token frequency. However, the absence of productivity effects in the lexical decision task suggests either that none of the measures developed really capture productivity or that productivity does not reliably affect on-line word recognition.

3.4.3 Morphological family

A word's morphological family consists of all the derivations and compounds that contain the stem of the word in question, i.e. the whole word in case of simple words. As described in section 2.5, the type count of the morphological family, its size, has a facilitatory effect on word recognition such that words with large families are recognised faster. Generally, there are no effects of the corresponding token count, the family frequency, over and above the effects of family size. In Danish, the morphological families are relatively large, cf. comparison below.

To obtain morphological family measures, both size and frequency, the lemma table (containing unique lemmatisations of all content words) was queried for all lemmas that contained the letter string of the stem in question. Stem allomorphs were also extracted, i.e. for *mis-greb* ('mistake'), the family included both *greb* and *grib*. The output of each query was then manually sorted to exclude spelling errors, abbreviations inside lemmas and accidental string matches, as for the affix frequency counts. Moreover, proper nouns were not counted, both because this seems to be the practice in the field (Baayen p.c.), because the connection between several morphemes when one is a proper noun is generally relatively loose, and because they are unsystematically lemmatised. Items with many proper nouns in their families were avoided where possible in order to minimise differences between the items. Finally, the counts excluded the target word as well as the stem of derived words and constituents of compounds as independent words.

The lemmatisation in *Korpus90/2000* is unsystematic with respect to whether or not linking elements are included in lemmatisations of compounds. Different lemmatisations — one with the linking element and one without — of the same compound occur frequently. For instance, a compound like *gen-brug-s-fabrik* ('recycling factory'), in which the linking element *-s-* is obligatory, is lemmatised both as *genbrug+fabrik* and as *genbrugs+fabrik* and therefore occurs twice in the raw type count. Because the type count is the most important family measure, these different lemmatisations had to be manually collapsed. This made the manual sorting considerably more time-consuming than it would otherwise have been, but it was crucial for the reliability of the family counts and for comparability to other languages where this may be less of a problem. For Experiment 1, over 125,000 lemmas were extracted by the automatic query of *Korpus90/2000*; these then had to be manually cleaned to obtain reliable morphological family counts for the 235 items. For Experiments 2 and 3, there was some repetition of stems, but otherwise the amount of manual sorting was similar, and for Experiment 3 even larger, because of the larger family sizes and because two family counts were necessary for the compounds.

From the sorted lists the morphological family size (type frequency) and frequency (token frequency) were extracted and used as predictors in the analyses of the lexical decision results.

Most morphological families, though especially the larger ones, contain a number of very low-frequency and somewhat odd words, particularly compounds such as *byg-ning-s-for-stå-else* ('building-s-understanding', family member of *bygg-eri* ('building') from Experiment 1). De Jong et al. (2000) investigated the effect of family size in visual lexical decision when family members with a frequency below various thresholds were excluded from the family. In most cases, family size was a stronger predictor when the full family size was included. The inclusion of the very low-frequency items in the morphological family counts does not imply that these are necessarily represented mentally for all participants; rather, it is assumed that the morphological family measures provide an approximation of how many words containing the item are known by the participants, and thus an index of the strength of the paradigmatic relations of the target stem in the mental lexicon. Seen in that light, the morphological family size is clearly parallel to the different measures of affix productivity, which attempt to measure the strength of the affix and its availability in word formation.

A prominent characteristic of Danish is the very extensive use of compounding which results in very large morphological families. In the experiments below, family sizes ranged between 1 and 2,103 (mean 262) in Experiment 1, between 1 and 2,165 (mean 353) in Experiment 2, and between 2 and 3,476 (mean 574) in Experiment 3, all values for the corpus of 43.6 million word tokens. Compared to the family sizes reported for Dutch and English in the literature (e.g. Moscoso del Prado Martín et al., 2004a), the family sizes for the Danish experiments are very high, although items with the very largest families were avoided to limit the necessary manual sorting. In order to assess whether this cross-linguistic difference was statistically significant, the morphological families of the items in Experiment 1 were compared to those of Dutch translation equivalents in Moscoso del Prado Martín et al. (2005). The sizes of the Dutch and Danish corpora are comparable, when adding the hapax legomena to the Dutch counts which were excluded by Moscoso del Prado Martín et al. (this addition was done by Harald Baayen). The resulting sample of overlapping items was quite small, with 18 members; nonetheless, the difference between Danish and Dutch was significant on a two-tailed Student's t-test ($t_{1,34}$ = -2.475, p = 0.024): the Danish morphological families (range 18 to 2103, mean 370) were significantly larger than the Dutch (range 15 to 261, mean 99). Since recent studies of Dutch (Tabak et al., 2005) show u-shaped family size effects with inhibition in the highest ranges, the fact that the Danish families are larger has potential implications for processing (see chapters 5 to 7).

In contrast, the Finnish items used by Moscoso del Prado Martín et al. (2004a) ranged in family size from 8 to 6029, with a mean of 620 and a median of 298, for a corpus of approximately half the size of *Korpus90/2000*. Although the Danish families are smaller than the Finnish, they are closer to the Finnish ranges and means than to the Dutch and English; the similarity between Danish and Finnish relative to Dutch and English would be based on rich compounding morphology, which both languages exhibit, while the difference probably is caused by Finnish derivation being more productive than Danish.

A further variable which can be understood both within the UP-framework and as a morphological family phenomenon is the effect of the number of morphologically related continuation forms introduced by Kemps et al. (2005a). The set of continuation forms for the simple words of Kemps et al. include both morphologically related and non-related forms though the related words are dominant, while the continuations of the prefixed words of Wurm et al. include only morphologically related words. For the items in the experiments reported below, only morphologically related continuation forms were included in the continuation measure and it is interpreted accordingly. A Perl-script was written to automatically extract the continuation forms for each item from the word form database. Word forms rather than lemmas were extracted, because the continuation count includes different inflectional forms which are collapsed in the lemma table. The output for each item was checked, and spelling errors and forms that were not morphologically related to the item were removed. Compared with the morphological family and affix frequency counts, relatively few forms were removed. From these lists, the type and token counts of the continuation forms for each item were produced, and Shannon's entropy (introduced in section 2.3.1) was calculated.

3.4.4 Uniqueness points

The uniqueness point (UP) of a constituent is defined as the phoneme at which it becomes different from all other words in the language, except continuations of the constituent itself. The UPs are based on broad phonetic transcriptions like those in section 2.3.2; these were done impressionistically, but Becker-Christensen et al. (2005) and (Hjorth, Kristensen, Lorentzen and Trap-Jensen, 2003-2005) were consulted when necessary. Purely phonemic transcriptions ignore very clear differences between pronunciations, while very fine-grained phonetic transcriptions were impracticable. Throughout, I operate with two different UPs words, as outlined in section 2.3.2: UP1 indicating where the word becomes distinguishable from unrelated competitors and the Complex UP (CUP) where the complex words deviate from related words sharing the same first constituent. In Experiment 3 (see chapter 7), the UP of the second constituent as an independent word is also drawn in.

The different UP-phonemes were determined by querying the word form table for a beginning-of-word marker and gradually larger parts of the word itself. UP1 occurred when the query returned only words that were morphologically related continuation forms of the first constituent, and the CUP when the query returned only words that were continuation forms of the whole word. So, for example for a word like *duft-ede* ('smelled', past tense), UP1 is the phoneme where *duft* deviates from other words in the corpus, except all continuations of *duft*, while the CUP is the point where the whole word *duftede* deviates from other forms in the corpus, including other continuations of *duft*, but not any continuations for *duftede* itself. For the UP of the second constituent (used in Experiment 3), the second constituent was treated as an independent word, with the UP occurring when the constituent as an independent word deviated from continuation forms. Alternative spellings of the phonemes were also checked. Both *stød* and primary stress were defined as distinguishing, the UP defined as the *stød*-phoneme and the vowel in the syllable carrying primary stress. In a few cases, the first or second constituent was homophonic with other words; in those cases, the UP was defined as the end of the constituent. Proper nouns were excluded from these as from the previous calculations.

After establishing the UP-phonemes in this way, the UPs were located in the speech signal, based on waveforms and spectrograms in the waveform editor CoolEdit 2000. The UP was defined as the middle of the time segment of the UP-phoneme, and was measured as ms from onset. In cases where the UP-phoneme was a stop, the beginning of the release noise was defined as the UP; for unreleased stops, the beginning of the closure was the UP. The UPs in ms for the words in the experiments are reported in Appendices A to C.

For the visual Experiment 1b, a purely orthographic UP1 was investigated. This was extracted in the same way as the phonological UP1, except that alternative spellings, stress and *stød* were ignored.

A somewhat related variable is word length. For the analyses of the auditory lexical decision data, this was measured as the duration of the speech signal in ms; for the visual data in Experiment 1b, length was measured as the length in letters.

3.4.5 Lexical neighbourhoods

A word's lexical neighbourhood consists of the words that can be produced by changing one letter (orthographic neighbourhood) or phoneme (phonological neighbourhood) in the target word. For Danish, the corpus resources only allows the extraction of orthographic neighbourhoods, a variable that serves as a control predictor. Given the relatively deep Danish orthography and the fact that there may be differing effects of phonological and orthographic neighbourhoods (Ziegler et al., 2003), this is not ideal, but the orthographic measure is nonetheless the best available way of controlling this factor.

To obtain neighbourhood density measures, all unique word forms were sorted into files of words with the same lengths, ranging from three, the length of the shortest item, to 15, the length of the longest item. A programme written in the programming language C (kindly provided by Harald Baayen) was run, extracting the neighbourhood density (the type count of the neighbours) for each of the words in the length-files, from which the relevant type counts could then be taken.

In the lower end of the letter length range, many strings are misspellings and abbreviations which, if counted, make the neighbourhood densities less comparable across the length range. The problematic items are generally of low frequency; more specifically, short words that only occur once in 44 million words are often nonword-like in some way, partly because hapaxes tend to be new or marginal words which in turn tend to be morphologically complex and relatively long. Therefore, a frequency threshold was applied, excluding hapaxes from the neighbourhood calculations for words of length three to seven letters. From a length of eight letters, a majority of the hapaxes do make sense and were included in the neighbourhood calculations.

The traditional measure of neighbourhood density does not say much about the processing of complex words, which tend to be relatively long and have few neighbours. Therefore, Baayen et al. (2007) introduced two additional measures: firstly, the number of words that are embedded in the target and secondly, the number of words in which the target is continuously or discontinuously embedded. Both counts included only morphologically unrelated words. An item like *ras-eri* ('rage', noun) has a number of embedded words, including for instance *se* ('to see') and *ase* ('to toil'). It is embedded in words like *fRASERIng* ('phrasing') and *industRiAliSERIng* (industrialisation). Following Baayen et al. (2007), the entropies of the frequencies of these embedded and embedding words were calculated and used as predictors in the analyses of Experiment 1 (see section 5.3). Since these orthographic measures were not even significant in the visual Experiment 1b, the measures were not included in the analyses of the auditory Experiments 2 and 3.

Chapter 4

Task, design and statistics

Parts of this chapter are also included in Laura Winther Balling (2008), A brief introduction to regression designs and mixed-effects modelling by a recent convert. *Copenhagen Studies in Language* 36, in press.

4.1 Outline

The present chapter discusses the methodological choices and considerations that were relevant to all three experiments reported in chapters 5 to 7. The choice of task is explained in section 4.2, then section 4.3 discusses the overall design of the experiments and section 4.4 the linear mixed-effects models used for the analyses and some more general statistical considerations.

4.2 Experimental task

All experiments reported in chapters 5 to 7 employed lexical decision (see also section 2.2). This choice of task was motivated by two main considerations: Firstly, lexical decision is probably the most widely used task in studies of lexical processing. This is particularly useful when investigating a language in which on-line lexical processing has not previously been studied, since the possibility of comparing the present work with previous studies is increased when the methods are the same, although the languages differ. This improves the chance of determining characteristics that are typical of the language, rather than of the task. Secondly, lexical decision allows the investigation of both form-related variables (like the UP-measures introduced above, section 2.3.2) and more semantic variables. Recent studies (e.g. Baayen et al., 2006) suggest that the strong frequency effects that are characteristic of the lexical decision task may be understood as semantic, since they reflect familiarity with particular concepts.

A criticism occasionally raised against lexical decision is that it is an artificial task (see Goldinger, 1996a): The judgement of whether a given string is a word is not part of everyday language processing. On the other hand, language processing in the laboratory is bound to be artificial to some extent, and participants understand the lexical decision task quickly and in an intuitive way. Moreover, although lexical decision is more sensitive to semantic factors and frequency, the results are in many ways comparable with other tasks such as naming (Baayen et al., 2007), gating (Wurm, 1997) and reading measured by eye-tracking (Kuperman et al., 2008b). A main difference is that naming and gating require the selection of a specific word, while lexical decision reflects more general activation. However, there are indications in the experiments reported below that participants perform relatively precise word identifications even in the lexical decision task. The possible artificiality of the task is a consideration when generalising from the results of the experiments to natural language processing. However, two aspects of the design of the present study reduce the problem: Firstly, random variation between the participants was statistically controlled (see section 4.4 below). Secondly, priming was not used as an experimental manipulation, instead naturally occurring variations like complexity type, different frequencies and UPs were used to explore lexical processing. When information about such distributional variables is available (as it is in Danish, albeit with the substantial modifications of existing resources described in chapter 3), the priming manipulation is unnecessary, except for the investigation of the very early stages of lexical processing with subliminal prime presentation. Cross-modal priming is argued by for instance Marslen-Wilson et al. (1994) to avoid a modality bias, but this may also be achieved by comparing auditory and visual processing as is done in Experiments 1a and 1b below.

4.3 Design

The design of all three experiments reported below differs from the design of most traditional studies of lexical processing by using a regression design in contrast to the traditional factorial designs. This choice of design and associated statistical analyses is motivated both negatively — by the potential problems with the alternative design option — and positively — by the advantages of regression analyses. The two are of course related, the advantages of regression designs being relative to the drawbacks of factorial designs.

Factorial designs are based on control between groups or conditions of items of everything except, in the simplest case, one variable of interest. As more and more variables have been found to affect word recognition (see review in chapter 2), this control between conditions has become increasingly difficult (as pointed out by Cutler as early as 1981). If the control between conditions is not adequate, it becomes questionable whether a between-conditions difference can be interpreted as an effect of the manipulated variable. The increasing number of potentially relevant control variables can also mean that relatively few items can be included in an experiment, considerably reducing the statistical power.

The manipulation of one variable between two conditions (again taking the simplest possible factorial design) is best suited to factors where the levels are to some degree naturally given, such as for instance different types of words. However, factorial designs are also often employed to investigate numerical variables, such as frequency, which must then be dichotomised. As shown by Cohen (1983), dichotomisation of continuous variables results in a substantial loss of statistical power (see also MacCallum, Zhang, Preacher and Rucker, 2002; Harrell, 2001). Moreover, from a linguistic point of view, the control of everything except one key variable can result in items becoming unrepresentative. For instance, stem and whole-word frequency are typically highly correlated, but a factorial design investigating one of these variables requires the dichotomisation of one and control of the other, which necessitates the choice of items for which the two are not highly correlated, i.e. items that may not be typical of words in the language.

This does not necessarily mean that the results of factorial studies are invalid, although this may be the case as demonstrated by Baayen, Levelt, Schreuder and Ernestus (2008b) in their reanalysis of the results of Baayen et al. (1997). In any case, the problems with factorial designs do indicate that alternative, more powerful techniques may be preferable. In recent years, regression designs have been emerging as such an alternative (e.g. Ford, Marslen-Wilson and Davis, 2003; Balota et al., 2004; Baayen et al., 2006). The English Lexicon Project of Balota et al. (2002) has probably furthered this development. Multiple regression techniques make it possible to assess multiple correlations of different independent or explanatory variables with the dependent variable at the same time. The regression analyses determine whether there are effects of each independent variable over and above the other independent variables included in the regression model. This logic also applies in the illustrations of the effects of the different independent variables in chapters 5 to 7: The partial effect of each variable is illustrated with all other variables held constant at their medians.

The possibility of including many different independent variables in the analyses has several benefits: Firstly, the fact that variables can be investigated instead of controlled makes multiple regression designs overwhelmingly more informative than factorial ones. For instance, the first experiment reported here (in chapter 5) simultaneously investigates the effects on lexical decision latency of complexity type, frequency, morphological family size, duration and uniqueness points, as well as a number of variables that turned out to be non-significant. Secondly, the absence of strict between-conditions control makes it possible to include many more items in the experiments, thus increasing the statistical power considerably. Thirdly, the artificial dichotomisation of graded variables is avoided, which also increases statistical power and may result in a more correct analysis (cf. Baayen et al., 2008b). Fourthly, effects of experimental context can be statistically controlled in a regression design: de Vaan, Schreuder and Baayen (2007) found that a given response was co-determined by the speed of the preceding responses. These factors cannot be experimentally controlled, but can straightforwardly be included as covariates in a multiple regression analysis and in this way statistically controlled. The correctness of the preceding responses is added as a covariate in all analyses; in Experiment 3a and 3b, the lexicality of the preceding response is included as an additional covariate, following Diependaele, Sandra and Grainger (2007).

The possibilities of investigating different lexical and other variables are not unrestricted in regression analyses. One restricting factor is the problem of collinearity between the different independent or explanatory variables. If variables are highly correlated, the unique contribution of each variable becomes impossible to assess. This collinearity restricts the questions that can be asked and, if unaddressed, can call into question the validity of the analyses. However, various remedies against such harmful collinearity are available, as discussed below (section 4.4.2). Moreover, one should be wary of outliers, which may show a stronger or weaker effect of a particular explanatory variable than is actually warranted. Both RTs and several of the corpus-based variables were therefore logarithmically transformed to reduce the skewness and avoid outliers.

4.4 Statistical models

4.4.1 Fixed and random effects

Various statistical choices exist for the analysis of data from a regression design. Of these, linear mixed-effects modelling seems to be the most powerful without being anti-conservative (see Baayen 2008a: 282-299). A statistical technique is conservative if it is likely to result in Type 2 error, i.e. not rejecting a false null hypothesis. A technique is anti-conservative if it is likely to result in Type 1 error, rejecting a null hypothesis which is in fact correct. Because Type 2 errors are cases where a significant effect comes out as non-significant and is therefore not interpreted, Type 2 errors are the least serious. Type 1 errors, where a non-significant effect is shown to be significant and therefore interpreted as a real result, are far more problematic. Therefore, while a statistical technique should not bee too conservative, it is more important that it is not anti-conservative. The linear mixed-effects technique employed in this study fulfils these requirements.

All analyses of the lexical decision results were done using the linear mixed-effects regression models available in the lme4 package (Bates and Sarkar, 2006, 2007) within the statistical computing environment R (mainly version 2.4.1, 2006). These models allow the modelling of non-linear as well as linear effects. The languageR package of Baayen (2008b) was also used.

Mixed-effects models include both random and fixed factors. Fixed factors are factors which have a fixed and low number of levels which exhaust the levels in the sampled population, such as the levels inflected, derived and compound words in the population of Danish complex words. The fixed factors are repeatable; for instance, new words can be chosen in the different categories of complex words. If a fixed factor is included in a regression model, one of the levels for the factor corresponds to the basic intercept of the model. The intercept is the value of the dependent variable when all variables in the analyses are zero; in terms of the typical illustration of effects in regression analyses, it is where the regression line crosses the vertical axis. Adjustment to the intercept for each of the other levels of the fixed factor are specified in the fixed-effects structure of the model. Fixed effects may interact with co-variates, i.e. numerical variables such as frequency, in which case the different levels of the fixed-effect factor show different slopes for the particular co-variate, e.g. frequency. The slope is the change in the dependent variable for a given change in the independent variable, i.e. the distance that the regression line covers on the vertical axis for a given change on the horizontal axis. The slope may be positive or negative and more or less steep.

Random factors are not repeatable and do not have a fixed number of levels. Typical random effects in word recognition studies are participant and item: Both participants and items are in principle sampled randomly from the relevant populations, each participant or item corresponds to a level of the variable which is not repeatable. Linear mixed-effects models can include random variation in two ways: as random intercepts and as random slopes. The random intercepts are adjustments to the basic intercept of the regression model for each level of the random factor, e.g. for slow vs. fast participants for the random factor participant. Random slopes are adjustments to the slope of a given numerical predictor for each level of the random effect. For instance, random slopes for frequency by participant are adjustments to the general slope for word frequency for each participant, i.e. for every level of the random effect 'Participant'. The random intercepts for participant, word and affix included in the analyses reported in chapters 5 to 7 take into account the random variation between participants, words and affixes; the random slopes for frequency by participant account for the difference between participants in terms of their slopes for frequency.

Fig. 4.1 illustrates the idea behind both random intercepts and slopes. The figure plots lexical decision reaction time as a function of word frequency for the participants in the visual Experiment 1b. The lines in each panel are lowess smoothers, which show the correlation between RT and frequency for each individual participant, but smoothing away some of the variance in the data to show the correlation more clearly (for the principles of this, see Baayen, 2008a: 38 and Venables and Ripley, 2002: 228-232). The different places in which the smoothers cross the vertical axis correspond to the different intercepts for the participants, which are accounted for in the statistical model as random intercepts with a mean of zero and an unknown variance. The mean of the random intercepts should be zero because the mean of the adjustments for all participants should correspond to the general intercept of the model. The precise intercept for each participant is not meant to be determined from fig. 4.1, but the figure illustrates the kind of differences betweeen participants that can be accounted for by the inclusion of random intercepts and random slopes.

Though the slopes of all the smoothers in fig. 4.1 are negative, corresponding to facilitatory effects of word frequency, the participants show clearly different profiles, with for instance participants number 1b05 and, even more so, number 1b20 showing much stronger frequency effects than the others.

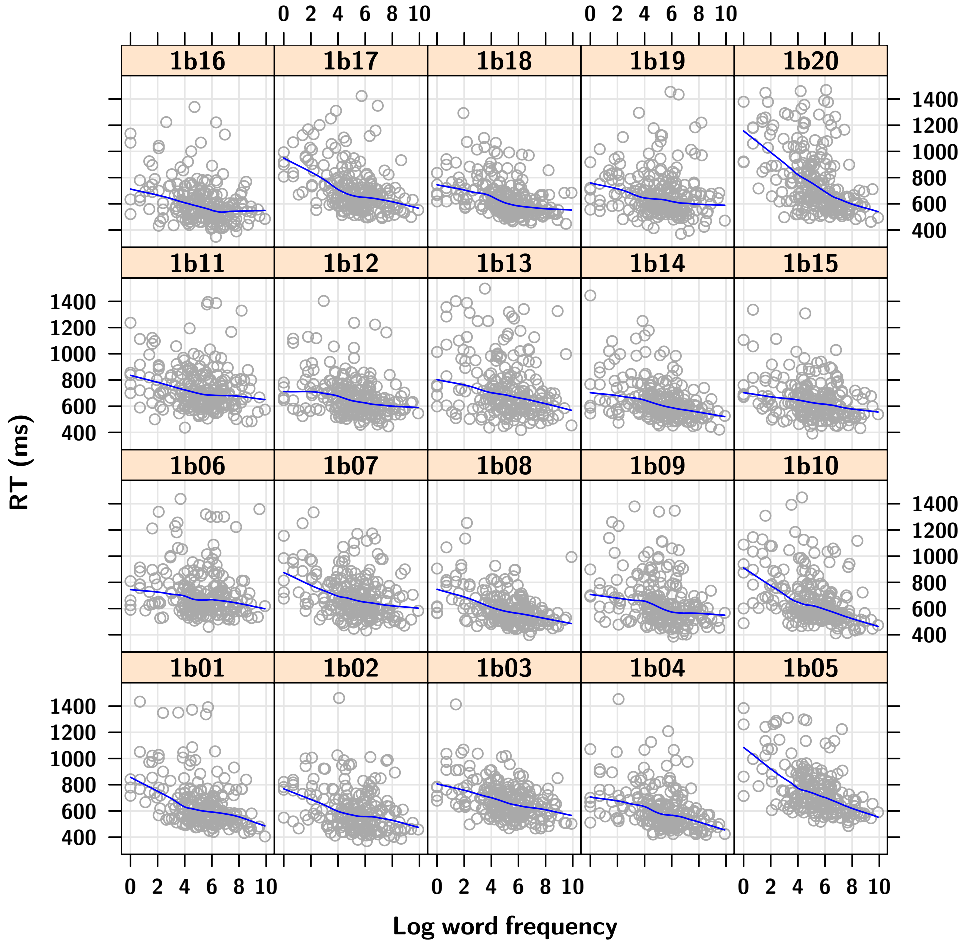

Figure 4.1: Conditioning plots of reaction time as a function of word frequency for each participant in the visual Experiment 1b (see further in chapter 5).

These differences can be modelled by random slopes for frequency for the participants. This example from Experiment 1b is particularly interesting because the initial analysis suggested an overall difference in frequency effects between males and females when the random slopes for frequency were not taken into account. Once this participant variation was accounted for by the inclusion of the random effect parameter, the sex difference disappeared. It turned out that the sex difference appeared because participant number 1b20, who showed the strongest frequency effect, happened to be male.

The example from Experiment 1b shows a major advantage of being able to include both fixed and random effects, namely that it becomes possible to assess whether group differences are significant over and above differences between individual participants. Another advantage is that a single analysis including random effects of participant and item can replace the usual separate analyses of the experimental results by participant and by item.

Whether or not a particular random effect is included in a given mixed-effects model is determined with the help of likelihood ratio tests. The model with the random effect in question is compared to the model without that random effect, in order to determine whether there is an increase in the goodness of the fit of the model to the data when the random effect is included in the model, and whether such an increase in goodness of fit justifies the loss of degrees of freedom caused by the inclusion of the random effect. The p-value of the likelihood ratio test indicates whether the difference in goodness of fit between the models is significant; if the p-value is small — indicating reason for surprise in the increase in goodness of fit given that only one parameter has been added — the random effect is justified. Although this example refers to the addition of one random effect, likelihood ratio tests may test the addition of more than one random effect parameter.

4.4.2 Collinearity

As mentioned, collinearity between the numerical predictors in the regression model is a problem, because it becomes impossible to determine which of the collinear variables are in fact predictive. As a consequence, models become unstable if collinearity is high. In order to assess the collinearity, the condition number, kappa or κ, is calculated. The condition number is a measure of how close the correlation matrix is to being singular, i.e. completely collinear. High and potentially harmful collinearity is indicated by κ-values of 30 or above (Baayen, 2008a). Many of the models show κ-values well above 100 for the correlation between the raw lexical variables, indicating that collinearity is a real problem. The specific steps taken to decorrelate the explanatory variables for each data set are described in the relevant method sections, but in general, three methods are employed:

Firstly, when the individual contributions of a cluster of correlated variables are not of central importance, a useful tool is principal components analysis (PCA, see Baayen, 2008a: 200ff). PCA transforms the variance of the original variables into principal components (PCs) which account for the same variance but are orthogonal to each other, i.e. completely uncorrelated. This technique is mainly used for the RTs to preceding items where the contributions of the individual measures (the RTs to the preceding item, the one before that and so on) are less important than the fact that this variance is included. If PCA is used in cases where the individual variables are more important, the contributions of the original variables can be gauged by inspecting the factor loadings of the original variables on the PCs, which in effect show the contribution of each original variable to each PC.

The second tool used is to take two highly correlated variables and construct linear regression models with one (the dependent variable) as a function of the other (the independent variable). For instance, in Experiment 1b, log whole-word frequency and length in letters are correlated; in order to tease apart the contributions of these two variables, a regression model is constructed with length in letters as a function of whole-word frequency. For the overall analysis, the dependent variable of the linear regression model, length in the example from Experiment 1b, is replaced with the residuals of the regression model, i.e. the variance in the dependent variable length that is not accounted for by the independent variable whole-word frequency. The correlation between the original dependent variable and the new residualised one is generally large, indicating that the residuals are a reasonable replacement. The new residualised variable is uncorrelated with the original independent variable (e.g. the residualised length is uncorrelated with whole-word frequency), so the two correlated variables in the original dataset are replaced with two uncorrelated variables, and the collinearity is reduced.

A third tool is specifically used for the measures of UP and duration in the experiments. The raw UP and duration variables are all measured from word onset, i.e. they measure overlapping parts of the auditory signal, which leads to collinearity between them. Part of this collinearity can be removed by only including the onset of the word in one of the measures. In the analyses reported in chapters 5 to 7, UP1 is measured as the distance from word onset to UP1, the Complex UP as the distance from UP1 to CUP, and word duration in ms as the distance from the Complex UP to word offset. This recalculation procedure is adopted from Baayen et al. (2007) who recalculate word duration and their single UP-measure in this way.

4.4.3 Variable selection

In regression analyses, a large number of variables can enter into an analysis, some of which may turn out be significant predictors of the observed data and some of which may not. This gives rise to a debate regarding whether the non-significant predictors should be retained in the final statistical model or not: On one hand, the statistical model becomes simpler if the number of variables is reduced and the patterns of results become much clearer, both because the model is simpler and because the inclusion of non-significant variables can sometimes obscure the effects that are really there and vice versa. On the other hand, there are potential theoretical problems with the validity of the final statistical model if a large number of models are constructed and one chosen (Harrell, 2001). In exploratory data analysis like the present, with no very clear-cut dichotomous hypotheses, the aim is to achieve the minimal model that adequately describes the data. In this case, the consensus seems to be that reducing the statistical model is acceptable (Baayen, 2008a), but this should be done in a careful and systematic way and preferably the removal of variables should be motivated by factors that are external to the statistical model, i.e. linguistic or psychological factors. Here, models including all potentially relevant predictors (the different lexical variables outlined in 3.4) were constructed and then reduced in step-wise fashion, reaching simpler and clearer models in which all predictors are significant.

Another issue to consider when selecting variables for the final analyses is the distribution of the variables. Although log transformed values for the variables are generally used, several of the variables include clear outliers. It is important to ascertain that the significance of the effects included in the final analyses is not caused by these outliers. In order to rule this out, the distributions of all significant variables were examined; if there were outliers, a new model was constructed which excluded these outliers, and the effect in question was only included in the final model if the model without the outliers also showed it to be significant.

4.4.4 Data filtering

It is standard in many studies to filter out reaction times that are more than two or three standard deviations from the person and items means. Instead of this a priori filtering, data points with large standardised residuals (outside the interval -2.5 to 2.5) in the models were excluded and the models refitted (cf. Crawley, 2002; de Vaan et al., 2007). This is both a more conservative method, in that fewer data points are removed than with the standard filtering, and one that substantially increases the model fit: Generally, the correlation between the fitted values of the regression model and the observed data in the experiments is around two percent higher when the model is trimmed by removing data points with large residuals, than when the data are filtered a priori based on subject and item means, while both in turn are considerably better than the unfiltered data. Moreover, variables that were significant because of outliers become non-significant in the models that are refitted to the trimmed data set.

Chapter 5

Experiments 1a and 1b

5.1 Introduction

The experiments presented in this chapter constitute a first exploration of on-line lexical processing in Danish, and were therefore designed to address several issues, drawing in both simple and different types of complex words, with a range of different characteristics. The complex words used in these initial experiments were derived forms. Derivation is understood as an intermediate category on a continuum of morphological types, sharing some of the regularity of inflection and some of the semantic richness of compounding (see section 2.6).

The two experiments, one auditory and one visual, also addressed a number of specific questions of which three were central:

Firstly, processing differences between the different types of words were investigated. One prediction based on the notion that morphological structure has a psychological functionality (see section 1.4) is that, other things being equal, complex words should be processed faster than simple words. Things, however, are not generally equal: Morphologically simple and complex words differ in a variety of ways, probably most noticeably in that complex words tend to be longer than simple words, but also in terms of for example morphological family size and frequency. These differences constitute one major motivation for the use of a regression design which allows statistical control of the different variables and thus permits the choice of typical words from each category, instead of words that must be matched between the categories (other, more technical reasons, for the choice of design are discussed in chapter 4).

Another question relating to the different types of words used in the experiments is whether the particle prefixed words differ from the standard prefixed and the suffixed words. What I term particle prefixes (see definition in section 2.6) are morphemes with a dual function as non-separable prefixes and independent particles or prepositions. An example is the morpheme *om* which can function as a preposition meaning 'about' or 'around', or a prefix with a meaning corresponding roughly to 'circum-' or 'sur-' in English, as in *om-egn* ('surroundings', literally 'about-region'). These particle prefixes have characteristics both of compound elements and of derivational morphemes and constitute a kind of borderline category between compounding and derivation which exhibits relatively rich semantics alongside relatively high frequency of occurrence in the prefix-like function. Particle prefixes are characteristic not only of Danish but also of other Germanic languages.

The differences between the simple and the three types of complex words are investigated in All Items Analyses which include variables that are relevant to both simple and complex words. Discovering any differences between the three types of complex words requires that morphological variables, such as constituent frequencies which are only relevant for the complex words, are also controlled; hence a second level of analysis considers only the complex words and the variables relevant to those.

Secondly, effects of lexical frequencies were explored. Word frequency is relevant for both simple and complex items, but the central comparison applies to the complex words only, namely what the contributions of whole-word vs. constituent frequencies are. The complex items were chosen with a wide range of both stem and whole-word frequencies and with different ratios between the two, since effects of relative frequency were found for English complex words by Hay (2001). Whole-word frequency effects are expected, while the question of stem frequency effects is much more open, as discussed in section 2.4. On a traditional view of stem frequency effects as characteristic particularly of low-frequency complex words (Alegre and Gordon, 1999), stem frequency effects would be predicted for the present items, which are of relatively low frequency. However, as shown by Baayen et al. (2007), very large datasets and powerful statistical methods are required in order to systematically detect stem frequency effects (see section 2.4); while larger than the datasets in most factorial studies, the present dataset may yet be too small to reveal significant stem frequency effects. In addition to the typically investigated stem frequency, the present experiments also explore whether there would be effects of the alternative morpheme frequency, that of the affix. Interactions between the different frequencies were also explored.

Thirdly, the experiments took into account morphological structure in the understanding of lexical competition by operating with the two different uniqueness points (UPs) for the complex words that were introduced in section 2.3.2. UP1 indexes the duration of the competition between the target and morphologically unrelated words. Complex UP (CUP) measures the duration of the competition between the target and morphologically related words (see section 2.3.2). This is a more fine-grained way of indexing the cohorts at different points in word recognition than the traditional single UP and should therefore result in a better understanding of lexical competition in auditory recognition of morphologically complex words.

In addition to these three main questions, the regression design allowed investigation of other variables, like morphological family size and junctural probability, to give as full a picture of the processing of derived words in Danish as possible. This is particularly relevant in a language which is underexplored in the domain of lexical processing. All relevant variables are outlined in section 5.2.

Though this thesis mainly considers the primary auditory modality, so much previous experimental work has been conducted in the visual modality that comparison between this work and many previous studies of morphological processing might not be immediately possible if only auditory word recognition was studied, since both the language and the modality would differ. Moreover, it is an important question to what extent visual and auditory processing are similar. For these reasons, the same experimental items were used in both an auditory and a visual version of the experiment.

Summing up, the experiments reported in this chapter compared the processing of morphologically simple and complex words and of different kinds of derived forms. Additionally, the roles of different frequencies and uniqueness points were considered, alongside a number of other lexical variables. A final question concerned the difference between auditory and visual processing with respect to the variables manipulated here.

5.2 Method

5.2.1 Materials

500 items were used in an auditory (Experiment 1a) and a visual (Experiment 1b) lexical decision experiment. Half of these items were words and half nonwords; of the words, 125 derived forms and 110 simple words were analysed as reported below. In addition, 15 compounds were included as fillers in order to vary the kinds of morphologically complex words encountered in the experiment, but these were not analysed. This diversity of the items was an attempt to reduce the risk of participants focusing on the relatively large number of derived words. In piloting as well as debriefing, most participants remarked on the nonwords and none mentioned the morphological complexity of the items.

For all words, it was attempted to avoid highly irregular spelling-to-sound mappings, pronunciations varying much between casual and careful speech, and homonymous or strongly polysemous words. For the simple words, letter strings that corresponded to the affixes used on the complex words were avoided. Letter strings that correspond to other affixes could not be avoided entirely. Words of different word classes were chosen: a group of nouns, a group of verbs, and a group of adjectives of which a few simple and all those carrying one specific affix (*-isk*, '-ish' or '-ic') can also function as adverbs. All words occurred in some inflectional form in *Korpus90/2000*, but in a few cases, not in the base form which was presented in the experiments. These words have a whole-word frequency of zero, although the lemma frequency of the whole word was always at least one.

For the complex words, some additional factors were taken into account: All complex items had free stems. Linking morphemes and stem allomorphy relative to the free word were generally avoided. Only bimorphemic derived forms were chosen, both as a homogeneity consideration in its own right and because semantic relations between morphemes are more complex with more than two constituents. These general restrictions were also applied in the subsequent experiments.

The affixes used were ten suffixes, seven prefixes and eight particle prefixes. Each of these was represented on five words in the experiment, thus keeping constant the number of times each affix occurred in the experiment. Most of the affixes were relatively salient in order to make possible the manual sorting that was necessary for the extraction of corpus-based productivity measures, see section 3.4.2. They were not homophonic with other affixes. The verbalising suffix *-ere* is homographic (but not homophonic) with the adjective comparative suffix, but none of the words with *-ere* were actually ambiguous. The affixes had different letters at the juncture with the stem, so as to obtain a range of different junctural probabilities. All affixes were relatively productive, with the exception of a few low-frequency affixes that were included for the sake of variation in productivity. The affixes are listed in Appendix D.

Complex items were chosen with a wide range of different stem and whole-word frequencies and different ratios between the two. Simple words generally have higher whole-word frequencies, but also for these, words with a wide range of different frequencies were chosen. Additionally, a range of other lexical variables were extracted as described in chapter 3. The semantic complexity of the whole words and of the stems was indexed by their respective numbers of dictionary meanings; this measure did not turn out to have any effect on processing latency. The complex words were all relatively semantically transparent. The different lexical variables are summarised in table 5.1, with those variables that are relevant for all items summarised for all 235 items in the upper part, while the lower part shows both morphological and other variables for the 125 complex items. The lexical variables for the individual words are listed in Appendix A.

There are two ways of constructing appropriate nonwords for a lexical decision experiment focusing on morphologically complex words: The same affixes as those used on the real words can be retained, or nonwords can be constructed that carry different affixes. The former option was chosen here, because it means that the specific affixes on the words would not in themselves be cues to wordness, although this has the drawback of doubling the number of times each affix occurs in the experiment to ten instead of five. The 250 nonwords were constructed by changing one to three phonemes in each word to form a nonword. In the case of the complex nonwords, the change was made to the stems while the affixes were retained. None of the nonwords occurred in *Korpus90/2000*.

Table 5.1: Means, standard deviations and ranges for properties of the items used in experiments 1a and 1b. Whole-word frequency refers to the frequency of the base form of the whole word, stem frequency to the lemma frequency for the stem, see section 5.3.3. The variables were converted to logarithmic scales for the statistical analyses, in order to reduce their skewness, but the non-transformed values are shown here for interpretability.

All items ($n = 235$)

	Mean	SD	Range
Whole-word frequency[a]	18	47	0 to 461
Morphological family size	262	349	1 to 2 103
Number of continuation forms	63	131	1 to 912
Number of meanings	1.9	1.4	1 to 10
Neighbourhood density	2.6	4.3	0 to 33
Mean bigram frequency[a]	27 739	16 108	2 471 to 88 334
UP1, ms	332	101	131 to 671
Length, ms	570	129	239 to 940

Complex items ($n = 125$)

	Mean	SD	Range
Whole-word frequency[a]	6	12	0 to 100
Stem frequency[a]	398	2 904	0.2 to 32 437
Affix type frequency[a]	11	13	0.4 to 51
Affix token frequency[a]	2 190	3 200	22 to 15 260
Morphological family size	341	405	1 to 2 103
Number of continuation forms	12	19	1 to 114
Number of meanings	1.3	0.7	1 to 5
Number of stem meanings	3.0	2.4	1 to 17
Neighbourhood density	0.5	0.9	0 to 4
Mean bigram frequency[a]	30 951	16 076	4 581 to 88 334
Juncture bigram frequency[a]	15 552	21 625	72 to 142 493
UP1, ms	289	85	131 to 579
Complex UP, ms	456	103	232 to 732
Length, ms	645	97	458 to 940

[a] Values for these variables are counts per million in *Korpus90/2000*.

In addition to the experimental words and nonwords, 30 words and nonwords were used for training and warm-up. These items had a similar composition to the experimental items, but carried different affixes in order not to introduce variations in the number of times each of the experimental affixes was encountered. All words and nonwords are listed in Appendix A.

5.2.2 Recording and preparation of auditory stimuli

The stimuli were read by a female native speaker of Danish who was unfamiliar with the purpose of the experiment. The stimuli were recorded on a Sony DAT-recorder (model TCD-D8), using a Sony electret condenser microphone (model EC-959a), in a sound-attenuated room. The reading lists mixed words and nonwords, to avoid overall intonational differences between the two sets of items. One reading filler was included at the beginning and two at the end of each list to avoid beginning- and end-of-list intonation on the items themselves. Seven words and seven nonwords occurred on each reading list. The speaker was instructed to read the stimuli carefully, and was told that real affixes would sometimes occur in nonwords. Hyphens and stress marks were inserted in the reading list when necessary to disambiguate pronunciation. The reading lists were digitised at a sampling rate of 22 kHz and a bit depth of 16 bit, and the words were segmented from the list and normalised for peak intensity.

5.2.3 Participants

21 volunteers (11 women and 10 men between the ages of 22 and 39, mean 26.3 years) participated in Experiment 1a, and 20 (12 women and 8 men between the ages of 21 and 38, mean 26.8 years) in Experiment 1b. Additionally, two volunteers participated in sessions that were interrupted because of the presentation programme breaking down; those data were not used. I recruited a roughly equal number of male and female participants in order to avoid that the sex of the participants might bias the results (cf. section 2.4), although males are generally harder to recruit in the female-dominated degree programmes whose students I had access to (English, linguistics and psychology). Each participant was tested separately in a sound-attenuated room. Most of the participants were students at the University of Aarhus, all were native speakers of Danish with normal hearing and normal or corrected-to-normal vision, according to self-report.

5.2.4 Procedure

Both the auditory and visual versions of the experiment were run using DMDX (Forster and Forster, 2006), on a portable computer. All appropriate preliminary tests for DMDX were run in TimeDX, using the instructions and suggestions provided in the on-line documentation for DMDX and TimeDX.

The stimuli occurred in a different pseudo-random order for each participant. The orders were generated using Mix (van Casteren, 2006), a programme which produces pseudo-random orders with certain constraints, ensuring in this case that no more than three words or three nonwords occurred in a row and that none of the affixes appeared on two consecutive trials.

In the auditory Experiment 1a, the stimuli were presented over headphones. Each stimulus was preceded by fixation point (a plus) displayed for 500 ms in the middle of the screen. The inter-stimulus interval (ISI) was set to 3000 ms (including the duration of the fixation point display), with time-out occurring 2500 ms after the beginning of each stimulus. The ISI was fixed rather than variable, because this seemed at first to give the smoothest running of the experiment. However, this set-up resulted in repeated cases of DMDX breaking down, delaying the experimental work substantially, so for Experiment 2, this part of the experimental set-up was changed. The fixed interval meant that the auditory Experiment 1a had a slower pace than the visual Experiment 1b and the auditory Experiment 2. Possible implications of this for the results of Experiment 1a are discussed in sections 5.3.1 and 5.3.4, and further addressed in Experiments 3a and 3b, see chapter 7.

In the visual Experiment 1b, each stimulus was preceded by a plus in the middle of the screen which was replaced with the stimulus centred around the same point after 500 ms. Each item was displayed for 2000 ms or until participants responded, giving an ISI which was variable but with a maximum of 2500 ms. The words were white on a black background, presented in a lower-case Courier New 18 point font on a 15 inch screen.

The time-course of the experimental sessions was as follows: on arrival, participants were instructed orally about the procedure of the experiment, then asked to sign a consent form (see Appendix F). They were then asked to adjust the volume (after hearing one of the words used as reading fillers when the stimuli were recorded), before written instructions in Danish appeared on the screen and the brief training consisting of 20 items started. The instructions ran as follows:

"You will hear/see both words and tokens that sound like words but are not. Decide as quickly and accurately as possible whether what you hear is a word in Danish or not. Press YES if it is a word, and NO if it is not. First, you will receive a brief training, then you can ask questions. Press the space bar when you are ready." (translated from the Danish original)

Participants pressed a button marked "YES" with their dominant hand, a button marked "NO" with the other hand. Handedness was determined by asking the participants whether they considered themselves right- or left-handed.

It was stressed that the participants should judge according to what they themselves recognised as words, rather than what could conceivably be a word in Danish. After the training, participants could ask questions (very few did) before the experiment itself started. Two breaks were inserted one third and two thirds through the 500 items; six warm-up items appeared at the beginning of the experiment and two after each break. The auditory experiment lasted about 30 minutes, the visual about 15 minutes.

Three pilot participants were tested for each experiment; the only adjustments made after piloting were the use of fixation points also in the auditory version, and the insertion of two rather than one break.

After the experiment, participants were asked to fill in a background questionnaire with questions about their language background. Information about sex, age and handedness was also requested. The questionnaire is reproduced in Appendix F.

5.3 Results and discussion

Errors constituted 4.3% of the auditory and 5.9% of the visual lexical decision responses; these were excluded from the RT analyses. Additionally, due to error rates over 30%, the responses to two simple and four complex items were removed for the analyses of the auditory RT data; these were *dølge* ('to hide'), *smæld* ('snap' or 'crack'), *efter-skælv* ('after-quake'), *sam-sende* ('send together' or 'broadcast together'), *van-held* ('mis-luck') and *van-skæbne* ('mis-fate'). The responses to one simple and five complex items were removed for the visual RT-analyses; these were *dølge*, *asket-isk* ('ascet-ic'), *gerrig-hed* ('miserly-ness'), *om-serv* ('re-service'), *sam-sende* and *van-held*. All in all, 5.4% of responses were removed for the auditory analyses and 7.0% of responses for the visual analyses. The mean correct RTs were 991 ms for auditory and 687 ms for visual lexical decision, measured from stimulus onset.

RT (log transformed to reduce skewness) and correctness (retaining all responses) were used as dependent variables in linear mixed-effects regression models (Bates and Sarkar, 2006) with participant and word as crossed random effects (see section 4.4.1). A stepwise backward variable selection procedure was used, removing non-significant variables from the models (see section 4.4.3). Both RT and several of the independent variables were logarithmically transformed to reduce the effect of outliers, taking the natural log of the relevant value, plus 1 for variables where any of the values were 0. All frequencies (derived, stem, family and bigram frequencies) were transformed as were P, P*, duration and UP-measures.

No a priori filtering of reaction time was performed; instead, data points with large standardised residuals in the models were excluded (between 2.4% and 2.8% of the data points) and the models refitted, as discussed in section 4.4.4. All tables and figures are based on the refitted models.

For the RTs, a single mixed-effects model was initially fitted for all items in both modalities. This overall model showed significant interactions between modality and a number of the predictors. On inspection, it turned out that the distributions of the residuals differed significantly between the

Table 5.2: Summary of regression model for auditory RTs to simple and complex words (the Auditory All Items Analysis), using contrast coding for Complexity with Simple as the reference level. In this and the following tables for the random effects, 'Groups' denote the main grouping factors (Word and Participant), and 'Name' specifies whether the standard deviation refers to random intercepts or to random slopes for some variable, in this case word frequency. Df = 4537.

Fixed Effects

	Estimate	MCMC mean	HPD 95% CI Lower	HPD 95% CI Upper	p(MCMC)	p(t)
Intercept	15.5405	15.5228	9.1358	21.5190	0.0001	<0.0001
Previous PC1	0.1027	0.1035	0.0866	0.1207	0.0001	<0.0001
Previous PC2	0.0504	0.0504	0.0309	0.0690	0.0001	<0.0001
Trial number	0.0001	0.0001	0.0001	0.0001	0.0001	<0.0001
Complexity: Complex	-9.0672	-9.0334	-16.5240	-1.9143	0.0156	0.0291
Log whole-word frequency	-0.0187	-0.0186	-0.0289	-0.0092	0.0004	0.0006
Complexity:Comp×Word freq	-0.0140	-0.0140	-0.0249	-0.0032	0.0114	0.0251
Log family size	0.0146	0.0146	0.0077	0.0214	0.0001	0.0003
Log UP1^{l}	-3.2217	-3.2159	-5.2753	-1.0734	0.0026	0.0073
Log UP1^{q}	0.2933	0.2929	0.1120	0.4675	0.0014	0.0039
Complexity:Comp×Log UP1^{l}	3.0980	3.0861	0.5346	5.5660	0.0164	0.0304
Complexity:Comp×Log UP1^{q}	-0.2643	-0.2633	-0.4858	-0.0514	0.0172	0.0320
Log UP1 to CUP	0.0345	0.0345	0.0255	0.0433	0.0001	<0.0001
Log CUP to offset	0.0165	0.0166	0.0099	0.0231	0.0001	<0.0001

Random Effects

Groups	Name	SD
Word	Intercept	0.0713
Participant	Intercept	0.0970
Participant	Log word frequency	0.0076
Residual		0.1518

^{l} Linear.

^{q} Quadratic.

two modalities (Kolmogorov-Smirnov D = 0.0594, $p < 0.001$), indicating differences between the two experimental tasks that were not captured with the current set of predictors. The difference may have to do with the fact that visual processing is affected by orthographic consistency as described in section 2.3.1; the various measures which are used to index this in English (cf. Balota et al., 2004; Baayen et al., 2006) were not available for Danish. Because of this difference in the residuals and for ease of interpretation, separate models were fitted to the data from the two modalities. Further motivating separate analyses is the fact that more variables are relevant for the complex than for the simple words (cf. table 5.1). Therefore, simple and complex items were analysed together and compared using those variables that apply to all items, while the complex items were also analysed in a separate model which also included those morpheme-related predictors that were only relevant for the complex words.

All in all, four different mixed-effects models were fitted to the RT data: one for all items and one for complex items in each modality. These are summarised in tables 5.2 to 5.5. The tables summarise the estimated coefficients for the different variables and factors in the first columns ('Estimate') and the associated p-value based on the t-distribution in the right-most columns ('p(t)'). This p-value is based on the upper bound of degrees of freedom (the number of observations minus the number of fixed effects parameters; this is the df-value mentioned in the captions), which is likely to be too high and therefore entails a risk of Type 1 error, i.e. incorrectly rejecting a "correct" null hypothesis. The tables also show the posterior distribution of the fixed effect coefficients, using Markov chain Monte Carlo (MCMC) sampling (see Baayen, Davidson and Bates, 2008a, chapter 7), which constructs multiple samples — in this case 10,000 — of the parameters in the model, based on the dataset and the given mixed-effects model. The mean estimates of these MCMC samples are shown in the second columns of the tables (as 'MCMC mean'); for significant effects, the means are generally close to the model estimates listed in the first column. The third and fourth columns are also based on the MCMC samples, showing the lower and upper bound of the 95% highest posterior density (HPD) intervals. These are the credible intervals within which 95% of the MCMC estimates for the given parameter lie, corresponding to the standard 95% confidence intervals based on the t-distribution, but providing superior accuracy. The p-values listed in the fifth column provide the corresponding MCMC-based significance levels; these are appropriately conservative, in contrast to the somewhat anti-conservative p-values based on the t-distribution with the upper bound as degrees of freedom. The p-values cited in the text are based on the MCMC simulations.

All the models are analyses of covariance which include both continuous predictors, such as frequency, and factors, such as complexity type (simple vs. complex words), except the Visual Complex Error Analysis (table 5.9) in which only continuous predictors were significant. The coefficients for the factor levels specify the estimated contrast coefficients for each of the levels, relative to one level of the factor, the reference level. The reference levels for all factors are specified in the table captions. For instance, in table 5.2 , the group of simple words is the reference level for the factor 'Complexity'; the contrast coefficient for 'Complexity: Complex' shows the adjustment to the intercept for the complex words. In this case, the contrast coefficient is

Table 5.3: Summary of regression model for auditory RTs to complex words only (the Auditory Complex Analysis), using contrast coding for the factor Affix Type with Prefix as the reference level. Number of continuations is residualised from log whole-word frequency and stem frequency from log family size. Df = 2379.

Fixed Effects

	Estimate	MCMC mean	HPD 95% CI Lower	HPD 95% CI Upper	p(MCMC)	p(t)
Intercept	5.6559	5.6557	5.3426	5.9766	0.0001	<0.0001
Previous PC1	0.1137	0.1135	0.0896	0.1370	0.0001	<0.0001
Previous PC2	0.0589	0.0588	0.0323	0.0845	0.0001	<0.0001
Trial number	0.0001	0.0001	0.0001	0.0002	0.0001	<0.0001
Type: Particle	0.0362	0.0361	-0.0007	0.0732	0.0570	0.0513
Type: Suffix	-0.0347	-0.0348	-0.0673	0.0003	0.0472	0.0413
Log whole-word frequency	-0.0348	-0.0348	-0.0417	-0.0279	0.0001	<0.0001
Log stem frequency^r	-0.0383	-0.0382	-0.0587	-0.0179	0.0001	0.0003
Word freq$\times$stem freq^r	0.0061	0.0061	0.0023	0.0100	0.0014	0.0020
Log family size	0.0193	0.0193	0.0104	0.0277	0.0001	<0.0001
Continuations^r	-0.0182	-0.0182	-0.0334	-0.0021	0.0236	0.0239
Log UP1	0.1997	0.1999	0.1490	0.2512	0.0001	<0.0001
Log UP1 to CUP	0.0301	0.0302	0.0204	0.0398	0.0001	<0.0001

Random Effects

Groups	Name	SD
Word	Intercept	0.0618
Participant	Intercept	0.0947
Residual		0.1500

^{l} Linear.

^{q} Quadratic.

^{r} Residualised.

negative, meaning that the complex words are recognised faster, when the values of all other variables are zero. For interactions between a factor and a numerical predictor, such as that between complexity and log whole-word frequency in the Auditory All Items Analysis (also in table 5.2), the main effect of the continuous variable, in this case 'Log Whole-Word Frequency', is specified for the reference level of the factor, in this case the simple words. The interaction parameter, 'Complexity:Comp$\times$Word Freq', shows the adjustment to the slope of the numerical predictor whole-word frequency for the specified factor level, namely the complex words. The whole-word frequency effect is significant for the reference level simple ('Log Whole-Word Frequency' is associated with a MCMC-based p-value of 0.0004), but significantly stronger for the complex words, as indicated by the negative estimated coefficient for

Table 5.4: Summary of regression model for visual RTs to simple and complex words (the Visual All Items Analysis), using contrast coding for the factor Complexity with Simple as the reference level and for Previous Response with Correct as the reference level. Number of continuations and length in letters are residualised from log whole-word frequency. Df = 4237.

FIXED EFFECTS

	Estimate	MCMC mean	HPD 95% CI Lower	HPD 95% CI Upper	p(MCMC)	p(t)
Intercept	6.7960	6.7973	6.5116	7.0639	0.0001	<0.0001
Previous PC1	-0.1180	-0.1186	-0.1373	-0.1015	0.0001	<0.0001
Previous PC3	-0.0229	-0.0234	-0.0463	-0.0020	0.0386	0.0434
Previous response: Error	0.0773	0.0770	0.0510	0.1009	0.0001	<0.0001
Complexity: Complex	-0.3964	-0.3993	-0.7461	-0.0362	0.0260	0.0379
Log whole-word frequency	-0.0480	-0.0480	-0.0550	-0.0405	0.0001	<0.0001
Log family size^l	-0.0427	-0.0427	-0.0731	-0.0130	0.0050	0.0085
Log family size^q	0.0052	0.0052	0.0018	0.0083	0.0014	0.0033
Continuations^r	-0.0194	-0.0193	-0.0275	-0.0107	0.0001	<0.0001
Length letters^r	0.0164	0.0164	0.0100	0.0229	0.0001	<0.0001
Log UP1	0.0034	0.0030	-0.0423	0.0497	0.9044	0.8934
Complexity:Comp×Log UP1	0.0642	0.0647	0.0023	0.1249	0.0394	0.0510

RANDOM EFFECTS

Groups	Name	SD
Word	Intercept	0.0515
Participant	Intercept	0.0775
Participant	Word frequency	0.0108
Residual		0.1762

^{l} Linear.

^{q} Quadratic.

^{r} Residualised.

the interaction parameter and the associated significant p-value. Some of the effects are significantly non-linear, such as the UP1-effect in the Auditory All Items Analysis summarised in table 5.2; for such quadratic effects, the corresponding linear term is included in the model even if it is not significant in itself. The effects that are discussed as linear in the following are those for which only the linear term was significant in the regression model where both the dependent variable RT and the independent numerical variables were log transformed.

Table 5.5: Summary of regression model for visual RTs to complex words only (the Visual Complex Analysis), using contrast coding for the factor Previous Response with Correct as the reference level. Number of continuations and length in letters are residualised from log whole-word frequency, juncture bigram frequency from log stem frequency. Df = 2214.

Fixed Effects

	Estimate	MCMC mean	HPD 95% CI Lower	HPD 95% CI Upper	p(MCMC)	p(t)
Intercept	6.4688	6.4681	6.1811	6.7415	0.0001	<0.0001
Previous PC1	-0.1206	-0.1208	-0.1469	-0.0936	0.0001	<0.0001
Previous response: Error	0.0969	0.0971	0.0621	0.1359	0.0001	<0.0001
Log whole-word frequency	-0.0394	-0.0394	-0.0465	-0.0326	0.0001	<0.0001
Stem frequency^r	-0.0148	-0.0148	-0.0232	-0.0060	0.0014	0.0007
Log family size^l	-0.0748	-0.0750	-0.1238	-0.0279	0.0024	0.0022
Log family size^q	0.0086	0.0086	0.0033	0.0135	0.0008	0.0008
Continuations^r	-0.0334	-0.0333	-0.0489	-0.0180	0.0001	<0.0001
Length letters^r	0.0154	0.0155	0.0064	0.0245	0.0006	0.0008
Log UP1	0.0656	0.0658	0.0215	0.1090	0.0038	0.0035
Juncture bigram frequency^r	0.0088	0.0088	0.0017	0.0167	0.0226	0.0211

Random Effects

Groups	Name	SD
Word	Intercept	0.0530
Participant	Intercept	0.0577
Residual		0.1877

^{l} Linear.

^{q} Quadratic.

^{r} Residualised.

The figures in this section show the partial effects of the predictors discussed. All effects are adjusted to the median values for the other continuous predictors and the reference levels in the relevant models, thus showing the effect of a given predictor with all other predictors held constant. The vertical axes show RT which is back-transformed to non-logged RT in order to make the RT-variable more interpretable. The back-transformation of RT in the plots may mean that effects which are linear in the analyses may not appear entirely linear in the plots; all effects are discussed on the basis of the models. The same range of 600 ms is used for the vertical axes in all plots to allow comparison of the magnitude of the effects; the same range is also used in Experiments 2 and 3.

In addition to the primary analyses of the reaction times, the error data were analysed, including all responses to all items. Four logistic regression models were fitted to the binomial error data, corresponding to the four different RT-models. The error analyses are summarised in tables 5.6 to 5.9, with estimated coefficients in the second column, and their associated standard error, z- and p-values in the subsequent columns. The percentage of correct responses was so high that the error analyses are not very powerful; therefore, their main function was to confirm the effects in the RT-analyses and to show whether any of the variables in the RT-analyses showed speed/accuracy trade-offs.

The lexical variables used in the current regression models (see table 5.1) were highly correlated, indicated by a κ of 157 for the raw variables (where κ-values over 30 indicate high and potentially harmful collinearity, see section 4.4.2). Various steps were taken to reduce this collinearity; these steps are described in more detail in section 4.4.2. RTs on previous trials were orthogonalised using principal components analysis. Additionally, the UPs and duration were recalculated, such that they did not measure overlapping parts of the signal: UP1 was included as distance from onset to UP1, the Complex UP (CUP) as distance from UP1 to CUP, and word duration as distance from CUP to word offset. These less collinear variables are included in the analyses discussed below, as indices of the UPs and duration. Finally, various lexical variables were decorrelated by fitting regression models where one variable was predicted by those variables with which it was highly correlated, and then using the residuals of the decorrelating regression models as predictors. In this way, the number of continuation forms and the length in letters (used in the analysis of the visual experiment) were both decorrelated from whole-word frequency, stem frequency was decorrelated from morphological family size and juncture bigram frequency from stem frequency; the decorrelated variables are marked as "Residualised" or "Resid." in tables and figures. With these decorrelated measures, collinearity was low, with κ below 10 for all models. In sum, it can be concluded that collinearity is not a problem for the analyses reported in tables 5.2 to 5.9.

The remainder of this section presents and discusses the results of the analyses which fall into three main groups: First, the context-related control variables are discussed in section 5.3.1. Then, the more content-based variables are considered in sections 5.3.2 (the difference between the complexity types), 5.3.3 (the patterns of frequency effects) and 5.3.4 (effects of morphological family). Finally, the form-related measures are discussed, uniqueness points and lengths in section 5.3.5 and junctural probability in section 5.3.6.

Table 5.6: Summary of the logistic regression model for correctness on all items in the auditory Experiment 1a (the Auditory All Items Error Analysis), using contrast coding for the factor Complexity with Simple as the reference level. Number of continuations is residualised from log whole-word frequency. Df = 4931.

Fixed Effects

	Estimate	SD	z	p(z)
Intercept	-0.2950	0.5804	-0.508	0.6113
Complexity: Complex	-1.5999	0.4150	-3.856	0.0001
Log whole-word frequency	-0.7316	0.0902	-8.111	<0.0001
Continuations^r	-1.4856	0.4162	-3.569	0.0004
Log word freq×Continuations^r	0.2598	0.0779	3.337	0.0008

Random Effects

Groups	Name	SD
Word	Intercept	1.3569
Participant	Intercept	0.8251

^{r} Residualised.

Table 5.7: Summary of the logistic regression model for correctness on complex words in the auditory Experiment 1a (the Auditory Complex Error Analysis), using contrast coding for the factor Affix Type with Prefix as the reference level. Number of continuations is residualised from log whole-word frequency. Df = 2619.

Fixed Effects

	Estimate	SD	z	p(z)
Intercept	-14.3841	5.5680	-2.583	0.0098
Type: Particle	-0.3708	0.4095	-0.905	0.3652
Type: Suffix	-2.1781	0.4735	-4.600	<0.0001
Log whole-word frequency	-0.6997	0.0943	-7.418	<0.0001
Continuations^r	-1.6398	0.4553	-3.602	0.0003
Log word freq×Continuations^r	0.2670	0.1184	2.255	0.0241
Log Complex UP	2.1864	0.9161	2.387	0.0170

Random Effects

Groups	Name	SD
Word	Intercept	0.9408
Participant	Intercept	1.0366

^{r} Residualised.

Table 5.8: Summary of the logistic regression model for correctness on all words in the visual Experiment 1b (the Visual All Items Error Analysis), using contrast coding for the factor Complexity with Simple as the reference level. Number of continuations is residualised from log whole-word frequency. Df = 4693.

FIXED EFFECTS

	Estimate	SD	z	p(z)
Intercept	-3.5585	2.1607	-1.647	0.0996
Previous PC1	0.7287	0.2298	3.171	0.0015
Complexity: Complex	-0.6481	0.3153	-2.056	0.0398
Log whole-word frequency	-0.5808	0.0624	-9.310	<0.0001
Log word freq×Continuations^r	0.1696	0.0572	2.964	0.0030
Continuations^r	-1.0638	0.2917	-3.647	0.0003
Log family size	-0.2461	0.0725	-3.392	0.0007
Log UP1	0.7233	0.3593	2.013	0.0441

RANDOM EFFECTS

Groups	Name	SD
Word	Intercept	0.8257
Participant	Intercept	0.5416

^{r} Residualised.

Table 5.9: Summary of the logistic regression model for correctness on complex words in the visual Experiment 1b (the Visual Complex Error Analysis). Number of continuations is residualised from log whole-word frequency. Df = 2494.

FIXED EFFECTS

	Estimate	SD	z	p(z)
Intercept	-8.0427	2.8172	-2.855	0.0043
Previous PC1	1.0737	0.3065	3.503	0.0005
Log whole-word frequency	-0.5101	0.0744	-6.855	<0.0001
Continuations^r	-1.3495	0.4034	-3.345	0.0008
Log word freq×Continuations^r	0.2157	0.0983	2.194	0.0282
Log family size	-0.2184	0.0915	-2.388	0.0169
Log UP1	1.3222	0.4852	2.725	0.0064

RANDOM EFFECTS

Groups	Name	SD
Word	Intercept	0.9393
Participant	Intercept	0.6558

^{r} Residual.

5.3.1 Context variables

In all analyses, there were significant effects of control predictors relating to the experimental context of each item, primarily of the RTs on previous trials. These variables are more informative about the nature of the task than about the mental lexicon, but the effects were strong and highly significant and it is therefore important to bring these variables under statistical control.

The RT on a given trial was predicted by the RTs on the preceding trials as well as by the correctness of the immediately previous response. RT on a given trial was generally longer when the RTs on previous trials were long and when the response on the immediately previous trial was an error. Since the previous RTs are naturally highly collinear, they were orthogonalised using principal components analysis (PCA), in order to avoid potentially harmful collinearity from this source (following de Vaan et al., 2007 and Baayen et al., 2007).

The principal components occur as Previous PCs in the tables and in figure 5.1. PC1, the principal component that accounts for most of the variance in the four previous RTs, had a significant effect in both the auditory and visual experiments as shown in the top panels of fig. 5.1. The factor loadings of the original predictors (corresponding to correlations between the PC and the original variables) were positive for the auditory PC1, so the upward slope in the top left panel of fig. 5.1 reflects an inhibitory effect: the longer the previous RT, the longer the current RT. For the visual PC1, the factor loadings were negative and the coefficient for the effect was also negative, meaning that PC1 — as in the auditory modality — reflects an inhibitory effect of the original predictors, in spite of the downward slope of PC1 in the top right panel of fig. 5.1.

The previous PC2 was only significant in the auditory task; this is illustrated in the left panel of the middle row of fig. 5.1. The factor loadings of the original variables on PC2 shows that this effect predominantly reflects RT on the immediately previous trial; RTs tended to be long, when the RT on the previous trial was also long. PC3 was only significant in the visual task. The effect of PC3 is seen in the right panel of the middle row of fig. 5.1; the factor loadings of the original variables were both positive and negative, so the effect of PC3 is not straightforwardly interpretable. However, it still serves to control the context-related variation.

In the error analyses, only the previous PC1 was significant and only for the visual experiment. The positive coefficients for the effect in the visual error analyses (see tables 5.8 and 5.9) reflect a facilitatory effect of the original variables: Participants were less likely to make an error when the previous RTs were long. This could be a direct result of the previous RTs being long or, what is perhaps more likely, it could be caused by the fact that the current RT was longer when previous RTs were long, increasing the probability of

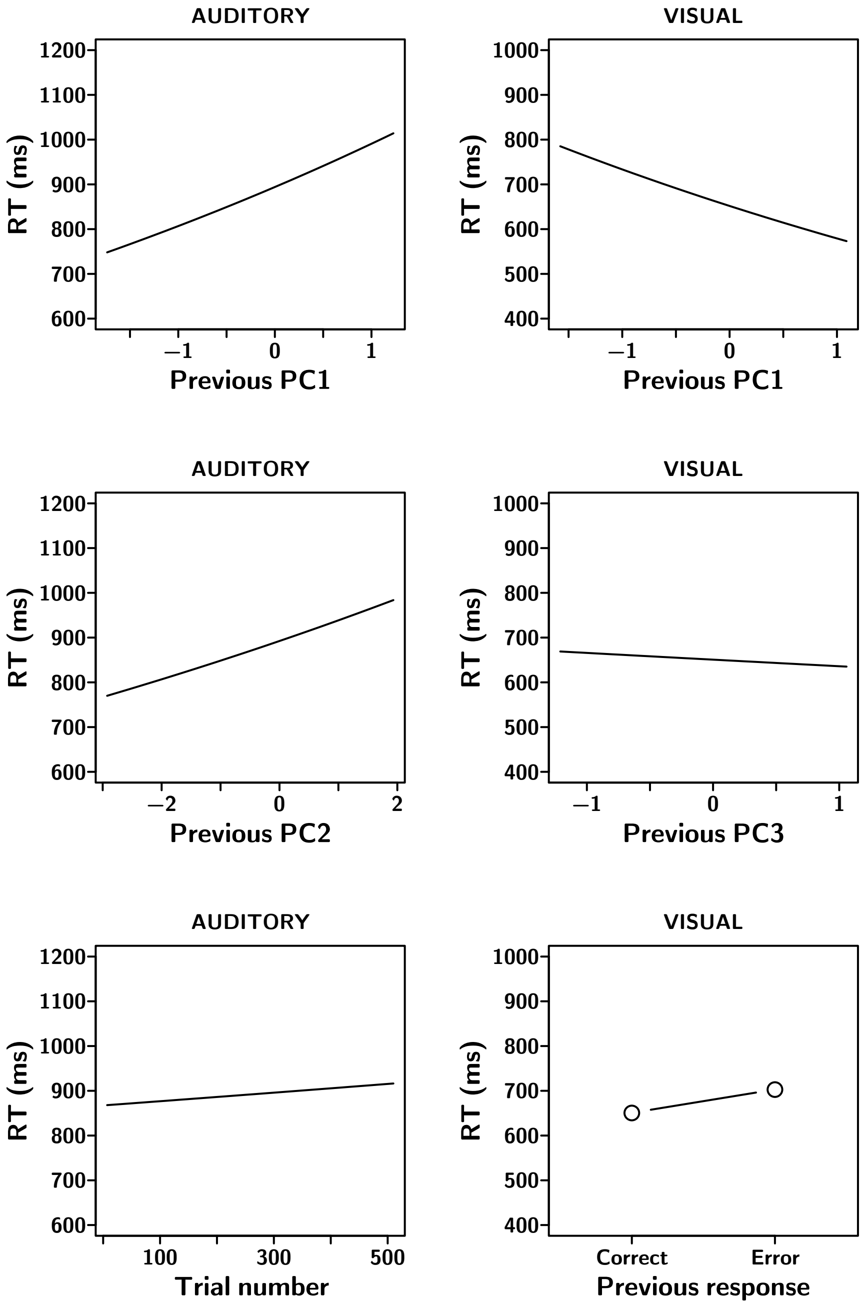

Figure 5.1: Partial effects of the context-related control variables Previous PC1, PC2 and PC3 as well as trial number and previous response correctness. The plots are based on the All Items Analyses. The panels to the left show the results of the auditory experiment, the plots to the right the results of the visual experiment. The vertical axes show RT in ms, while the horizontal axes in the top two rows show the scales of the principal components based on previous RT. The line between the two circles in the bottom right panel is drawn to make the difference between the categories more clearly discernible.

a correct response. The fact that the effect of this on correctness was only significant in the visual experiment is probably a result of the variable ISI and the resulting faster pace of this experiment.

In all cases, the effects of RTs on previous trials reflect temporal consistency across trials. The facilitatory effects of RT on previous trials are the analytical reflex of what participants termed a good flow in responding.

Another effect of experimental context is shown in the bottom left panel of fig. 5.1. This is an inhibitory effect of trial number, a weak fatigue effect which was only significant in the slower auditory experiment. The difference in the pace of the experiments is also the likely reason why the correctness of the previous responses was only significant in the visual task, such that RTs were slower if the previous response was an error, as shown in the bottom right panel of fig. 5.1. This effect probably reflects a consciousness on the part of the participants of having made an error, resulting in more careful, slower responding on the trials immediately following an error.

Finally, I investigated whether there was an effect of affix repetition, by adding a variable that indexed how many times the specific affix on each complex word had previously been encountered in the experimental list. A facilitatory effect of this would correspond to an affix priming effect, but the variable was solidly non-significant in all analyses (p-values > 0.5) and therefore not included in the final models summarised in tables 5.2 to 5.9.

This section and fig. 5.1 are based on the analyses of all items but almost identical patterns were found in the analyses of complex items, with the exception that, for the complex items in the visual experiment, Previous PC3 was not significant, a fact that can be ascribed to the reduced statistical power of that analysis relative to the Visual All Items Analysis.

5.3.2 Complexity types

In the All Items Analyses, two complexity types are compared, namely the simple and the complex words. The inclusion of this complexity type variable addresses the first question outlined in the introduction, namely whether the complex words have an advantage over the simple words. In terms of mean RTs, the group of simple words were recognised faster than the group of complex words (944 vs. 1033 ms in the auditory experiment, 646 vs. 725 in the visual). However, this difference in the group means could be due to distributional factors such as length and frequency. The inclusion of the complexity type factor in the regression models makes it possible to investigate whether there is a difference between simple and complex words once co-variates such as length and frequency are controlled.

In both the Auditory and the Visual All Items Analyses (summarised in tables 5.2 and 5.4), the simple words are the reference level, with the contrast coefficient for complex indicating the estimated size and significance of the difference between the group of complex words and the reference level simple words, when all numerical variables are set to zero. In both experiments, complexity interacted with one or more co-variates. These interactions must be considered, because they indicate that the complexity type difference varies with different values on the interacting co-variates. The difference in the intercept between the complexity types only reflects the difference when the values of all co-variates are zero.

In the auditory Experiment 1a, complexity type interacted with both frequency and UP1, while complexity type interacted with UP1 in the visual Experiment 1b. In the top left panel of fig. 5.2, the UP1-effects are plotted separately for the complex (solid line) and the simple words (dashed line), for the auditory Experiment 1a. The line for the simple words is always higher than the line for the complex words, indicating an advantage for the complex words, but this advantage is small for the middle of the UP1-range. UP1 is the point where competition from unrelated lemmas is resolved. It is plausible that the role of morphemic information in the recognition of complex words is stronger when this competition is resolved early; this in turn means that the advantage for complex words — which would rely on the availability of both morphemic and whole-word information — is stronger for words with early UP1. A similar advantage for the complex words when UP1 is early is observed for the complex words in Experiment 1b, as illustrated in the bottom left panel of fig. 5.2. The differences between the UP1-effects are further discussed in section 5.3.5.

The interaction of complexity type with frequency in Experiment 1a is illustrated in the top right panel of fig. 5.2. Again, the simple words (shown by the dashed line) are recognised slower than the complex words (the solid line), and this difference becomes larger for higher-frequency words. Overall, the auditory Experiment 1a shows an advantage for the complex words which, however, is small for the middle range of UP1-values. The interaction is further discussed in section 5.3.3.

The error analyses confirm this general complexity advantage, showing significantly fewer errors for complex than for simple words both in the auditory and in the visual modality.

Crucially, the complexity type effects are not artefacts of affix repetition speeding up the recognition of complex words, since there was no effect of the number of times an affix had previously been encountered in the experiment (see the previous section, 5.3.1).

Since both verbs, nouns and adjectives were included, the question arises whether it plays a role that the verbal base form in Danish — the form in which the simple verbs were presented in the experiments — is not the root, but the infinitive which for consonant-final roots consists of the root plus a schwa. As argued in section 3.2, this form is the base form for verbs in Danish;

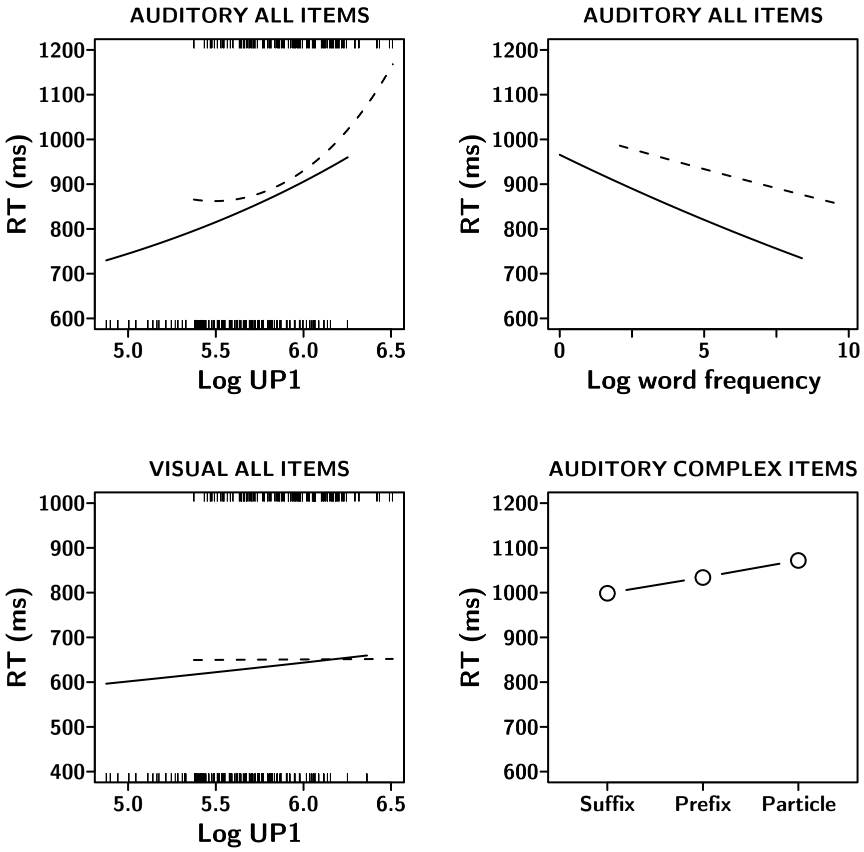

Figure 5.2: Partial effects of complexity type in the auditory Experiment 1a and the visual Experiment 1b. The simple words are represented by dashed lines, the complex by solid lines. The top row shows the differences between simple and complex words in Experiment 1a for different values of UP1 (left) and frequency (right) when other variables are held constant at their medians. Similarly, the bottom left panel shows the difference between the simple and complex words in Experiment 1b for different values of UP1. The bottom right panel shows the effect of affix type in the Auditory Complex Analysis, the lines between the circles representing the categories are included to make the difference between the categories more clearly discernible. Both the panels to the left show the distribution of UP1-values for the items: the tick marks on the top bar of the panel show the distribution of UP1-values for the simple words, the tick marks on the bottom bar, the UP1-values for the complex words.

the schwa is devoid of content and phonotactically motivated. This justifies the classification of these infinitival forms as simple. None of the analyses showed any effects of word class. More specifically, neither the Auditory nor the Visual All Items Analysis showed an interaction between complexity type and word class which would have indicated that the complexity advantage is different for verbs than for nouns and adjectives.

The All Items Analyses summarised in tables 5.2, 5.4, 5.6 and 5.8 operate with the simple complexity type division between simple and complex words for the most parsimonious analysis of the data. However, if affix type (the difference between prefixed, particle prefixed and suffixed words) is incorporated in the All Items Analyses, it emerges that the complexity advantage is carried more by the suffixed words than by the prefixed and particle prefixed in both modalities. This suggests that the complexity advantage relies more on the stem than on the affixes, which is unsurprising given that stems are generally longer and more informative than affixes.

The analyses of the complex words in the auditory modality (see tables 5.3 and 5.7) reveal a similar pattern: The suffixed words are recognised faster than the prefixed words (which are the reference level), while the particle prefixed words are recognised more slowly than the prefixed words. The difference to the prefixed words is just significant for the suffixed words ($p = 0.0472$) and almost significant for the particle prefixed ($p = 0.0570$). This pattern is illustrated in the bottom right panel of fig. 5.2. The corresponding error analysis shows that the suffixed words were also significantly less error-prone than the prefixed (see table 5.7). In contrast to the All Items Analyses discussed above, there were no interactions involving complexity type in the Complex Analysis. Two aspects of the complexity type results need to be explained: the difference between the suffixed and both types of prefixed words, and the difference between the particle prefixed and both other types of words.

The difference between suffixed words on one hand and prefixed and particle prefixed words on the other can be explained in the same way as the fact that the suffixed words showed a stronger complexity advantage relative to the simple words: If a complexity advantage depends mostly on the activation of the stem rather than the affix alongside the whole word, then the fact that the stem is word-initial and carries primary stress in the suffixed words explains their advantage both relative to the simple and relative to the prefixed and particle prefixed words. In addition to this explanation based on facilitatory characteristics of the suffixed words, the difference can be understood with reference to inhibitory properties of prefixed and particle prefixed words: Both types of prefixes have large word-initial cohorts and these word-initial cohorts are likely to be semantically much more diverse for the prefixed and particle prefixed words than for the suffixed words, resulting in a disadvantage for prefixed and particle prefixed words. Given that no affix type frequency effect was observed, the semantic diversity of the prefix cohorts is probably the more important of the two characteristics.

The second question is why this disadvantage relative to the suffixed words is larger for the particle prefixed than for the prefixed words or, phrased differently, why the particle prefixed words are recognised more slowly than both other complex types, though the difference to the prefixed words is not quite significant. One reason could be semantic: Because of the generally richer semantics of the particle prefixes, derived forms carrying particle prefixes are arguably slightly less transparent than those carrying suffixes or standard prefixes. Since less transparent words are recognised more slowly (see e.g. Wurm, 1997), this could explain why the particle prefixed words are recognised more slowly than the prefixed and suffixed words. Another possible explanation is that the large and diverse word-initial cohorts of both types of prefixed words are more inhibitory for the particle prefixed words because the cohorts also contain the particle as an independent word, which might be the preferred interpretation early in the word, triggering time-costly reanalysis when the stem is encountered. This could hold for the six out of eight particles that do not vary allomorphically when occurring as prefixes, while it is less plausible for the particles *af-* ('away', 'from') and *bag-* ('back', 'behind') which have different vowel phonemes in the two different functions. The stems of suffixed words can, at least from a phonological point of view, also function as independent words, but the stem of a suffixed word is more likely to be helpful than the particle of a particle prefixed word, because the semantic information is richer and more closely related to the whole-word meaning in transparent complex words.

There was no effect of affix type in the Visual Complex Analyses (see tables 5.5 and 5.9). The differences between prefixed and suffixed and between prefixed and particle prefixed words in the auditory Experiment 1a were understood in terms of the gradual availability of the signal over time which is characteristic of the auditory modality. Therefore, it makes sense that the difference between the three types of complex words should not be strong enough to reach significance in the visual modality: When the whole word is present at once and may be read in one fixation, the advantage of the most informative constituent occurring first disappears, and the risk of initially parsing the particle as an independent word, triggering later reanalysis, should also be minimised.

In sum, the hypothesis that morphological structure aids word recognition is confirmed by the present experiments, though the complexity advantage is weak for some UP1-values. The overall result is replicated in Experiments 2 (see section 6.3.2) and 3 (cf. section 7.3.5). These results are further discussed in these sections and in chapter 8.

5.3.3 Frequencies

Frequency was a main manipulation in this experiment, as it is in the field more generally, see section 2.4. There were solid effects of whole-word frequency in all analyses, as well as a number of interesting interactions which are described in the three following subsections: First, the whole-word frequency effects in the All Items Analyses are discussed, with a focus on the interaction between complexity type and whole-word frequency in the auditory modality. Then, I describe the effects of stem and whole-word frequency in the analyses of the complex items and the interaction between them in the auditory modality. Finally, two more marginal interactions are briefly considered, between frequency and continuation forms in the error analyses and between frequency and length in the visual modality. All effects of frequency on RT are depicted in fig. 5.3.

In addition to the fixed effects of frequency, the two analyses of RT to all items included random slopes for frequency, in effect taking into account differences between the participants in terms of the effects of corpus-based frequency on recognition time. The inclusion of these was supported by likelihood ratio tests (p-values < 0.001). Such tests did not support the inclusion of random slopes for frequency in the Complex Analyses (p-values > 0.5). This suggests that the variance that the random slopes account for in the All Items Analyses is accounted for by the morphological variables that are included in the Complex Analyses.

Initial analyses of the data from the visual experiment suggested an interaction between sex and whole-word frequency but this turned out to be very clearly driven by a single outlier, a male participant with an extremely steep slope for frequency, see also fig. 4.1 on p. 97 and discussion in section 4.4.1. When random slopes for frequency by participant were included, the interaction between sex and frequency was non-significant. In the auditory analyses, there was a tendency towards stronger frequency effects for female than for male participants, similar to the significant sex differences found in Experiments 2 and 3 (see chapters 6 and 7) and by Ullman et al. (2002). However, in this experiment, the interaction was not significant once random slopes for frequency were added, indicating that the variation between participants was stronger than the difference between the sexes.

Complexity by frequency

The analyses of simple and complex items showed strong effects of whole-word frequency in both experiments: More frequent words elicited significantly faster RTs and fewer errors. The simple main effect of whole-word frequency for all items in the visual experiment is shown in the top left panel of fig. 5.3. In this experiment, the frequency effects were the same for simple and complex items.

In the auditory experiment, the whole-word frequency effect differed between the simple and the complex words, as shown in the top right panel of fig. 5.3: the slope of the solid line representing the complex words is steeper than that of the dashed line representing the simple words. In table 5.2, the term 'Log whole-word frequency' refers to the effect of frequency for the

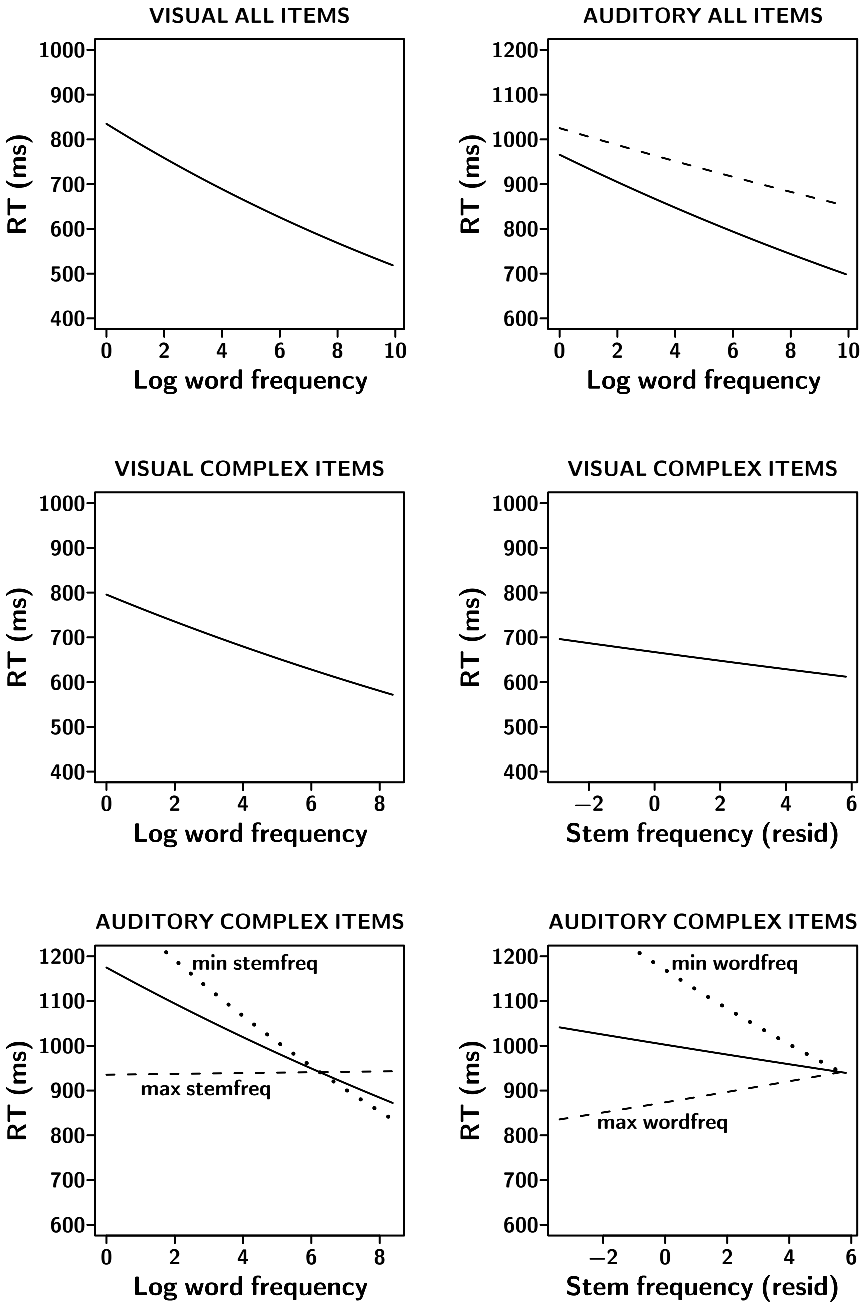

Figure 5.3: Partial effects of stem and whole-word frequency. The top two panels show the effects of whole-word frequency in the All Items Analyses; for Experiment 1a shown in the top right panel, this effect differs between simple words (the dashed line) and complex words (the solid line). The middle panels show the main effects of whole-word and stem frequency for the complex items in the visual Experiment 1b. The bottom panels illustrate the interaction between stem and whole-word frequency for the complex items in the auditory Experiment 1a: the left panel shows the effect of whole-word frequency for words with minimum (dotted), median (solid) and maximum (dashed) stem frequency; the right panel shows the effect of stem frequency for words with with minimum (dotted), median (solid) and maximum (dashed) whole-word frequency. The convention of using dotted lines for minimum values, solid lines for median values and dashed lines for maximum values is used where relevant throughout the thesis.

simple words which are the reference level; the effect is significant, but the interaction term shows that the effect is significantly stronger for the complex words. No such effect is found in the error analysis.

It was argued in section 2.4 that the whole-word frequency of complex words can be understood as a combinatorial probability, the probability of the morphemes in the complex words co-occurring. At the same time, it can be understood as a straightforward string probability, as it is for the simple words. The stronger whole-word frequency effects for the complex words in Experiment 1a support this dual interpretation: If whole-word frequency has both a combinatorial and a whole-word component for the complex words, but only a whole-word component for the simple words, the effect becomes stronger for the complex word because it indexes two probabilities or two sources of information.

There is evidence that frequency effects may at least partially be artefacts of the age at which words are acquired (see section 2.4). Age of acquisition is also a possible factor in the frequency effects observed in the present experiments, but there are at least two reasons why age of acquisition should not be the only factor: Firstly, as reasoned by Baayen et al. (2007), age of acquisition is less likely to be a confounding factor for low-frequency words, which many of the present items are. Secondly, the fact that the frequency effects are stronger for complex words, which are generally likely to be acquired later than simple words, implies that there is an effect of frequency over and above any effect of age of acquisition. Age of acquisition is not further considered in the following experiments.

Whole-word and constituent frequencies

While only whole-word frequency is relevant in the analyses of all items, two additional frequencies are relevant when considering complex words, namely the frequencies of their constituents, the stems and the affixes. As described in section 2.4, stem frequency has received extensive attention in the literature as the typical index of morphological processing, but as will be shown in chapter 6, affix frequency is another possible measure of morphemic processing. Affix frequency appears to be more relevant for suffixed than for prefixed words. In the present experiments, there were no effects of affix frequency or of any of the measures of affix productivity described in section 3.4.2, but the analyses of complex items show effects of both stem and whole-word frequency in both modalities.

The simplest case is the visual Experiment 1b where there were significant main effects of both whole-word and stem frequency, with neither entering into any interactions. The effects are visualised in the middle panels of fig. 5.3, showing that, as expected, whole-word frequency had a stronger effect than stem frequency. This shows that both combinatorial and string probabilities play a role, and that both whole-word and morpheme information contribute to the visual word recognition process. In both the visual and the auditory analyses, stem frequency was decorrelated from morphological family size with which it is somewhat correlated.

The picture is more complicated for the auditory experiment, in which stem and whole-word frequency interact. The interaction is shown from two perspectives in the bottom two panels of fig. 5.3. The bottom left panel plots the effect of word frequency for words with different stem frequencies: The solid line represents the straightforward facilitatory effect of whole-word frequency for words with median stem frequency. The dotted line shows that this effect is stronger for words with low stem frequency. It is attenuated for words with high stem frequency as shown by the almost flat dashed line. The bottom right panel shows the effect from the opposite perspective: For words with median whole-word frequency (shown by the solid line), the effect of stem frequency is facilitatory, but weaker than the corresponding effect of whole-word frequency for words with median stem frequency. The bottom right panel also shows that the facilitatory effect of stem frequency is much stronger when whole-word frequency is low (the dotted line), while the effect becomes inhibitory for words with high whole-word frequency (the dashed line). The division into minimum, median and maximum values is used in fig. 5.3 to illustrate the interaction; the statistical models do not operate with such a division. This pattern is remarkably similar to the interaction between affix and whole-word frequency in Experiment 2; possible interpretations of both interactions are discussed in sections 6.3.3 and 8.3.

The interaction between stem and whole-word frequency is reminiscent of Hay's (2001) argument that the roles of the two frequencies depend not on absolute but on relative frequency: if stem frequency is high, the whole-word frequency has little effect, and when whole-word frequency is high, stem frequency is inhibitory. However, Hay's precise prediction is not upheld: she argues that if stem frequency is higher than whole-word frequency, there should be effects of stem frequency and vice versa. A binary factor indicating whether stem or whole-word frequency is the highest did not interact with either frequency variable.

As mentioned, the frequency measure used for the whole word is the frequency of its base form, mainly because this is the form that is actually encountered in the experiments. In contrast, the best stem frequency predictor is the frequency of the stem lemma: in the analysis of the visual RTs, only the lemma-based stem frequency measure was significant, and for the auditory RTs, it was a stronger predictor than the base-form measure. In the complex words, the stem frequency is supposed to index familiarity with the stem lemma, hence the lemma frequency is a more relevant predictor than the frequency of a specific form. This difference in the significance of the two stem frequency variables is suggestive, but not solid enough to support any strong conclusions about the structure of the mental lexicon; the materials were not designed to tease these effects apart.

Other frequency interactions

In addition to the more solid interactions discussed in the previous subsections, two relatively marginal interactions are worth mentioning: firstly, an interaction between whole-word frequency and the number of continuation forms in the error analyses, secondly, a borderline significant interaction between whole-word frequency and length in letters in the visual modality.

The coefficients for whole-word frequency in tables 5.6 to 5.9 reflect a facilitatory effect, with fewer errors for more frequent words. However, this is modulated by the significant interaction between whole-word frequency and the number of continuation forms, with whole-word frequency effects being stronger for words with fewer continuation forms and absent for the words with the maximum number of continuation forms. Seen from the opposite perspective, there was also a facilitatory effect of the number of continuation forms, but only for the words with the very lowest whole-word frequency. In other words, the effect of frequency is much more robust than the effect of continuations. Because the two measures were correlated in this dataset ($R^2 = 0.4749$), they were decorrelated for the analyses; the continuation-variable in these models is the residuals of the decorrelating regression model.

The relation between the continuation forms and the whole-word frequency can be understood in terms of conditional probabilities: The continuation forms that are activated as the word is heard or read restrict the search space for the correct target word considerably relative to the full lexicon; the stronger this set of continuations is (as indexed by the continuation forms variable), the more efficient this restriction is. What the interaction seems to imply is that the restriction of the space of lexical candidates that is indexed by the continuation forms variable is more facilitatory when the whole-word frequency is low, indicating a low a priori probability of the target being that word. Conversely, when the probability of the word is high a priori, as indexed by a high whole-word frequency, it seems to matter less to what extent the set of continuation forms help restrict the set of lexical candidates. Since this pattern is only observed in the error analyses, it should be interpreted cautiously.

In the initial analyses, an interaction was found in the visual modality between length in letters and frequency, such that the overall facilitatory frequency effect was stronger for long words and attenuated for short words; conversely, the inhibitory length effect was stronger for low-frequency words and attenuated for high-frequency ones. In other words, readers were more sensitive to frequency in long words than in short; this could both be because long words are more difficult and thus a facilitatory factor such as frequency becomes more important. Another possible reason is that the frequency effect has more time to emerge in the longer words, which take longer to recognise even if they are frequent. Seen from the opposite perspective, low-frequency words are more vulnerable to inhibitory length effects, whereas the high-frequency ones are familiar enough to readers to be processed so holistically that length has a reduced effect. However, this effect is not included in the final model summarised in table 5.4, because the effect, which was relatively weak to start out with, disappeared when words carrying two long and atypical (in being bisyllabic and containing full vowels) suffixes, *-agtig* and *-mæssig*, were not included in the analysis. It is nonetheless noticeable that this trend is similar to the interaction between whole-word frequency and length observed by Moscoso del Prado Martín et al. (2004a) and the effects found in an eye-movement study of Finnish by Bertram and Hyönä (2003).

5.3.4 Morphological family

This section discusses two different morphological family effects: first, the effect of the standard morphological family, which is a type count of derived forms and compounds that contain the stem of the target, and then the set of morphologically related continuation forms, which to some extent constitute a subset of the family. The morphological family of a word like *aften* (evening) includes derivational forms and compounds like *aften-lig* ('evening-like'), *aften-tur* ('evening walk') and *premiere-aften* ('opening night'), while the continuation forms of *aften* encompass only those cases where *aften* is at the beginning of the word, i.e. *aften-lig* and *aften-tur* but not *premiere-aften*, but additionally include inflectional forms like *aften-er* ('evening-s').

Considering first the visual modality, there was a non-linear effect of morphological family size which is illustrated in the top left panel of fig. 5.4. The effect is facilitatory up to a certain point, but flattens out and then becomes inhibitory for families larger than about 150 members. This non-linear family size effect is parallel to the effects observed in the regression studies of Tabak et al. (2005) for Dutch and Baayen et al. (2006) for English. The facilitatory effect reflects that form-meaning mappings for the morpheme in question become more systematic as families become larger (Feldman et al. 2006: 60). However, above a certain family size, so many family members are activated that they constitute noise and therefore slow down the recognition of the target. Moreover, the larger families naturally tend to be semantically more loosely connected, which should detract from the facilitatory family effect given the semantic nature of the effect (cf. section 2.5). The inhibitory component is stronger in the visual Experiment 1b than in the Dutch data of Tabak et al. (2005); this may be caused by the fact that Danish morphological

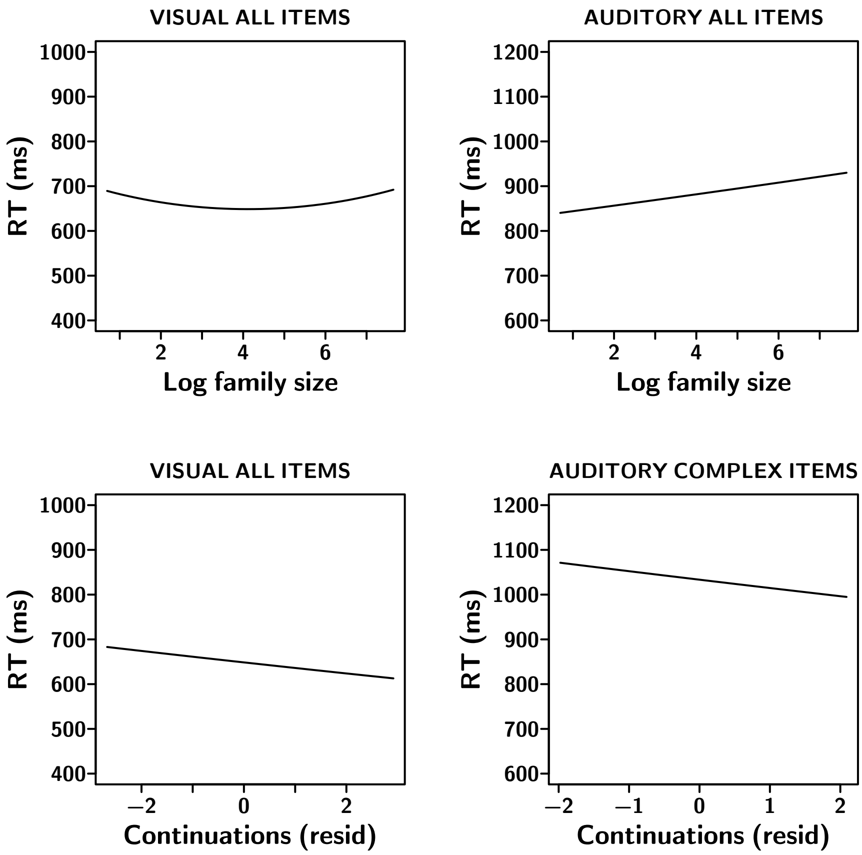

Figure 5.4: Partial effects of morphological family size and number of continuation forms, with results from the visual experiment in the left panels and from the auditory experiment in the right panels. The plots of the family size effect are based on the analyses of all items. The auditory effect of the continuation forms, shown in the bottom right panel, is significant in the analysis of complex items only, while the visual effect in the bottom left panel is based on all items. The continuations are residualised from whole-word frequency.

families are overall substantially larger than Dutch and English ones, cf. the comparison of Danish and Dutch in section 3.4.3 and of Swedish (which is similar to Danish in this respect) and English in Josefsson (2005).

Turning to the auditory experiment, this showed a straightforward inhibitory effect, as illustrated in the top right panel of fig. 5.4. The larger the family of a word, the longer it took to recognise in auditory lexical decision. In other words, it appears that morphological family members do not aid the activation of the target, but instead compete for recognition with the target in the auditory modality. This contrasts with the facilitatory effects reported in the literature (see section 2.5, though the effect has previously mostly been observed for visual word recognition) and the facilitatory effect observed in Experiment 2 (see section 6.3.4). This cannot be an issue of the Danish families being much larger than in those languages for which family effects have previously been found (excepting Finnish), since the same items and families were used in the visual experiment and similar or larger ranges in the following experiments.

The top panels of fig. 5.4 are based on the analyses of all items, but the same pattern appears in the analyses of the complex items, see tables 5.3 and 5.5. In the error analyses, the family effect was not significant for the auditory modality: Although a large family delays the response relative to items with smaller families, it does not make an error response more likely. In the visual modality, the inhibitory component for the largest families disappeared in the error analyses, leaving a simple facilitatory effect with fewer errors for words with larger families.

The family effects shown in the top panels of fig. 5.4 raise two related questions: Firstly, why is the auditory family size effect exclusively inhibitory, with no initial facilitation? And secondly, why do the auditory and the visual experiments show different family effects although the same items were involved?

The nature of the apparent competition effect in the auditory modality was investigated by splitting the families into those family members which were of higher frequency than the target and those (the majority) which were of lower frequency. If the family effect is a competition effect, one could expect it to be primarily driven by the strongest members of the family, i.e. those which are more frequent than the target. Meunier and Segui (1999a) found that words with a large number of higher-frequency family members were recognised more slowly than those with fewer family members of higher frequency, though it should be noted that Meunier and Segui's competitors are only a subset of the family as defined here. In the present experiments, the distinction between family members with higher vs. lower frequency did not confirm Meunier and Segui's result: the number of lower-frequency family members was a much better predictor than the number of higher-frequency family members, probably because the former most closely approximates the full family size. In other words, this comparison did not result in any further insights into the nature of the supposed family competition effect in the auditory experiment. The larger morphological families in Danish could make the frequency rank of the target in its family less distinctive than it is in the smaller families in French, resulting in this difference between the French results of Meunier and Segui and the present result from Danish.

Another approach is to consider the semantic coherence of the family: if the family effect is generally facilitatory because the family members activate appropriate semantic representations, then semantically less coherent families should be less facilitatory, or even more likely to be inhibitory. Large families are likely to be less semantically coherent than smaller ones. Moscoso del Prado Martín et al. (2004a) found that although there was a facilitatory effect of the very large Finnish families on the recognition of complex words, this was predominantly carried by the subset of the stem family which consisted of family members of the whole word, what Moscoso del Prado Martín et al. term the dominated family. Moscoso del Prado Martín et al. do not claim that the dominated vs. non-dominated contrast is explanatory in itself, but they argue that the dominated family provides an index of the semantically most closely connected part of the whole family. In the analysis of the present auditory experiment, it was the non-dominated family, i.e. the semantically more distant members, that were responsible for most of the inhibition; however, the normal family size measure was still the strongest in terms of variance explained. The same pattern applied in the visual experiment, where the family effect was u-shaped rather than inhibitory. The fact that the non-dominated family accounted for most of the inhibition could be an effect of the fact that the correlation between the full family size and the non-dominated family size was almost perfect ($R^2 = 0.96$), while the correlation between the full family size and the dominated family was much smaller ($R^2 = 0.09$) though still significant. In sum, the finding from Finnish that the non-dominated family is less facilitatory than the dominated family members is replicated in Danish, but the distinction does not help explain the difference between the modalities.

A reason for the difference between Experiments 1a and 1b could be the pace of the experiments: The fixed ISI in the auditory Experiment 1a slowed down the experiment, which could mean that an initially facilitatory effect of morphological family in the early stages of word recognition could become inhibitory if slower responding increases the activation of the family members so that they become competitors to the target. However, since the two experiments differ both in modality and ISI, it is impossible to determine whether the ISI is indeed the cause of the difference. Therefore, ISI was directly manipulated between two auditory experiments, 3a and 3b (see chapter 7), but these failed to show any significant family size effects at all. This suggests that the different paces of the present auditory and visual experiments are not the reason for different family effects after all.

That words are distributed in time in the auditory modality and in space in the visual is a fact often appealed to when explaining differences in results between the two modalities. In the visual modality, the distribution in space means that it should be immediately apparent that a number of the family members are not compatible with the input, with the result that the family members are not direct competitors, but instead aid the activation of the target through semantic overlap, at least up to a certain point. However, this also applies to some extent in the auditory modality: When hearing a word, none but the onset-aligned members of the family should enter into the direct lexical competition process. Since the family members that are onset-aligned, namely the morphological continuation forms, had a facilitatory effect in both modalities (see further below), this difference between the modalities does not help explain the difference between the results.

Morphological family size effects have been demonstrated in a range of languages, but mainly for visual word recognition. These studies have shown facilitatory effects of family size or non-linear effects similar to that observed in the present visual Experiment 1b, though inhibitory effects may also found under certain conditions (Colé et al., 1997, see section 2.5). For auditory word recognition, demonstrations of morphological family size effects are much scarcer: as mentioned, Meunier and Segui (1999a) found an inhibitory effect of a morphological family measure which included only affixed forms, probably at least partly because compounding is less extensive in French than in Danish (Josefsson, 2005). An effect of the morphological family size as defined here has only been found in auditory word recognition by Baayen et al. (2002) and by (Wurm et al., 2006) for Dutch simple and complex words, respectively; in both cases, the effects were facilitatory. In contrast, (Baayen et al., 2007) found effects of family size in a visual but not a parallel auditory experiment. This means that the family effect in the auditory modality is in conflict with fewer studies than appears if considering both modalities.

A partial explanation for the inhibitory component being stronger for the morphological family effects in Danish than in other languages could be that another measure is accounting for the facilitatory effect of family activation. A good candidate for this is the number of morphologically related continuation forms of the word, both because the continuations are semantically more closely related to the target and to each other, and because the stem occurs in the same position in the signal in the continuations as in the target. The latter fact is arguably more relevant in the auditory modality, but the effect of continuations was significant in both modalities. The type count of continuation forms is to some extent a subset of the traditional morphological family, but not entirely contained in the morphological family since the continuation forms also include inflected continuations. This difference aside, the continuation forms do clearly constitute a more closely related set of morphological family members and are therefore more likely to show facilitatory effects. The fact that the set of continuations are partly contained in the set of family members makes it useful to residualise the morphological family size from the number of continuations; the residualised variable can be understood as an index of the part of the morphological family which is not onset-aligned with the target and in that way misfit the signal. The effects of morphological family remain stable in the analyses when morphological family size is replaced with the residualised measure, indicating that the inhibition observed is caused by the words that are not onset-aligned with the target.

In the visual modality, both the All Items Analysis and the Complex Analysis showed significant facilitatory effects of the number of continuation forms. This effect is shown in the bottom left panel of fig. 5.4. In the auditory modality, the effect was only significant in the Complex Analysis; this effect is shown in the bottom right panel of fig. 5.4. In the All Items Analysis, an interaction between complexity type and continuations approached significance, with marginally significant facilitatory effects for the complex words. The effect is shown in the bottom right panel of fig. 5.4: the more continuations a word has, the faster it is recognised, so this effect does constitute a facilitatory family effect. The effect was also significant in the error analyses where it interacted with frequency as discussed in section 5.3.3 above.

The family size is the type count of morphological family members; an alternative predictor is the corresponding token count, the family frequency. The type count was the strongest predictor in all analyses, in line with previous results, e.g. Schreuder and Baayen (1997) and Ford et al. (2003), but the two measures are highly correlated, in general and in this dataset, making it impossible to disentangle the contributions of the size vs. the frequency of the morphological family in these experiments. A similar case holds for the continuations: here the type count was a better predictor than both the token count and the entropy calculated across the continuations, but the materials were not designed to separate the effects of these measures.

In sum, the inhibitory effect of morphological family size effects in the auditory experiment remains difficult to understand, while the u-shaped effect in the visual experiment is more in line with previous studies. The issue is discussed further in sections 6.3.4 and 7.3.2 and in chapter 8.

5.3.5 Uniqueness points

The present analyses include the two different uniqueness points, UP1 and CUP, that were introduced in section 2.3.2 and illustrated in fig. 2.2 on p. 33. UP1 is the point where competition from morphologically unrelated competitors is resolved, the CUP or Complex UP where competition from related competitors is resolved. For a suffixed word like *accept-ere* ([ɑgsɛbˈteʔʌ], 'to accept', literally the noun 'accept' plus a verbalising suffix), which was also used as an example in section 2.3.2, UP1 is the phoneme [ɛ]. At this point, *acceptere* deviates from all other words in the language except continuation forms of the stem *accept*; the competitors up to this point include morphologically unrelated words like *aksiom* ([ɑgsiˈoʔm], 'axiom'). The CUP occurs at the phoneme [e], where the whole word *acceptere* deviates from other forms of the stem *accept* such as *accept-abel* ([ɑgsɛbˈtaʔbəl], 'acceptable'). UP1 indexes competition from morphologically non-related words and CUP competition from morphologically related words. Similarly for a particle prefixed word like *op-takt* (['ʌbtɑgt], 'upbeat'), UP1 is the [ɑ] where the target deviates from overlapping words, like *optiker* (['ʌbtigʌ], 'optician') which shares the onset but in which the onset does not function as the particle prefix *op-*. The CUP of *op-takt*, is the next phoneme, [g] where *op-takt* deviates from words sharing the prefix *op-*, for instance *op-tag-er* (['ʌbtɑʔ], present tense of the verb 'record' or 'take up').

For suffixed words, UP1 corresponds to the traditional UP; for prefixed words, it is the CUP that corresponds to the traditional UP. This means that for a mixed set of complex stimuli like the present, the traditional UP is no longer a single entity. However, the fact that both new UPs show significant effects independently of each other indicates that there are different stages of competition in the recognition of morphologically complex words, making it advantageous to operate with two different UPs. Moreover, they account for these two stages of competition in a uniform way across the different types of complex words.

Generally, UP1 either precedes the CUP or the two coincide. In the present dataset, there are two words where the CUP (the uniqueness from related competitors) precedes UP1 (the uniqueness from non-related competitors), namely *van-røgt* ('neglect', literally 'mis-care') and *van-skæbne* ('unhappy fate', literally 'mis-fate'). These two words pose a problem in relation to the collinearity of the UP and word duration measures. The collinearity between the two UPs and the raw duration of the words can be drastically reduced by measuring the two UPs and duration as three non-overlapping parts of the signal: the distance from word onset to UP1, the distance from UP1 to CUP and the distance from CUP to word offset. However, if two values of UP1 to CUP distance are negative, it becomes impossible to use the logged values for this variable, in turn making its distribution highly non-normal. Therefore, the CUP for the two words in question was shifted to coincide with the UP1. The CUP is then the point where all competition has been resolved, though in almost all cases, the competitors that are compatible until just before the CUP are morphologically related. In the definition of Wurm (1997), these two words would be CRUP words, even though the distinction between UP and CRUP is not identical to that between UP1 and CUP, see section 2.3.2.

This distinction between UP1 and CUP applies only to the complex words, since for simple words, continuation forms of the stem, which constitute the competition cohort on which the CUP is based, are also continuations of the whole word and must therefore be excluded from UP-calculations. However, for purposes of control, both uniqueness points were included in the analysis of all items, with the value of the distance from UP1 to CUP set to zero for the simple words, although it is not ideal to operate with a variable that has so many zero-values. The real test of the two UPs is the analysis of the complex items. The inclusion of both UPs in the All Items Analysis is further motivated by the fact that the complex types differ with respect to the correspondence between the new UPs and the traditional UP, making it difficult to determine which of the UPs for the complex words should be included in the All Items Analysis.

This section presents and discusses first the auditory experiment, where the uniqueness point effects are the largest and most robust effects. Then the smaller but still significant effects of UP1 and length in the visual modality are discussed.

Auditory experiment

In the auditory modality, both uniqueness point measures showed significant inhibitory effects, with longer RTs for words with later UP1 and CUP.

In the All Items Analysis, the effect of UP1 interacts with complexity, as also discussed above. The effect is shown again in the top left panel of fig. 5.5. The simple words are illustrated by the dashed line and show no effect for relatively early UP1s, but strong inhibition for later UP1s. The complex words, shown by the solid line, show a linear inhibitory effect of UP1. For both types of words, the UP1-effect reflects competition from non-related cohort members.

The initial flatness of the UP1-effect for the simple words could indicate that a certain amount of input is required for a lexical decision to be made; this could reduce the benefit of a very early relative to a slightly less early UP1. This initial flatness is absent for the complex words. In stress-timed languages like Danish (cf. section 3.2), the first constituent is sometimes compressed in complex words relative to simple words (Davis et al., 2002). This may in turn mean that more information is available in complex words with very early UP1s in ms, than for simple words with similar UP1s in ms, resulting in no initial flatness for the complex words. The difference could also be caused by the complexity advantage described above: when competition from unrelated words is resolved early, as indicated by an early UP1, the morphemic information may play a stronger role in the recognition of the complex words, making the benefit of an early UP1 stronger for complex than for simple words.

Competition from morphologically related competitors also had an inhibitory effect on the recognition of the complex words, as evidenced by the significant effect of the distance from UP1 to CUP, which was used as an index of CUP in order to reduce collinearity. This measure was included in both the All Items and the Complex Analyses, in order to control this source of variation between the complex words, but for the simple words in the All Items Analysis, the value of this variable is zero. This means that the real test of the CUP-effect is the Complex Analysis, which showed a

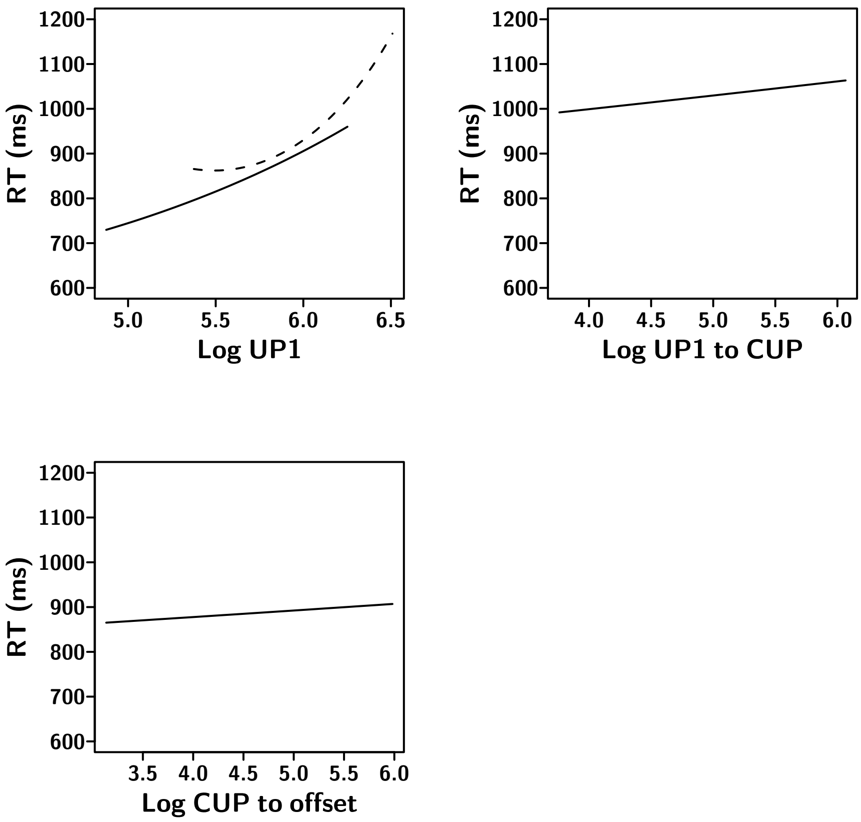

Figure 5.5: Partial effects of uniqueness points and duration for the auditory Experiment 1a. To reduce collinearity, CUP is measured as distance from UP1 to CUP and duration as distance from CUP to offset. The units on the horizontal axes are log ms. The dashed line in the top left panel shows the effect of UP1 for the simple words, the solid line the UP1-effect for the complex words. Note that the ranges of the horizontal axes in the top right and bottom left panels are restricted to exclude a small number of outliers with a value of the relevant predictor of zero.

clear inhibitory effect of UP1 to CUP distance for all three complex types, as shown in the top right panel of fig. 5.5. The inhibitory effect of UP1 to CUP distance is a result of competition from morphologically related competitors.

The new UPs reflect a division between competition from unrelated and related words. If such a division is real, no differences between the different affix types are expected in terms of the effects of the new UPs. The Complex Analysis clearly confirms this: there were no significant interactions between affix type and either of the UPs. The fact that the types differ with respect to the correspondence of UP1 and CUP to the traditional single UP shows no effect.

The bottom panel of fig. 5.5 shows the effect of the distance from CUP to word offset, which indexes word duration. The effect was inhibitory, such that longer words elicited longer RTs. The effect is shown for all items in the bottom panel of fig. 5.5. It was not quite significant in the Complex Analysis and therefore not included in the final model summarised in table 5.3; this is probably an issue of statistical power.

In the error analyses, only the CUP showed a significant effect in the analysis of errors on complex items, see table 5.7, such that there were more errors for word with later CUP. There were no UP-effects in the All Items Error Analysis, see table 5.6.

The inhibitory effects of both UPs show the harmful effects of competition: the longer both non-related (measured by UP1) and related words (measured by CUP) are consistent with the input, the more difficult it is to recognise the word. The new information provided by this new contrast between two UPs is that this competition comes not only from entirely separate lemmas that are considered for the traditional UP for simple and suffixed words, but also from words that share one constituent with the target. The fact that both UPs affect recognition indicates that listeners are sensitive to the morphological structure of the complex items. Furthermore, the significant effect of the CUP shows competition between those second constituent candidates that are possible given that the first constituent has been processed; in other words, the CUP-effect suggests a similar kind of conditional processing as that shown by the effects of CRUP (see section 2.3.2). The division is further explored for suffixed words in Experiment 2 and for prefixed, particle prefixed and compounds words in Experiment 3. In contrast to the continuation forms of the first constituent, morphologically related words that overlap with the whole target, i.e. the continuation forms, support the recognition of the target as described in section 5.3.4 above.

Visual experiment

In the visual experiment, only UP1 was significant, and the effect was much smaller than the uniqueness point effects in the auditory experiment. However, it is not surprising that the effect was smaller in the visual than in the auditory modality, rather what is surprising is that this auditory measure was significant at all in the visual experiment. UP1 showed a significant effect in both the Visual All Items Analysis (see table 5.4) and in the Visual Complex Analysis (see table 5.5). The interaction between complexity type and UP1 in the All Items Analysis indicates that the inhibitory effect of UP1 was only significant for the complex words. The effect is shown in the left panel of fig. 5.6, where the solid line represents the inhibitory UP1-effect for the complex words, while the dashed line represents the simple words,

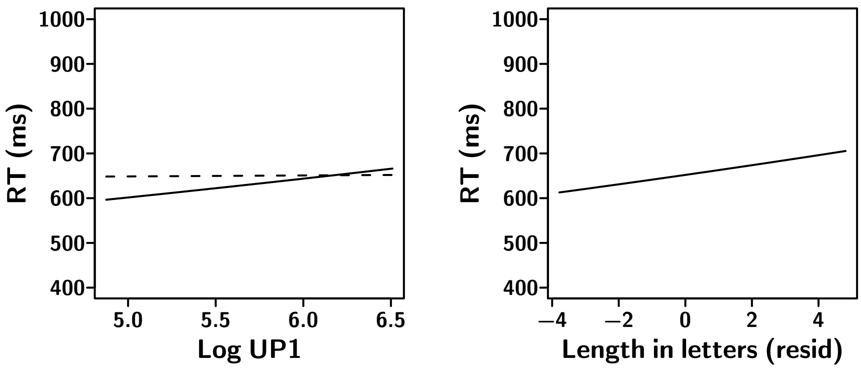

Figure 5.6: Partial effects of UP1 and length in letters for the visual Experiment 1b. In the left panel, the dashed line represents the simple words and the solid line the complex words.

for which there was no significant effect of UP1. In the error analyses, only the complex words show an effect of UP1, confirming the effect found in the RT-analyses.

The difference between the complex and the simple words could partly be due to the higher probability for the complex words of being read in two fixations, triggering some element of left-to-right processing — on which the UP-effects rely — also in the visual modality. However, even the complex words are not long enough to systematically require two fixations, so this is perhaps not the sole explanation. Another explanation could be that complex words that have an early UP1, i.e. become unique from unrelated competitors early, benefit more from morphemic information during recognition than complex words with a later UP1, resulting in this UP1-effect for the complex words.

The effect of UP1 indicates some degree of auditory recoding of the visual stimulus as part of the recognition, in accordance with the reading model of Lukatela, Eaton, Lee, Carello and Turvey (2002), at least for the complex words. The effect of UP1 in log ms in the visual experiment was not an artefact of a visual UP1, defined as the letter at which the written word deviates from all other written words in the language (except, as usual, its own continuation forms) and counted as the position in the string of the UP1-letter (following Baayen et al., 2007). The visual UP1-measure was neither predictive on its own, nor did it affect the significance of the auditory measure. This confirms the idea of auditory recoding of the visual stimulus in reading.

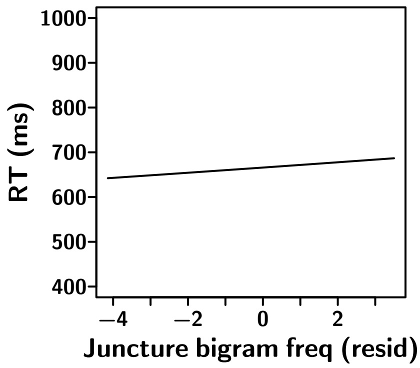

Figure 5.7: Partial effect of the frequency of the bigram at the morpheme juncture of the complex words for the visual modality.

5.3.6 Junctural probability

Turning to the effect of length, the measure used to index this was length in letters, decorrelated from whole-word frequency. The effect of length was inhibitory, with longer words eliciting longer RTs, as shown in the right panel of fig. 5.6. As could be expected, the error analyses showed no effect of length: longer words were not more error-prone.

A final significant predictor is the frequency of the bigram that straddles the juncture between stem and affix in the complex words. The bigram juncture frequency is an index of how marked or salient the juncture is: the higher the frequency of the bigram, the more likely that bigram is to occur inside a word, and the less marked the juncture is as a juncture. Conversely, the lower the juncture bigram frequency, the more salient the boundary and the more likely it is to be perceived as a morpheme boundary (Hay, 2002). The frequency of the bigram crossing the juncture is thus inversely related to the juncture probability.

In the visual experiment, the frequency of the juncture bigram had an inhibitory effect, such that words with less frequent bigrams at the juncture were recognised faster than words with more frequent juncture bigrams. This means that words with more marked junctures, indexed by low-frequency juncture bigrams, were easier to recognise. The effect is illustrated in fig. 5.7. It seems that the more clearly the constituent morphemes can be parsed and perceived, the more they contribute to the recognition. This could be understood as a reflex of an advantage for words with several sources of information, both morphemic and whole-word based, that is also evident in the advantage for complex over the simple words. The significance of the effect confirms the off-line findings of Hay (2002) and is similar to the effect observed for derived words by Baayen et al. (2007). It should be noted that the effect size is relatively small.

The fact that this measure is only significant in the visual modality is presumably due to the fact that the measure is orthographic, and thus tied to the visual modality in a relatively deep orthography like the Danish. A parallel effect of diphone frequency would be predicted for the auditory modality, but the available resources do not allow the testing of this. Apparently, the bigram frequency measure is not a good enough index of diphone frequency to reach significance in the analysis of the auditory experiment.

A couple of related visual measures did not show significant effects, not even in the analysis of the visual experiment. The mean frequency of the bigrams of the items did not show significant effects. The same holds for the orthographic neighbourhood density (Coltheart's N) and the embedded and embedding entropies (see sections 2.3.1 and 3.4.5).

5.4 Summary and conclusions

Since several of the results are further investigated in the subsequent experiments, the wider implications of the present findings are better understood after these experiments have been presented and are therefore discussed in chapters 6 to 8. An exception is the comparison of auditory and visual results which is most relevant here because the next experiments are exclusively auditory. This section first sums up the main findings of the experiments presented in this chapter, before comparing the results of the auditory and the visual experiments.

5.4.1 Main results

The present experiments confirmed the importance of drawing in effects of experimental context, namely the RT and correctness on previous trials as well as general training or fatigue effects. These variables had strong and significant effects, thus bringing under statistical control a considerable amount of variance that would otherwise be noise in the data. The possibility of controlling these variables constitute one major advantage of employing a regression design.

The predicted complexity advantage was found in both modalities: the complex words were recognised significantly faster than the simple words. Complexity type interacted with UP1, such that the advantage was mainly found when UP1 was early. This suggests that early uniqueness from unrelated competitors, indexed by UP1, allows morphemic information to contribute to the recognition of the complex words.

The complexity advantage was stronger for the suffixed than for prefixed and particle prefixed words, a difference also reflected in the effect of affix type in the Auditory Complex Analysis. The suffixed words were recognised faster than both types of prefixed words which could be due to the fact that the stem carries more information, resulting in a processing advantage for words in which the stem is encountered first. Moreover, the large and diverse word-initial cohorts of both types of prefixed words could inhibit the recognition of prefixed words. The fact that the particle prefixed words were recognised more slowly than both other types of complex words was explained in terms of the possible lower semantic transparency of the particle prefixed words and the possibility that the particle as an independent word functions as a competitor in auditory word recognition.

The complexity advantage indicates that information from both the whole word and the constituent morphemes contribute to the recognition of the complex words. This interpretation is supported by the presence of facilitatory frequency effects for both the whole word and the stem. In the auditory modality, these two frequency effects interacted, such that the whole-word frequency effect became attenuated for words with higher stem frequency, disappearing for words with very high stem frequency. Correspondingly, stem frequency was most facilitatory for words with the lowest whole-word frequency, and inhibitory for words with higher whole-word frequency. This indicates that whole-word and morphemic sources of information are not used independently, an effect replicated in Experiment 2 and further discussed in section 6.3.3. The stronger frequency effects for complex than for simple words in Experiment 1a suggests an added advantage from the fact that whole-word frequency can function as a conditional probability for the complex words.

While the u-shaped family effect in the visual experiment is in line with recent evidence, the inhibitory effect of family size in the auditory modality is surprising, both in view of previous evidence and in comparison with experiments 2 and 3a and 3b. A number of possible reasons for this finding are explored in Experiments 3a and 3b and in the comparison of the experiments in chapter 8. One possible factor contributing to this result is the facilitatory effect of the number of morphologically related continuation forms, which may be accounting for the facilitatory effect of morphological family in the auditory experiment.

Further, two different uniqueness point measures (introduced in 2.3.2) were investigated in the analyses. In the auditory experiment, the effects of both UP1 and Complex UP were significant, showing the sensitivity of the processing system to different cohorts at different points in the lexical competition process, with UP1 indexing competition from non-related words and CUP competition from related words. The roles of UP1 and CUP are further explored and confirmed for suffixed words in Experiment 2, see 6.3.5.

Finally, the visual experiment confirmed, though not strongly so, that words with more marked junctures between the stem and the affix were recognised more easily than those with less marked junctures, as evidenced by the inhibitory effect of juncture bigram frequency in the visual modality. This indicates that not only the strength of the stem, indexed by the stem frequency effect, but also its parsability in the complex word affects the recognition of the word: The clearer the boundary between the morphemes, the more the morphemes can contribute to the recognition, and the faster the word can be recognised.

5.4.2 Auditory and visual processing

There were two purposes of running both an auditory and a visual version of this experiment: Firstly, since most studies of morphological processing investigate the visual modality, the use of the same items in both an auditory and a visual experiment allowed some assessment of whether differences relative to the predominantly visually based literature were due to the language being Danish or to modality differences. Secondly, parallel studies of visual and auditory recognition of the same words are not common (though see Baayen et al., 2007 and Taft, Hambly and Kinoshita, 1986), but they may illustrate important differences and similarities in word recognition between the two modalities.

Generally, the overlap between the auditory and visual modality was substantial, with effects of the different frequencies, complexity types, family measures and UPs showing up in both experiments. However, there was a clear difference between the modalities in terms of the family effects, with linear inhibition in the auditory modality and inhibition only for the larger families in the visual modality. This is a case where the conclusion would have been different without the visual results, namely that the family size effect is inhibitory in Danish. As it is, it is clear that there are also modality differences, although it is unclear how the family effect difference could relate to the differences between auditory and visual word recognition. The difference between the modalities would probably follow if the inhibitory auditory effect in this experiment and the difference to the facilitatory effect in Experiment 2 could be explained with reference to the ordering of the constituents. The ordering of the constituents could play a role if the stem family is facilitatory when the stem is initial as in suffixed words and inhibitory when the stem is final (and generally unstressed) as in prefixed words. This is further discussed in chapter 8. The effect of the other family measure, the number of morphologically related continuation forms, was facilitatory in both modalities and similar magnitudes.

The differences between the pattern of context effects seems to be more a matter of inter-stimulus interval than of modality: the slower auditory experiment shows a weak fatigue effect, while the visual experiment shows stronger sequential effects, with a significant effect of whether the previous response was correct or not.

All frequency measures, including family and junctural probability, were based on the corpus of written Danish, as was the neighbourhood density; however, some of the measures approximated spoken language better than others. The most clearly visual measures were mean bigram frequency and orthographic neighbourhood measures, which were non-significant for both experiments, and the junctural bigram frequency which was only significant in the visual experiment. This is unsurprising given the relatively deep orthography of Danish. The most phonological measures, namely the uniqueness points, were significant in both modalities, though weaker for the visual experiment. This is in line with the results of Baayen et al. (2007) and suggests that auditory memory representations are activated during visual processing.

Chapter 6

Experiment 2

Results and interpretations of Experiment 2 are also reported in Laura Winther Balling & R. Harald Baayen (2008), Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, in press. The dataset is available as "danish" in the languageR function library of Baayen (2008b).

6.1 Introduction

This experiment investigates the recognition of inflected words in Danish and compares inflected with derived and simple words. The comparison of all three complexity types establishes to what extent the complexity advantage found in Experiment 1 (see section 5.3.2) generalises to other items. More specifically, the present experiment investigates whether the complexity advantage holds for inflected words as well as derived words in comparison to simple words. Results of factorial studies of Finnish suggest that this may not be the case (see section 2.6.1), but it is an open question whether this also applies in a language with a less rich morphology and in a regression study that allows the choice of relatively typical examples of the different categories.

In addition, the design allows the comparison of regular inflected and less regular derived words. Some researchers (e.g. Pinker, 1997a; Clahsen, 1999; Ullman, 2001, see sections 1.3 and 2.4) argue that regular morphological forms are processed in a fundamentally different way from all less regular forms, including both most derivational forms and irregular inflection. The hallmark of this rule-based processing of regular complex words should be clear stem frequency effects and no whole-word frequency effects for regular words, because at least low-frequency regular words lack whole-word memory representations. Also in theoretical morphology, a categorical distinction between inflected and derived words is often posited (see section 2.6.2). The question posed by the present experiment is whether a categorical distinction between inflected and derived words is a stronger predictor of word recognition than the graded variables that differ systematically between the two classes, such as semantic transparency and affix productivity. Based on both theoretical considerations and previous evidence, the hypothesis is that the effects of the graded differences will outperform the categorical difference between inflected and derived words, but a categorical effect may still be observed.

These comparisons aside, the inclusion of inflected words broadens the perspective of this work, both because these constitute a somewhat different area of lexical organisation in Danish, and because the question of how inflected words are recognised is strongly debated. One set of lexical variables that are of particular interest in this connection is whole-word vs. constituent frequencies, which are also investigated here. Generally, much emphasis is placed on stem frequency as the main diagnostic of decompositional, morpheme-based processing, but since many studies fail to find effects of stem frequency (see section 2.4), alternative measures should be sought. One such alternative is the converse constituent frequency, the frequency of the affix, which is investigated here. The inclusion of both inflected and derived words results in a relatively wide range of affix frequencies which should be ideal for obtaining an effect of this measure.

A further main objective of this experiment was to confirm the role of the two different uniqueness points that were introduced in section 2.3.2 and experimentally documented for a set of different complex words in Experiment 1 (see section 5.3.5). The suffixed items of the present experiment are more uniform with respect to their relation to the traditional UP than the mixed prefixed and suffixed items of Experiment 1.

Finally, a question arising from the results of Experiment 1 is what the role of morphological families is in Danish word recognition. The auditory Experiment 1a suggested that larger morphological families make word recognition slower, in contrast to previous evidence. Hence, the words chosen for the present experiment had similar morphological family sizes to those in Experiment 1, in order to investigate whether the inhibitory effect, which is surprising given effects reported in the literature, is replicated. The semantic transparency ratings that were obtained for purposes of comparing inflected and derived words could also inform the understanding of the morphological family size effect.

6.2 Method

6.2.1 Materials

90 derived, 70 inflected and 72 simple words and 232 pairwise matched nonwords were presented in an auditory lexical decision experiment. The three categories were comparable in terms of the ranges of morphological family size, while duration and UPs necessarily varied between the categories, because simple words are generally shorter than complex ones, and most of the inflected words shorter than most of the derived. In terms of lexical frequencies, the word frequencies of the simple words were comparable to the stem frequencies of the complex words, but somewhat higher than the whole-word frequencies of the complex words. This is a natural consequence of the fact that simple base forms tend to be more frequent than inflected forms and, especially, derived words. While the stem frequencies of inflected and derived words were similar, the whole-word frequencies of the inflected words were higher than those of the derived, reflecting the fact that for whole words of a given frequency, most of the inflectional forms are likely to be more frequent than most of the derived forms. The complex items were chosen to vary in relative stem and whole-word frequency. The lexical variables are summarised in table 6.1; all items are listed in Appendix B.

As for the items in Experiment 1, the following characteristics were avoided: highly irregular spelling-to-sound mappings, pronunciations varying

Table 6.1: Means, standard deviations and ranges of lexical variables of all items in the top table and of complex items alone in the bottom table. Whole-word frequency refers to the frequency of the base form of the whole word, stem frequency to the lemma frequency for the stem. These continuous variables were converted to logarithmic scales for the statistical analyses, in order to reduce their skewness, but the non-transformed values are shown here for interpretability.

All items ($n = 230$)

	Mean	SD	Range
Whole-word frequency[a]	23	58	0 to 495
Morphological family size	353	425	1 to 2 165
Number of continuation forms	62	164	0 to 1 390
Neighbourhood density	3.9	5.4	0 to 28
Mean bigram frequency[a]	31 564	17 927	2 401 to 74 478
UP1, ms	315	84	154 to 633
Length, ms	548	114	186 to 820

Complex items ($n = 157$)

	Mean	SD	Range
Whole-word frequency[a]	17	46	0 to 388
Stem frequency[a]	70	118	<1 to 652
Affix token frequency	7 943	9 100	199 to 29 328
Affix type frequency	320	712	1 to 2 337
Morphological family size	369	454	1 to 2 165
Number of continuation forms	10	40	0 to 458
Semantic transparency rating	4.9	1.6	1.3 to 7.0
Neighbourhood density	2.0	2.8	0 to 19
Mean bigram frequency[a]	36 566	17 624	9 984 to 74 478
Juncture bigram frequency[a]	36 042	27 684	24 to 160 470
UP1, ms	296	75	154 to 478
Complex UP, ms	467	95	261 to 682
Length, ms	594	91	364 to 820

[a] Values for these variables are counts per millon in *Korpus90/2000*.

much between casual and careful speech, and pseudo-cases of the affixes used on the complex items. The complex words were bimorphemic and had free stems, avoided linking morphemes and stem allomorphy, with the exception of regular stress- and *stød*-changes, schwa-deletion and regular coarticulatory phenomena like assimilation of /n/ to /m/.

The 70 inflected words carried seven different affixes, all of which were chosen to represent the default inflectional form for the given function. There were four verbal, two nominal and one adjectival affix (see Appendix D for a list of the affixes used).

There were 90 derived forms, with nine affixes occurring on ten words each. Because all Danish inflections are suffixations, the derivational affixes used were all suffixes, in order to enable comparison between the two categories. Seven of the derivational suffixes were also used in Experiment 1 (see Appendix D). Productive derivation in Danish mostly forms nouns and adjectives rather than verbs (which are most often formed by conversion), so the nine derivational affixes were one verbalising, three nominalising and five forming adjectives (some of which can also function as adverbs). The derived words differed in semantic transparency, which was rated in a pre-test, see below.

Each affix was represented on ten different items in the experiment. There are more items per affix in this experiment than in Experiment 1, both because Experiment 1 showed no evidence of affix priming effects, and because the debriefing of participants after Experiment 1 suggested that what was consciously noticed was the nonwords rather than the morphological structure of the items. The effect of this increased number of presentations of each affix was assessed by including a variable indexing affix repetition.

Affixal homonymy is extensive in the Danish inflectional system, at least partly because all forms are phonetically relatively weak, with schwa as the most frequent vowel. Therefore, affixal homophony could not be avoided for the inflected words, with five out of seven inflectional affixes but none of the derivational affixes being homophonous with other affixes. However, a variable indicating whether or not each affix was homophonous had no main effect and entered into no interactions. Moreover, given the stems they occur with, the affixes were unambiguous, with two sets of exceptions: Five of the items carrying the verbal past tense suffix *-ede* could also be interpreted as the plurals of adjectivally used past participles of the same verbal stems, though the past tense form intended is the dominant one in all cases. The verb *lov-ende* ('promis-ing') is homophonous with one possible pronunciation of the definite noun plural *lov-e-ne* ('the law-s'), but again the intended form is the more frequent. The pattern of results remained the same when these potentially slightly ambiguous words were excluded. Neither the limited whole-word homophony nor the more extensive affix homophony for the inflectional forms constitutes a confounding factor in this experiment.

Derivational suffixes tend to be phonologically more salient (in terms of length and vowel quality) and to have more clearly marked morpheme boundaries than inflectional suffixes. Longer derivational affixes were avoided in order to alleviate the problem of differences in salience, while stems ending in vowels or semi-vowels like /r/ and /w/ were avoided for the inflectional forms to alleviate the problem of juncture strength differences. The fact that these differences could not be controlled entirely becomes less problematic because they work against the hypothesis of no categorical difference in processing between inflection and derivation. As in Experiment 1, the frequency of the bigram at the stem-affix juncture was included in the analyses, as the best available index of juncture salience, but no significant effects of this were found.

The simple group encompassed 72 words, roughly equally divided between verbs, nouns and adjectives to reflect the approximate balance between the three word classes for the complex words.

232 nonwords were constructed, in a manner similar to Experiment 1. The affixes were always retained on the nonwords. Additionally, there were 30 training and warm-up items with a similar composition to the items in the experiment.

Two items were not included in the analyses: one definite noun, *film-en* ('movie-the'), because the affix was almost inaudible, and one present participle, *storm-ende* ('storm-ing'), because of homophony with the plural form of the related noun. *Storm-ende* is more problematic than the homophonous items mentioned above, because the *-ede* forms were still forms of the same verb, while the intended verb *lov-ende* was only homophonous with one of two possible pronunciations of the noun *lov-e-ne*. Therefore, *storm-ende* was removed, while the other homophones were retained in the analyses reported in tables 6.2 to 6.5. However, as mentioned, the pattern of results remained the same when all potential homophones were excluded. Additionally, one derived item, *nævn-ing* ('juror'), was removed from the analyses of the complex items only, because an incorrect stem form was used in the semantic transparency questionnaire, making the semantic rating invalid.

Semantic transparency

The semantic transparency of all complex words was investigated using a questionnaire, applying the procedure of for example Wurm (1997). Participants were asked to rate the semantic relatedness of the complex words and their stems, on the assumption that this reflects the perceived semantic transparency of the complex words, while avoiding any technical terminology. Seven-point Likert scales were used for the rating, with '1' used for words whose meanings were not related at all and '7' indicating closely related meanings. A pseudo-suffixed word and its pseudo-stem (*hov* ('hoof') and *hovere* ('to gloat') in which *-ere* looks and sounds like a verbalising suffix, but does not function as such) were used as an example of two non-related meanings and a noun and its plural form as an example of two meanings that are closely related (*hest* and *hest-e*, 'horse(-s)'). Respondents were instructed to focus on the meaning rather than the form of the words, to try to use the whole scale, and to use the rating 1 (= not related) to mark cases where they did not recognise one or both words, based on the assumption that, out of context, such unknown words are exactly opaque.

Participants were recruited via a weblog (unrelated to the research project) and the questionnaire was run on-line (on http://www.hostedsurvey.com/ in January 2007). The questionnaire took between 20 and 30 minutes to complete. Only completed questionnaires were used; these were filled out by 22 women and 14 men between the ages of 20 and 54 (mean 32.9 years). None of the respondents participated in any of the other experiments. Two completed questionnaires were discarded because the respondents had grown up as bilinguals; all other participants had Danish as their first language.

Two versions of the questionnaire were constructed: version A had half the pairs occurring with the stem first and half with the complex word first, and version B with the opposite order. Excepted from this were cases where the stem was homonymous with another word, for which the whole word always occured first. Each participant received a random order of the pairs. The word pairs that were rated are listed in Appendix B.

The ratings were averaged for each item and the resulting mean was used as the semantic transparency measure in the analyses reported below.

6.2.2 Recording and preparation of stimuli

The procedure for recording and preparing stimuli was identical to that used for Experiment 1a, see section 5.2.2.

6.2.3 Procedure

The procedure of the lexical decision experiment was identical to that used for Experiment 1a, except that the inter-stimulus interval was not fixed. Instead, the presentation programme moved on to the next item when subjects responded or, in the absence of a response, 2500 ms after stimulus onset. This made the pace of the experiment faster than the pace of Experiment 1a. One advantage of this was technical in a much smoother running of the experiment; the other that the experiment required less time to be invested by the subjects, who were volunteers. The experiment took 15 to 20 minutes to complete.

6.2.4 Participants

The participants were 22 volunteers, aged between 19 and 37 years (mean 26.5). There were nine men and 13 women. All had Danish as their first language and reported normal hearing. The participants were asked to fill in the same background questionnaire that was used for Experiment 1; the questionnaire is reproduced in Appendix F.

6.3 Results and discussion

For the reaction time analyses, all error responses were removed; these amounted to 3.7% of responses. Additionally, two items were removed because of error rates over 30%: one very low-frequency and semantically opaque derived item, *tøj-eri*, and one relatively low-frequency simple item, *rank*. All in all, 5.0% of the responses were removed due to errors or high error rates on specific words. The reaction times were not a priori filtered, instead the regression models were trimmed to exclude data points with large standardised residuals, see discussion in section 4.4.4. For the All Items Analysis, 2.5% of correct responses were removed in this way, and for the Complex Analysis 2.8% were removed. Mean correct RT was 874 ms measured from word onset.

All responses to all 230 items were retained in the All Items Error Analysis, and all responses to all 157 complex items in the Complex Error Analysis.

As for Experiment 1, linear mixed-effects model with crossed random effects for participant and item were fitted to both reaction times and errors. The analyses tested a number of variables; non-significant variables were removed using a backwards stepwise elimination method, to reach the models summarised in tables 6.2 to 6.5, which show the same statistics as the tables in Experiment 1, see pp. 111ff. As in Experiment 1, there were two levels of analysis: one set of analyses for RT and correctness for all items, using variables that were relevant to both simple and complex words, and one set for the complex items, adding variables pertaining to the morphological structure of the complex items. In addition to the word and participant random intercepts, the Complex Analysis included random intercepts for affix and random slopes for whole-word frequency (for the latter see section 6.3.3).

Figures 6.1 to 6.5 show partial effects of the different significant variables, adjusted to the medians of other continuous predictors and to the reference levels that are specified in the models in tables 6.2 and 6.3. The same range of 600 ms is used on all vertical axes, as it was in chapter 5.

The collinearity between the lexical predictors was high, indicated by a κ of 72 for the variables in the All Items Analysis and a κ of 147 for the variables in the Complex Analysis when the variables involved were not decorrelated. However, when the same steps were taken to decorrelate the variables here as were taken in Experiment 1, the collinearity of variables in the analysis of RT to all items and both error analyses was low, with κ-values below 10. For the Complex RT Analysis, κ was only slightly higher at 13. As in Experiment 1, the UPs and duration were included as log time in ms from onset to UP1, from UP1 to CUP and from CUP to offset. Variables decorrelated by replacing the standard variable with the residuals of a decorrelating regression model are marked as “Residualised”. In this

Table 6.2: Summary of regression model for reaction times to all words (the All Items Analysis), using contrast coding for the factors Previous response with Correct as the reference level, for Complexity Type with Simple as the reference level and for Word Class with Adjective as the reference level. Df = 4713.

Fixed Effects

	Estimate	MCMC mean	HPD 95% CI Lower	HPD 95% CI Upper	p(MCMC)	p(t)
Intercept	9.8999	9.9113	6.9224	12.8035	0.0001	<0.0001
Affix repetition	-0.0042	-0.0042	-0.0059	-0.0026	0.0001	<0.0001
Previous PC1	-0.0001	-0.0001	-0.0001	-0.0001	0.0001	<0.0001
Affix repetition×PC1	0.0000	0.0000	0.0000	0.0000	0.0188	0.0220
Previous PC2	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001
Previous response: Error	0.0288	0.0287	0.0087	0.0476	0.0020	0.0030
Complexity type: Derivation	-0.2609	-0.2610	-0.3600	-0.1579	0.0001	<0.0001
Complexity type: Inflection	-0.2781	-0.2782	-0.3831	-0.1768	0.0001	<0.0001
Word class: Noun	0.0200	0.0198	-0.0016	0.0399	0.0594	0.0556
Word class: Verb	0.0287	0.0285	0.0061	0.0509	0.0116	0.0128
Log whole-word frequency	-0.0149	-0.0149	-0.0193	-0.0101	0.0001	<0.0001
Log UP1^{l}	-1.3669	-1.3709	-2.4213	-0.3594	0.0104	0.0090
Log UP1^{q}	0.1413	0.1416	0.0515	0.2331	0.0028	0.0021
Log UP1 to CUP	0.0729	0.0730	0.0531	0.0928	0.0001	<0.0001
Log CUP to offset^{l}	-0.0160	-0.0161	-0.0372	0.0036	0.1204	0.1259
Log CUP to offset^{q}	0.0051	0.0051	0.0012	0.0092	0.0116	0.0122

Random Effects

	SD
Word	0.0536
Participant	0.0641
Residual	0.1311

^{l} Linear.

^{q} Quadratic.

way, semantic transparency was decorrelated from morphological family size, morphological family size from affix and whole-word frequency, and affix frequency from whole-word frequency.

The following sections describe the significant effects of complexity type (section 6.3.2), frequencies (6.3.3), family size and semantic transparency (6.3.4) and uniqueness points and duration (6.3.5). Other variables were tested but not found significant; the most important of these are discussed below, but it should be noted here that mean bigram and juncture bigram frequencies were also checked but found to have no significant effects and therefore excluded from the final models. This is not surprising, given the orthographic nature of these measures and the results of Experiment 1a.

Table 6.3: Summary of regression model for RT to complex words only (the Complex Analysis), using contrast coding for the factors Previous Response with Correct as the reference level and for Sex with Female as the reference level. Df = 3216

Fixed Effects

	Estimate	MCMC mean	HPD 95% CI Lower	HPD 95% CI Upper	p(MCMC)	p(t)
Intercept	5.5841	5.5823	5.0605	6.0903	0.0001	<0.0001
Affix repetition	-0.0035	-0.0034	-0.0051	-0.0016	0.0001	0.0001
Previous PC1	-0.0002	-0.0002	-0.0002	-0.0001	0.0001	<0.0001
Affix repetition×PC1	0.0000	0.0000	0.0000	0.0000	0.0002	0.0005
Previous PC2	0.0000	0.0000	0.0000	0.0001	0.0006	0.0005
Previous response: Error	0.0335	0.0323	0.0090	0.0554	0.0060	0.0048
Family size^r	-0.0081	-0.0080	-0.0139	-0.0017	0.0092	0.0155
Semantic transparency^r	-0.0079	-0.0078	-0.0160	0.0000	0.0544	0.0740
Sex: Male	0.0047	0.0046	-0.0449	0.0524	0.8482	0.8770
Log whole-word freq	-0.0130	-0.0127	-0.0414	0.0147	0.3456	0.0001
Affix frequency^r	-0.0424	-0.0421	-0.0664	-0.0177	0.0018	0.0011
Sex:M×Log word freq	0.0017	0.0011	-0.0404	0.0475	0.9514	0.5914
Sex:M×Affix freq	0.0258	0.0255	0.0032	0.0468	0.0238	0.0180
Word freq×Affix freq	0.0095	0.0094	0.0046	0.0143	0.0002	0.0002
Sex:M×Wordfreq×Affixfreq	-0.0065	-0.0064	-0.0108	-0.0018	0.0078	0.0048
Log UP1	0.2511	0.2513	0.2172	0.2867	0.0001	<0.0001
Log UP1 to CUP^l	-0.1866	-0.1859	-0.3781	0.0150	0.0666	0.0798
Log UP1 to CUP^q	0.0274	0.0273	0.0071	0.0480	0.0100	0.0132
Log CUP to offset	0.0149	0.0148	0.0084	0.0214	0.0001	<0.0001

Random Effects

Groups	Name	SD
Word	Intercept	0.0443
Affix	Intercept	0.0280
Participant	Intercept	0.0637
Participant	Word freq	0.0047
Residual		0.1308

^{r} Residualised.

^{l} Linear.

^{q} Quadratic.

6.3.1 Context variables

This experiment confirms the finding from Experiment 1 of strong and significant effects of context-related control variables. There were significant effects of the two strongest of four principal components based on the reaction times to the four previous items (Previous PC1 and Previous PC2), as well as of the correctness on the immediately previous trial and of the number of times an affix had been encountered in the experiment. These effects are

Table 6.4: Summary of logistic regression model for correctness on all words (the All Items Error Analysis). Df = 5056

FIXED EFFECTS

	Estimate	SD	z	p(z)
Intercept	-3.5363	0.3810	-9.283	< 0.0001
Previous PC1	0.0006	0.0003	2.009	0.0445
Trial number	0.0016	0.0006	2.619	0.0088
Log word frequency	-0.2653	0.0690	-3.846	0.0001
Log neighbourhood density	0.3846	0.1402	2.743	0.0061

RANDOM EFFECTS

	SD
Word	1.1739
Participant	0.3657

Table 6.5: Summary of logistic regression model for correctness on complex words only (the Complex Error Analysis), using contrast coding for the factor Sex with Female as the reference level. Df = 3428.

FIXED EFFECTS

	Estimate	SD	z	p(z)
Intercept	-2.9381	0.4156	-7.069	0.0000
Previous PC1	0.0011	0.0004	2.494	0.0126
Sex: Male	-0.5466	0.4926	-1.110	0.2672
Log whole-word frequency	-0.2820	0.0880	-3.204	0.0014
Sex:Male×Log word freq	0.2089	0.1009	2.070	0.0384

RANDOM EFFECTS

	SD
Word	1.0100
Participant	0.3512

illustrated for the All Items Analysis in fig. 6.1. The affix repetition value was set to 1 for all simple words, in order to control this variable in the All Items Analysis, while it varied between 1 and 10 for the complex words.

Previous PC1 and PC2 are shown in the top panels of fig. 6.1. The original previous RT-variables had negative factor loadings on PC1, which means that low PC1-values are correlated with long RTs on the previous trials and high PC1-values with short RTs on previous trials. Therefore, the downward slope of the line in the top left panel of the figure reflects an inhibitory effect of the previous reaction times: RTs were longer when RTs

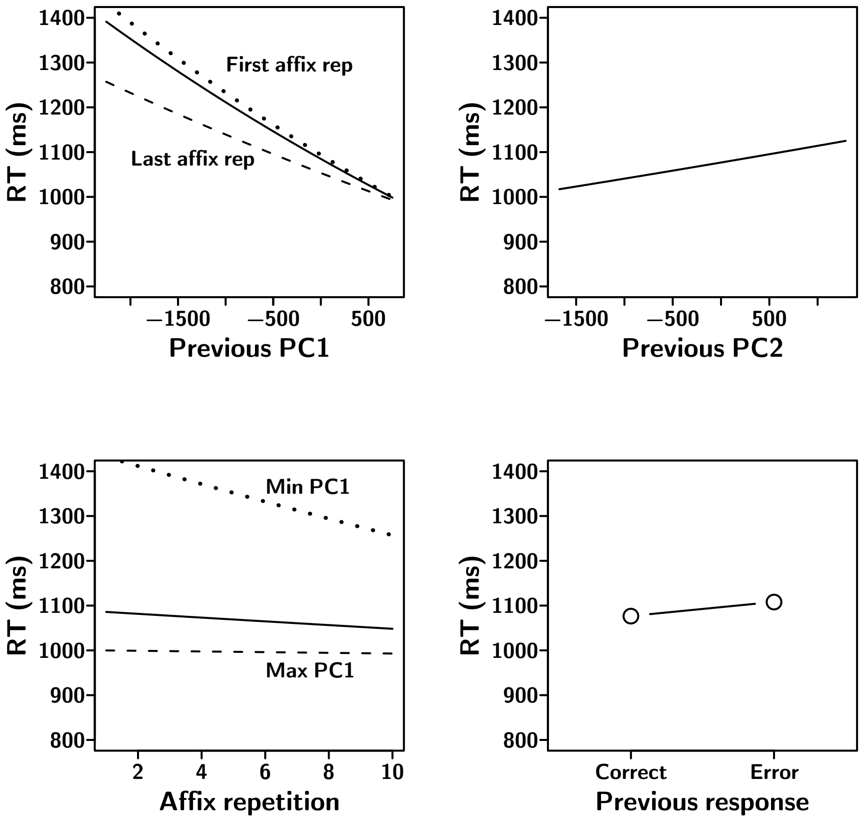

Figure 6.1: Partial effects of the control variables Previous PC1 and PC2, affix repetition and previous response. The line between the two circles in the bottom right panel is drawn to make the difference between the categories more clearly discernible.

to previous items were longer. The correlation of PC2 with the previous RT-variables was positive, so the the upward slope of the PC2 shown in the top right panel of fig. 6.1 also reflects an inhibitory effect, though this was weaker than the effect of PC1.

The effect of PC1 was modulated by the number of times an affix had been repeated, indicated by the interaction between PC1 and affix repetition in both the All Items and the Complex Analyses (see tables 6.2 and 6.3). The effect of the previous RTs was stronger at the first occurrence of an affix than at the last. This is shown in the top left panel of fig. 6.1 where the dotted line shows the effect of PC1 for the first occurrence of an affix and the dashed line the effect for the last occurrence. The bottom left panel of fig. 6.1 illustrates the effect of affix repetition which varied with RT on the previous items. The plot shows that this repetition effect is noticeable mainly for cases where the PC1 was low, corresponding to slow reactions to the previous items; in other words, affix repetition primarily plays a role when responses are slow. The reason that the affix repetition effect was significant in this experiment but not in Experiment 1 is probably that each affix occurred twice as often on words in Experiment 2 (ten times) as in Experiment 1 (five times). The fact that this variable is statistically controlled is important, since it means that none of the other effects are artefacts of affix repetition.

The correctness of the previous response was also significant, with longer RT when the previous response was an error, shown in the bottom right panel of fig. 6.1. This confirms the finding in Experiment 1 that this factor plays a larger role when inter-stimulus interval is variable, and hence typically shorter, than when it is fixed. It probably reflects that participants were generally aware of having made an error, and slowed down the response to the following item as a consequence.

Fig. 6.1 is based on the analysis of all items but the same pattern is observed in the analysis of the complex items alone. In the error analyses, trial number was significant in the All Items Analysis (see table 6.4) with participants becoming slightly more error-prone as the experiment proceeded; this probably reflects fatigue. Previous PC1 was significant, though not strong, in both error analyses. Since PC1 was negatively correlated with previous reaction times, the positive coefficients for PC1 in the error analyses mean that the slower participants were on the preceding trials, the fewer errors they made.

6.3.2 The complexity advantage

In Experiment 1, the three types of complex words were generally recognised faster than the simple words, although this effect was strongest for the suffixed derived forms and interacted with whole-word frequency and UP1. This complexity advantage was strongly confirmed in the present experiment: other relevant variables being statistically controlled, both the inflected and the derived words were processed significantly faster than the simple words (p-values = 0.0001), see table 6.2, where the coefficients for 'Inflected' and 'Derived' show the difference to the 'Simple' type which is the reference level. This advantage is illustrated in fig. 6.2. The error analysis shows no effect of complexity type, but this is not surprising since the binary error variable tends to be less sensitive than the graded RT.

The effect of repeated exposure to the affixes was controlled by the inclusion of the variable indexing how many times each affix had been encountered in the experiment previous to each trial, as described in the previous section. This rules out the possibility that the complexity advantage observed is an artefact of affix priming.

This complexity advantage indicates that the recognition of complex words can draw both on information relating to the whole word and to the constituent morphemes. Using a route-metaphor, it seems that the

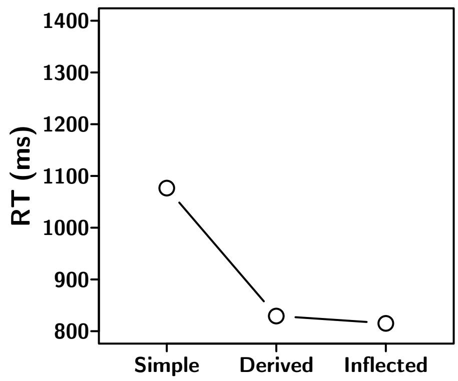

Figure 6.2: Partial effect of Complexity Type, showing the advantage for derived and inflected over morphologically simple words in the All Items Analysis. The lines between the circles are drawn to make the difference between the categories more clearly discernible.

recognition of complex words can proceed both via a whole-word and a morpheme route, resulting in faster recognition. This advantage is further discussed in section 8.2.

Crucially, the advantage is observed for both inflected and derived words. This is interesting in two ways: Firstly, the fact that both the highly regular inflected words and the less regular derived words are recognised faster than the simple words shows that the complexity advantage is not restricted to words that can be rule-generated, and hence cannot exclusively be the product of an efficient rule-mechanism. Secondly, the fact that the derived words are faster than the simple words, although the derived forms vary in semantic transparency suggests that the complexity advantage cannot be exclusively semantic, since it is maintained although the semantics of the roots is generally less available for the derived than for the inflected words.

The similar coefficients and overlapping HPD intervals for the inflected and derived words in table 6.2 indicate that there is no difference between the two categories, as also suggested by fig. 6.2. However, there are variables that cannot be included in the analysis of all items because they apply only to the complex words; therefore the real test of differences or similarities between the inflected and derived words is the Complex Analysis. The Complex Analysis showed no difference between inflected and derived words ($p > 0.4$), and the complexity type factor is therefore not included in the final model summarised in table 6.3. This is contrary to what the DP-Model of Ullman (2001) would lead one to expect, with its fundamental distinction between regular inflectional morphology and less regular derivational morphology. Crucially, the complexity type variable did not enter into any interactions with whole-word frequency (effects of which Ullman and colleagues predict for less regular forms like the present derived forms) or stem frequency (effects of which are predicted for regular inflected words). There were significant effects of the graded variables affix frequency and (marginally) semantic transparency (see sections 6.3.3 and 6.3.4 below) that vary between typical inflection and typical derivation, but no categorical difference between inflected and derived words. In other words, it seems that the differences between regular and less regular morphology is best accounted for as continuous rather than categorical, and thus not easily fitted into a framework with two distinct mechanisms.

The fact that the items are relatively typical exemplars of the relevant categories strengthens the validity of the conclusions drawn and constitutes a major advantage of the regression design used here and in the other experiments.

With regard to the definition of the different affixes as derivational vs. inflectional, this is largely uncontroversial for the affixes used here, with the exception of the adjectival superlative form *-est* which is categorised as inflectional (following Arndt, 1997, Diderichsen, 1987 and the lemmatisation of *Korpus90/2000*). Adjective superlative forms are sometimes categorised as derivational; for an experimental study doing this, see Marshall and van der Lely, 2006. The analyses were also run without words with this affix and showed the same pattern as the one reported here.

Another effect of word type was that of word class: In the analysis of all items, the adjectives were recognised faster than both nouns and verbs, though only significantly so relative to the verbs. This could be a consequence of adjectives probably being more likely than verbs to occur in isolation in everyday language use, and therefore being slightly more natural and recognised faster in the lexical decision task which, of course, presents words in isolation. This differs from the advantage for nouns observed by Sereno and Jongman (1997) and Baayen et al. (2006). The effect of word class did not reach significance in the analysis of RT to complex words or in either of the error analyses.

6.3.3 Frequencies

The analyses of all items showed a significant effect of whole-word frequency on both response latency and correctness (see tables 6.2 and 6.4): High-frequency words were recognised faster, as is usual in word recognition, particularly in the lexical decision task (Balota, 1994), and they were also less error-prone. More informatively, in the analysis of RT to complex items, whole-word frequency interacted with affix frequency (see table 6.3), and this interaction differed as a function of the sex of the participants. In the error analysis, the interaction between whole-word and affix frequency did not reach significance, but whole-word frequency varied between the sexes, with stronger frequency effects for the females.

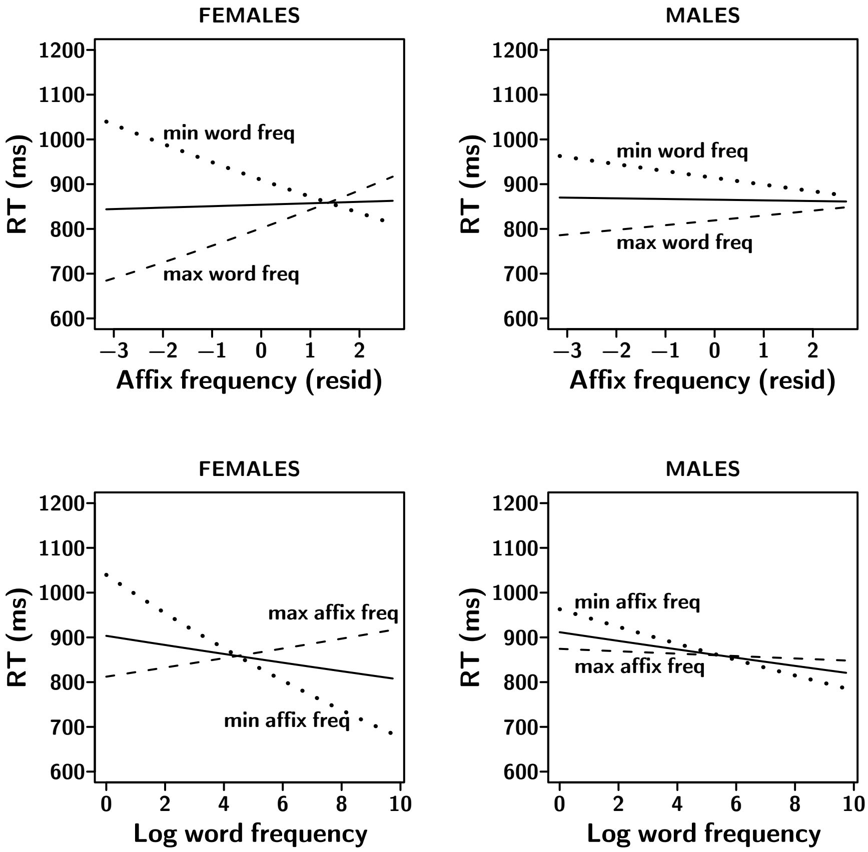

Figure 6.3: Partial effects of affix and whole-word frequency for the complex items. The top panels show the effect of affix frequency split between words with minimum (dotted lines), median (solid lines) and maximum (dashed lines) whole-word frequency, for women (left) and men (right). The bottom panels show the effect of whole-word frequency split between words with minimum (dotted lines), median (solid lines) and maximum (dashed lines) affix frequency, for women (left) and men (right). Affix frequency was residualised from whole-word frequency.

The Complex but not the All Items Analysis also included random slopes for whole-word frequency by participant. The effect of whole-word frequency differed more between participants when considering only the complex items than when including also the simple items; this varying sensibility to whole-word frequency for complex words suggests that participants differed in their reliance on the whole-word vs. constituent-based processing of these complex words. Interestingly, the difference between the sexes was found over and above these individual differences.

The interaction effects in the analysis of complex items are illustrated in fig. 6.3: The top panels show the effect of affix frequency for words with minimum (dotted lines), median (solid lines) and maximum (dashed lines) whole-word frequency, for females (left) and males (right). For words with median whole-word frequency, the effect of affix frequency is practically absent. However, there is a clearly facilitatory effect of affix frequency for words with low whole-word frequency, and a clearly inhibitory effect for words with high whole-word frequency. The effects are stronger for females than for males.

The bottom panels of fig. 6.3 demonstrate for females (left) and males (right) how the whole-word frequency effect is modulated by affix frequency: The solid line represents the whole-word frequency effect for words with median affix frequency, a straightforward facilitatory effect. For words whose affixes have the lowest token frequency, represented by the dotted line, the facilitatory effect of whole-word frequency is noticeably stronger than for the median frequency affixes. In contrast, for words with maximum affix frequency, shown by the dashed line, the effect of whole-word frequency reverses to become slightly inhibitory for the women (the left panel) and flat for the men (the right panel).

In contrast to Experiment 1, there were no significant effects of stem frequency. This suggests that the exclusive reliance on stem frequency as the measure of decompositional, morpheme-based processing is misguided, as argued in section 2.4. The present results show that it is useful to consider the converse constituent frequency, that of the affix.

The interaction between affix and whole-word frequency observed here is very similar to the interaction seen in the auditory Experiment 1a between stem and whole-word frequency. In the present experiment, all items are suffixed, so the interaction between whole-word and affix frequency is an interaction between whole-word and second constituent frequency. Experiment 1 included both prefixed and suffixed words, but the majority of the complex items were prefixed, suggesting that the interaction between whole-word and stem frequency also represents an interaction between whole-word and second constituent frequency. This is confirmed by two separate subanalyses of the data from Experiment 1a, of the suffixed words alone and of the two types of prefixed words together: the interaction between stem and whole-word frequency remained significant for the prefixed words, but was completely absent for the suffixed words when analysed alone, although a three-way interaction between affix type, whole-word and stem frequency did not reach significance in the Complex Analysis for Experiment 1a. In sum, Experiment 1a and Experiment 2 are consistent in that they both show interactions between whole-word and second constituent frequency. In both experiments, both frequencies showed the strongest effects when the other frequency was low (the dotted lines in fig. 6.3 and in fig. 5.3 on p. 126); it seems that recognition relies more on the type of information indexed by one of the frequencies when the information indexed by the other frequency is less available. Conversely, when both frequencies are strong, the effects are attenuated (in the case of whole-word frequency) or even reversed (in the case of second constituent frequency), as shown by the dashed lines in figures 6.3 and 5.3. As facilitation is expected for frequency effects, this pattern leaves the question why the effects disappear or become inhibitory when both frequencies are high.

In order to answer that question, it is first necessary to consider why it is the second constituent frequency that interacts with whole-word frequency. Since the onset of the second constituent is later than that of the whole word, it is at first sight surprising that it is primarily the second constituent that interferes with the whole-word frequency. However, although processing of the first constituent of the whole word can begin earlier than the processing of the second constituent, many aspects of the interpretation of the first constituent and thus of the whole-word depends on the second constituent. On a syntactic level, parts of the interpretation of the first constituent *John* in *John ate the apple pie* only become possible after the second constituent *ate* has been processed; for instance, the second constituent provides information that *John* is to be interpreted as an agent. In the complex word *apple pie*, something similar happens: we only know that *apple* refers to a cooked mass rather than a fruit on a tree after hearing the second constituent *pie*. In derived and inflected words, stems are often word class ambiguous, referring to rather different albeit related concepts; in almost all the present items and also often in general language use, any ambiguity is resolved when a derivational or inflectional suffix is encountered. The ambiguity of the stem may also be semantic, in which case the ambiguity — ranging from homonymy over polysemy to collocational differences — may be resolved when an affix is encountered. For instance, the base *bære* ('carry' or 'bear') also has a more figurative sense ('endure'), but the derivation *bær-bar* ('carry-able', often referring to laptop computers) is generally used in relation only to the literal sense. Monomorphemic words benefit from multiple senses (Rodd et al., 2002), but for complex words stem ambiguity may delay commitment to an interpretation until the second constituent disambiguates the stem.

On this background, the frequency interaction could arise as follows: When whole-word frequency is high, it not only means that the whole-word information is strong, but also that the constituents that make up the complex words are likely to co-occur, i.e. whole-word frequency can be understood as a combinatorial probability (see section 2.4). When the first constituent is encountered, it will activate not only its own representation — or in exemplar-terms, memory traces of the constituent as an independent word — but also other morphologically complex words in which the same constituent functions as a first constituent. There may be subphonemic cues that indicate that the constituent is a constituent rather than an independent word. If the whole-word frequency is high, the probability of the two constituents co-occurring is high and the complex word in question is likely to be a relatively strong candidate among the words activated on hearing the first constituent; this speeds up recognition as shown by the facilitatory effect of whole-word frequency in both Experiment 1a and 2. When the second constituent is encountered and processed, this affects the interpretation of the first constituent. When the frequency of that second constituent is high, combinatorial probability could matter less because the second constituent is relatively dominant. Conversely, when the second constituent is weak, recognition and interpretation would rely more on the strength of the morpheme combination indexed by the whole-word frequency, i.e. resulting in stronger effects of whole-word frequency when second constituent frequency is low (the dotted line in the bottom panels of fig. 6.3).

Seen from the opposite perspective, low whole-word frequency, indexing low combinatorial probability, means that a strong second constituent, indexed by high constituent frequency, aids processing, resulting in the stronger facilitatory effect of second constituent frequency when whole-word frequency is low, represented in the top panels of fig. 6.3. It seems that when other cues to second constituent identity are weak, i.e. when whole-word frequency is low, the strength of the second constituent itself provides cues to the interpretation of the combination. Conversely, when other cues are strong, i.e. when whole-word frequency is high, a strong second constituent is no longer helpful to recognition. This could be an effect of competition from other words involving the same constituent: the overall familiarity with the second constituent as a second constituent becomes inhibitory when more precise cues to the whole-word interpretation are strong. Importantly, the dashed line representing high whole-word frequency remains well below the other lines in the top panels of fig. 6.3 and the bottom right panel of fig. 5.3 on p. 126 up until the highest second constituent frequency values. This means that most of the words with high whole-word frequency are processed faster than those with lower whole-word frequency, even though the second constituent frequency effect becomes inhibitory.

Given this interpretation, the extent to which such an interaction arises in a given experiment will depend on how much the first constituents of the complex items in question are modified by their second constituents. These modifications can be quite subtle and therefore difficult to measure and control, hence the effects and their directions may differ between experiments.

The interpretation of the precise direction of the frequency effects is relatively speculative. However, three important points remain irrespectively of the specific interpretation of the direction of the effects: Firstly, the effects of whole-word and constituent frequencies are not independent contrary to what is argued in dual-route race models like Schreuder and Baayen (1995) and Bertram et al. (1999). Secondly, although interpretation starts very early based on the first constituent, it is later modified by the information contained in the second constituent. Thirdly, the interaction is compatible with the hypothesis that whole-word frequency should not only be interpreted as a string probability but also as the combinatorial probability, measuring the probability of the two constituents co-occurring.

Within the DP-Model, the processing delay for very high-frequency words with very high-frequency constituents could arise because information from two different neural substrates (Ullman, 2001, 2004) would have to be integrated at roughly the same point in time. However, in view of the fact that inflected and derived forms did not differ categorically, the explanation in terms of integration of information from different mental mechanisms is perhaps unlikely to be correct.

A further question is why there is an asymmetry such that the affix frequency effect becomes more inhibitory when word frequency is high, than the word frequency effect does when affix frequency is high, i.e. the asymmetry between the dashed lines in the top panels vs. the bottom panels in fig. 6.3 (and similarly for stem and whole-word frequency in Experiment 1a, see fig. 5.3 on p. 126). This is likely to be because the word has more bottom-up support than the affix and is encountered earlier than the affix in auditory processing in the suffixed words in this experiment; correspondingly in Experiment 1a, the onset of the stem is later than the onset of the whole-word for the prefixed and particle prefixed words. This makes the whole-word more resistant to interference from the second constituent, with the result that the whole-word frequency is not reversed but merely attenuated. In contrast, the second constituent frequency becomes inhibitory because this has less bottom-up support than the word and hence is more vulnerable to interference.

Turning to the difference between the sexes, this is best discerned by comparing the left panels of fig. 6.3 with the right panels: The effects and their interactions are much more pronounced for females than for males. This is somewhat similar to the results of Ullman et al. (2002), who report whole-word frequency effects for regular inflected words for female but not for male participants, under various conditions. However, in contrast to the results of Ullman et al., the frequency effects in the present experiment *were* significant for the males, but significantly weaker than for the females; from the perspective of the DP-Model, this could be related to the fact that the present experiment includes both regular and irregular complex forms. More critically for the DP-Model, the effects that were stronger for the females were not only of whole-word frequency, but both of whole-word and affix frequency. If the interaction is a result of the superior verbal memory of women (Kimura, 1999), this indicates that the constituent frequencies are also a product of verbal memory, rather than the result of a rule-based process in the procedural memory system.

The interaction between sex and whole-word frequency in the complex error analysis is more in line with the results of Ullman et al.: The interaction reflects that whole-word frequency had a significant effect only for the female participants. For both RT and correctness, interactions between sex and frequency were only found for the complex items, i.e. in the case where whole-word frequency may be specifically understood as an index of whole-word- as opposed to morpheme-based processing. It is therefore not surprising that no interaction was found for in the All Items Analyses.

In Experiment 1, a tendency for stronger frequency effects for complex words for female than for male participants became solidly non-significant when individual differences at the participant level were included in the form of random slopes. This was not the case in the present analyses: the three-way interaction between sex, whole-word and affix frequency remained significant when random slopes for frequency were included, and no other random slopes had explanatory value (according to likelihood ratio tests). In this case, as in Experiment 3 (see section 7.3.4), the group difference is not an artefact of individual differences.

The affix frequency measure used here was the token frequency of words in the corpus in which the affixes occurred in the outermost layer of morphological structure. Inflectional suffixes are generally found in the outermost layer of morphological structure, and for comparability (as well as for practical reasons, see section 3.4.2), the frequency counts for the derivational affixes were also restricted to such cases. Moreover, the affix frequencies were extracted as potential measures of affix productivity, for which the outermost layer of structure is the most relevant. In the analysis, the token frequency count can be replaced with a type count, with which it is highly correlated, without loss of significance. However, the pattern for the token frequency was clearer and more robust, and hence the token frequency was used as a measure of affix strength and productivity. The affix frequency measures are also correlated with the productivity measures P and P*, however, these showed no significant effects in the analyses.

6.3.4 Semantics and morphological families

In the analysis of all items, there was no effect of morphological family size for either RT or correctness. In the analysis of RT to complex items, the effect of family size was significantly facilitatory, see table 6.3. This is illustrated in the left panel of fig. 6.4: words with more family members were recognised faster. This is in line with previous evidence in the literature (see section 2.5) but contrasts with the inhibitory effect found in Experiment 1 (see section 5.3.4). This difference between the experiments is discussed in chapter 8, where the results of Experiment 3 can also be drawn in. The effect in the present experiment on its own is straightforwardly interpretable: The

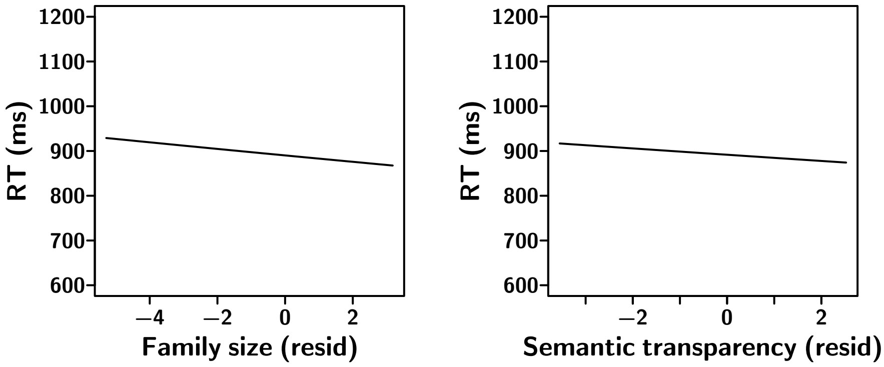

Figure 6.4: Partial effects of morphological family size and semantic transparency in the analysis of complex items. Morphological family size is residualised from affix and whole-word frequency and semantic transparency from morphological family size.

existence of more family members makes it easier to activate the semantics of the target word and hence makes the target easier to recognise.

Curiously, the effect of morphological family size only became significant in the analysis of RTs to complex items when semantic transparency and affix frequency were included in the analysis. This holds also when semantic transparency was decorrelated from morphological family size with which it was somewhat correlated, and is thus not caused by that correlation. Rather, it suggests that there is some other connection between these variables. The connection between morphological family and semantic transparency is relatively straightforward: if the morphological family size effect is semantic in nature, as argued for instance by De Jong et al. (2000), then the control of semantic transparency in a sense controls how available the semantics of the stem is and allows the morphological family size to emerge. It is less clear why the affix frequency variable must be included in the model in order for the family effect to become significant. It could, however, have to do with the fact that affix frequency has a very diverse effect on different items, see the top panels of fig. 6.3 and discussion in section 6.3.3, so that only when this source of variation is controlled, other effects emerge.

The difference between the analyses of all items and complex items is not simply a difference between simple and complex items in terms of morphological family effects, firstly because there was no interaction between complexity type and family size in the All Items Analysis, secondly because the family size effect only became significant for the complex items when semantic transparency and affix frequency were included in the analysis.

There was no effect of continuations in this experiment, so it seems that the standard family size measure is enough to account for the facilitatory effect of morphological family activation, contrary to Experiments 1 and 3. This is further discussed in section 7.3.2 and chapter 8.

Semantic transparency was only marginally significant, but the effect went in the expected direction: the more transparent the combination of stem and suffix, the faster the recognition of the complex word. Other things being equal, relatively transparent words such as *stolt-hed* ('pride', literally 'proud-ness') were recognised faster than opaque words such as *lang-som* ('slow', literally 'long-ish')), with intermediate cases such as *motiv-ere* ('motivate', literally the noun 'motive' plus a verbalising suffix) in-between. In spite of not quite reaching significance, the variable was included both because it was so close to being significant and in order to control this difference between inflected and derived words. The effect of semantic transparency of the stem-suffix combination is parallel to the results of Wurm (1997). It is also in line with the results of Marslen-Wilson et al. (1994) who dichotomised the semantic transparency variable and observed priming effects for transparent but not for opaque words. Since the variable was not dichotomised in the present experiment (for the reasons discussed in chapter 4), the results are not directly comparable; however, there were no interactions between semantic transparency and either constituent frequency, which would have corresponded to the difference in priming effects observed by Marslen-Wilson and colleagues. The difference may be the result of the present opaque items being less opaque than those employed by Marslen-Wilson et al., or of the possibility that in unprimed processing opaque words can undergo decomposition.

6.3.5 Uniqueness points

As in Experiment 1, the present analyses operate with two different UPs for the complex words (introduced in section 2.3.2). UP1 is the point at which a suffixed word deviates from all words in the language except those which are continuations of the same stem; for suffixed words, UP1 corresponds to the traditional UP. The exclusion of stem continuations from the UP-calculation was originally applied (Marslen-Wilson, 1984) in order to avoid that the UP would occur after word offset for most words. However, it is highly relevant which particular inflectional, derivational or compound form of the stem the listener is actually hearing, so a measure of the competition between these continuations is appropriate. The Complex UP (CUP) is such a measure: it defines the point at which the whole complex word deviates from other words in the language, including the continuation forms of the stem, but excluding the continuations of the whole word. Thus, UP1 measures competition between unrelated stems or lemmas, while CUP measures the competition

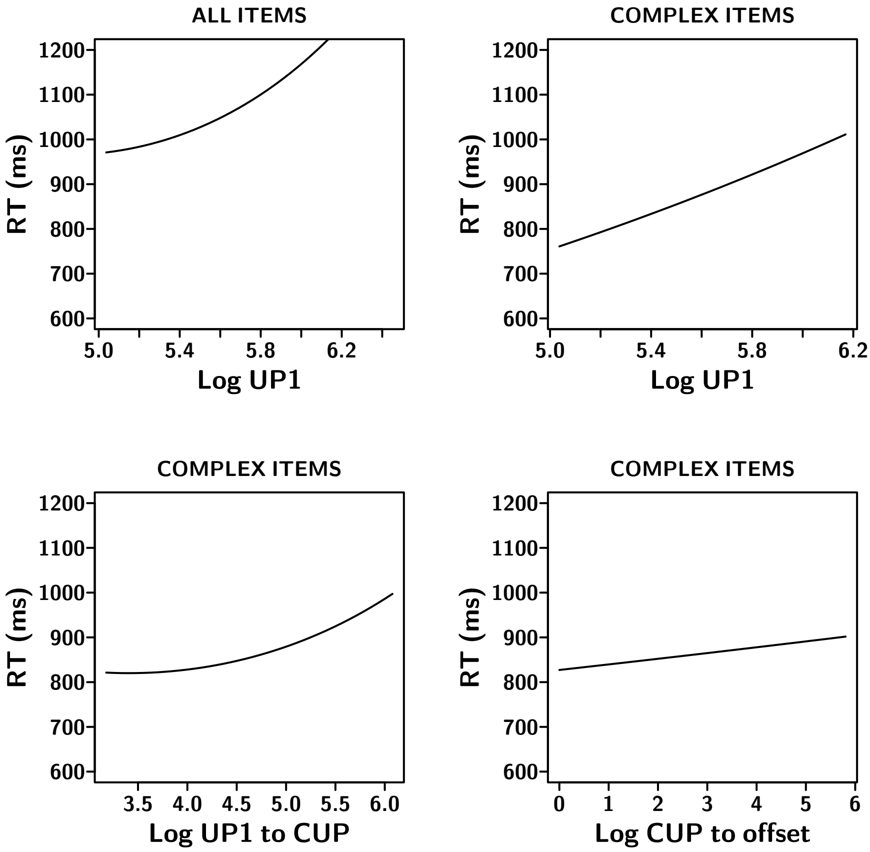

Figure 6.5: Partial effects of UP1, CUP and duration. The units on the horizontal axes are log ms. The top left panel shows the effect of UP1 for all items, while the remaining three panels show the effects for the complex items of UP1, UP1 to CUP and CUP to word offset.

between the morphologically related continuations of the target stem. UP1 and CUP are shown for a suffixed words in the top panel of fig. 2.2 on p. 33.

As in Experiment 1, both UP1 and CUP were included in the All Items Analysis, although the measures overlap for the simple words. Also similarly to Experiment 1, both uniqueness points and duration were recalculated to measure three non-overlapping parts of the auditory signal, in order to reduce collinearity: UP1 was included as time in log ms from word onset to UP1, CUP as time in log ms from UP1 to CUP and duration as time in log ms from CUP to word offset.

In the All Items Analysis, all three measures were significant: The effect of UP1, corresponding to the traditional UP for all three complexity types, was non-linear, as shown in the top left panel of fig. 6.5. Such a non-linearity did not reach significance for the complex items alone, see the top right panel of fig. 6.5 which shows a clear linear effect of UP1 for the complex items. This difference could be due to the All Items Analysis having greater statistical power. It is not a question of UP1 being earlier for the simple than for the complex items. The initial flatness in the All Items Analysis could be due to the fact that a certain amount of acoustic information is necessary for the UP effect to become relevant, as seems to be the case for the simple items in Experiment 1. The fact that the UP1-effect is non-linear when the simple items are included and linear when only the complex items are analysed is the same in both experiments.

The distance from UP1 to CUP is included in the All Items Analysis to keep this source of variation under control. However, the validity of the CUP-construct can only be assessed in the analysis of the complex items, where both uniqueness points are relevant, and where the CUP specifically measures the competition between morphologically related continuation forms of the stem. There was a significant effect of UP1 to CUP distance for the complex items, which was non-linear such that the inhibitory effect became stronger when the distance to the CUP was greater, see the bottom left panel of fig. 6.5. A similar non-linearity is observed in Experiment 3 and discussed in section 7.3.3. The CUP-effect shows that stem continuations compete in the recognition process, indicating that the participants are performing a rather precise word identification, rather than just relying on the uniqueness of the word from unrelated competitors to make a lexical decision.

There was no additional effect of the number of post-offset continuation forms of the whole word which Wurm et al. (2006) found to be significant. This could be because the suffixed inflected forms have few or no continuation forms.

The distance from CUP to word offset was significantly inhibitory in both analyses, with a quadratic term in the All Items Analysis. This difference between the analyses could be due to the range of CUP to word offset values being larger for the full dataset, though it is not the cases of zero distance from CUP to offset alone — which are most frequent for the simple words — that are responsible for the effect. The non-linearity can be understood in two ways: If the CUP to offset measure is understood literally, it indicates that the duration of the word beyond the CUP only begins to slow responses down if they are above a certain duration. If distance from CUP to offset is understood as an index of duration, it means that words must have a certain duration before the duration begins to be time-consuming.

The error analyses showed no effects of UPs or duration. The All Items Error Analysis did include a significant effect of neighbourhood density. Although this did not interact with complexity type, the reason that it is significant only when all items are included is probably the larger lexical neighbourhoods for simple words. The positive sign of the coefficient for neighbourhood density (see table 6.4) indicates that words with larger neighbourhoods are more error-prone. Like the effect of CUP, this indicates that participants are making relatively precise word identifications which are more difficult for words with denser neighbourhoods. It is unclear why the effect should be significant in the error analysis only. The fact that it is only significant in the present experiment, not in Experiments 1 or 3, may reflect a coincidental better correspondence between the visual neighbourhood density measure that could be extracted from the corpus and the phonological neighbourhoods of the items.

6.4 Summary

There are four main results of the present experiment:

Firstly, there was strong evidence of a processing advantage for both inflected and derived words over simple words, while there was no difference between the two complex types. This indicates that the morphological structure of the complex words is functional, but the non-difference between regular inflected and less regular derived words indicates that this functionality cannot be understood in terms of a rule-mechanism that is more efficient than whole-word-based processing.

Secondly, this experiment replicated the interaction between whole-word and constituent frequency which was found in Experiment 1. While the analysis of the RTs in Experiment 1a showed an interaction between stem and whole-word frequency, the present Experiment 2 showed an interaction between affix and whole-word frequency. The interaction in Experiment 1a was carried by the prefixed and particle prefixed words which means that in both experiments, the interaction was between whole-word and second constituent frequency. It shows that whole-word and morphemic processing routes do not operate independently and suggests that whole-word frequency is best understood as a conditional probability.

Thirdly, the effect of morphological family size was facilitatory for the complex words, and as such contrasts with the inhibitory effect found in Experiment 1a. This could be due to differences in constituent ordering or in the paces of the experiments, or it could be related to the role of morphological continuation forms. This issue is further investigated in Experiment 3, see section 7.3.2.

Finally, this experiment strongly supported the validity of operating with two uniqueness points: UP1 is the uniqueness point of the first constituent; this UP corresponds to the traditional UP for suffixed words and indexes competition between distinct lemmas that are unrelated to the target. The Complex UP is the uniqueness point of the whole complex word, which indexes competition from morphologically related stem continuation forms.

The combination of these two measures gives a more accurate picture of lexical competition in morphologically complex words, showing the sensitivity of the processing system to shifting probabilities over time, to the differences between related and unrelated competitors and to the constraints imposed by the first constituent on the recognition of the second constituent. The different UPs and other form-related measures are further documented in Experiment 3, see section 7.3.3.

The wider implications of these results and their relation to Experiments 1 and 3 are discussed in chapter 8.

Chapter 7

Experiments 3a and 3b

7.1 Introduction

One of the most surprising and challenging outcomes of the previous experiments was the discrepancy between the effect of morphological family size which was inhibitory in Experiment 1a but facilitatory in Experiment 2, and the fact that the linear inhibitory family effect in Experiment 1a contrasts with a substantial body of literature (see section 2.5). The most clear-cut difference between Experiments 1a and 2, which where both auditory, was the difference in inter-stimulus interval (ISI) and consequently in the pace of the experiments. This could help explain the difference in the effect of morphological family size in the following way: When ISI is variable and the pace of the experiment therefore relatively fast (like Experiment 2), a correct yes-decision is more likely to be based on a general familiarity with the word, whereas the fixed and therefore generally longer ISI in Experiment 1a probably strengthens the reliance on precise identification of the actual target. This does not mean that all correct decisions in Experiment 1a are necessarily based on precise identifications and even less that all decisions in Experiment 2 are made without precise word identification; indeed the significant effect of Complex UP in Experiment 2 suggests that relatively precise identification often does take place, but the speededness of Experiment 2 is nonetheless likely to favour faster responses to words that are familiar because of their large families. This may help explain why Experiment 2 shows a facilitatory effect of family size while Experiment 1a does not. It does not explain the direct inhibitory effect of morphological family size observed in Experiment 1a. However, this could also be caused by the longer ISI: the potentially slower processing in Experiment 1a is likely to be a result not of early automatic word recognition processes being slower, but of later decision or verification processes being elongated. Such late processes are not likely to be speeded up by general familiarity with the morphological family; instead activated morphological relatives could act as competitors, especially if — as is the tendency for larger families — they are semantically diverse and therefore contain many members that are relatively distant from the target.

Because the items and the affixes differed between Experiments 1a and 2, the hypothesis that ISI is responsible for the difference in family effects could not be directly tested for those datasets. Instead, the ISI, and with that the pace of experiment, was directly manipulated for a new set of items in the two experiments reported in this chapter: Experiment 3a had a variable ISI, resulting in a faster pace, while Experiment 3b had a fixed ISI of 3000 ms like Experiment 1a and therefore a slower pace. The fixed ISI made the pace of the experiment slower (cf. section 7.2.3), but the effects of ISI on actual response time were not entirely clear, see further in section 7.3.1. The first main question addressed by the present experiments is whether the morphological family size effect varies as a function of fixed vs. variable ISI. Another family effect that could vary with the pace of the experiment is the number of morphologically related continuation forms, since this had a significant effect in the slower Experiment 1a but not in the faster Experiment 2. This measure was also investigated in the experiments reported in this chapter.

Both Experiments 1a and 2 included derived forms which were therefore also used in the present experiment in the form of a group of prefixed and a group of particle prefixed words. Additionally, compounds were included; this widens the scope of this work to encompass the three main kinds of morphological operations, inflection, derivation and compounding. A further motivation is that productive compounding is a strong characteristic of Danish (see sections 3.2 and 3.4.3). More specifically, it is a question to which extent listeners are sensitive to the morphological structure of the compounds which differ from derived words by both constituents generally being formally and, particularly, semantically salient. In order to assess the contributions of the constituents, lexical frequencies and morphological families of both compound constituents and the whole compounds were investigated.

The final group of words were morphologically simple, included in order to dilute the proportion of complex words in the experimental lists. Since the simple words were included in any case, they were chosen to be similar in various ways to the complex categories (see section 7.2.1), in order to determine whether the complexity advantage generalises to compound words.

The second main purpose of these experiments was to replicate the effects of UP1 and CUP for prefixed and particle prefixed words and to investigate whether these measures are also significant predictors of recognition time for compounds. The measures can be applied without modifications to compound words, as outlined in section 2.3.2. Given the strong effects of these measures for both derived and inflected words in the previous experiments, the effects are hypothesised to generalise to compounds. If the UP-effects are indeed significant for the compounds, the combination of UP1 and CUP provides an accurate and uniform description of the distinct competition from unrelated and related words during the recognition of all types of complex words.

Another question that was addressed by the present experiments is the extent to which effects of morphological families for affixes, stems and compound constituents are parallel. It is obvious that compounds have two constituents for which the morphological family can be established in a straightforward way, but the words that contain a given prefix or particle can also be understood as the morphological family for that prefix or particle. A similarity between effects of the first constituent families of compounds (in which the first constituents are content words), particle prefixed words (whose first constituents can function as independent words but are not defined as content words) and prefixed words (whose first constituents are bound) would suggest that the morphological family effect is not exclusively semantic, while a semantic interpretation of the family effect would be strengthened if first constituent family effects were observed mainly or only for the compounds. However, the experiments showed no significant effects of first constituent family on RT for any of the types and no difference in the family effect between the types in the error analysis, so this aspect of the results cannot further the understanding of differences and similarities between the different complex types. The question is not further discussed.

In addition to the main questions concerning family effects and uniqueness points, the analyses control the effects of various context variables, like the analyses of the previous experiments. The effects of different frequencies are also investigated. The effects of whole-word and second constituent frequencies were included in the analysis of all complex words. The effects of both first and second constituent frequencies were studied in a subanalysis of compounds and particle prefixed words, in which both first and second constituents can function as independent words and have independent lexical frequencies.

7.2 Method

7.2.1 Materials

200 words and the same number of matched nonwords were presented in two auditory lexical decision experiments: Experiment 3a with a variable ISI and therefore a faster pace, and Experiment 3b with a fixed ISI and slower pace. The words fall into four groups, with 50 words in each: compounds, prefixed, particle prefixed and simple words. The items are listed in Appendix C, and the means, SDs and ranges for the different lexical variables summarised in table 7.1.

Since the main question of these experiments was whether the ISI affected the effect of morphological family size, all four groups of words were chosen to have similar ranges of family size. As the standard family size measure for the prefixed and particle prefixed words is that of the second constituent, the parallel measure for the compounds was also taken to be the second constituent family. However, the size of the first constituent family was also indexed and investigated for the complex words. The first constituent families were restricted to family members in which the constituent which was first in the target also occurred as the first constituent. For example, for a compound item like *damp-koge* ('to steam', literally 'to steam-boil'), the first constituent family included words like *damp-maskine* ('steam engine'), but not cases like *vand-damp* ('steam', literally 'water steam'). There are two reasons for this: Firstly, De Jong et al. (2002) found that the position-specific

Table 7.1: Means, standard deviations and ranges for properties of the items used in Experiments 3a and 3b. Whole-word frequency refers to the frequency of the base form of the whole word, second constituent frequency to the lemma frequency of that constituent as an independent word. The variables were converted to logarithmic scales for the statistical analyses in order to reduce their skewness, but are quoted here as non-transformed values for interpretability.

ALL ITEMS (n = 199)

	Mean	SD	Range
Whole-word frequency[a]	11	30	0 to 193
Family size, second constituent	574	606	2 to 3 476
Number of continuation forms	62	177	0 to 1 246
Neighbourhood density	2.3	4.9	0 to 29
Mean bigram frequency[a]	27 370	15 876	1 445 to 101 015
UP1, ms	321	98	141 to 670
Length, ms	704	140	354 to 1 210

COMPLEX ITEMS (n = 149)

	Mean	SD	Range
Whole-word frequency[a]	4	17	0 to 186
Second constituent frequency[a]	401	2 658	0.2 to 32 437
Family size, first constituent	304	354	5 to 1 476
Family size, second constituent	611	606	4 to 3 476
Number of continuation forms	17	78	0 to 924
Neighbourhood density	0.4	0.8	0 to 6
Mean bigram frequency[a]	26 124	13 991	1 445 to 70 722
Juncture bigram frequency[a]	10 033	16 712	23 to 83 663
UP1, ms	297	83	141 to 558
Complex UP, ms	499	112	290 to 847
Length, ms	747	122	450 to 1 210

[a] Values for these variables are counts per millon in *Korpus90/2000*.

family was a better predictor of compound recognition than a general family count. Secondly, it had the practical reason that the family counts for the prefixes and particles were based on the productivity measures which only counted occurrences as a first constituent (see 3.4.2); for comparability, the same method had to be applied to the compound families. In fact, there was a further potential difference between the types, namely that the type and token frequencies for both types of prefixes includes only productive uses, while there was no such distinction for the compounds. However, there are no indications that this makes a difference, since there were no significant effects of the first constituent family on RT for any of the types.

Restricting the second constituent families in a similar way was virtually impossible, because the available corpus is not parsed for derivational and compound structure, so this was not done, although De Jong et al. (2002) found that the position-specific family counts were better predictors than the general counts both for the first and the second constituent. Moreover, it seems likely that the position-specificity of the first constituent matters more than for subsequent constituents in a language like Danish where so many words, including many compounds, have more than two morphemes and stress often falls on the first morpheme: it is plausible that the salience of the first constituent differs much more from that of subsequent constituents than those second, third and fourth constituents differ from each other. There is a sense in which the grammatical head of compounds — i.e. the second constituent for the present bimorphemic compounds — is different from other constituents, but given the complete absence of an effect of the second constituent family (see section 7.3.2 below), this is unlikely to play a role.

The simple words, stems and second constituents had a higher range of morphological family sizes than the items in the previous experiments: ranging from 2 to 3476 family members, in contrast to 1 to 2103 in Experiment 1 and 1 to 2165 in Experiment 2. Items with larger families were included both in order to maximise the chances of detecting a family effect and because many appropriate items happened to have rather large families. As usual, the automatically extracted string matches for the family were manually cleaned to include only actual family members. The first constituent families ranged from 0 to 1476; this is relatively low because only productive uses of the affixes were counted as part of the affix families, see section 3.4.2.

For all words, similar considerations applied as in previous experiments. Very irregular pronunciations were avoided as were pronunciations varying much between casual and careful speech, homonymous or strongly polysemous whole-words and pseudo-cases of the affixes used on the complex items. The complex words were bimorphemic, had free stems and avoided linking morphemes and stem allomorphy, with the exception of regular stress- and *stød*-variations and schwa-deletion. Only relatively semantically transparent complex words were chosen. The item choice was less restricted by considerations of word class of the constituents than for previous experiments, since the word class of the stems did not play a role in the previous experiments, and since such considerations would have made the item choice substantially more difficult. The constituent frequency used in the analyses was always that of the stem in the appropriate word class, except in cases where the affix changed the word class.

The compounds were a mixture of noun-noun, noun-adjective, noun-verb and adjective-verb compounds. There were six adjectives, 22 nouns and 22 verbs; these proportions were repeated within each of the other item categories in the experiment. It was attempted to select compounds where both constituents had a transparent relation to the whole word (cf. Libben, Gibson, Yoon and Sandra, 2003). Many compound verbs in Danish carry one of a relatively restricted number of verbs which have rather broad and vague meanings, but only two such compound verbs were included as stimuli, namely *plan-lægge* ('to plan', literally 'to plan-lay') and *pant-sætte* ('to pawn', literally 'to place as a security/pawn'). For these two compounds, the transparency of the head relative to the whole is perhaps not quite as high as could be desired. The remaining 20 compound verbs had semantically heavier heads, e.g. *sol-bade* ('to sunbathe').

The particle prefixed words carried five different particles, each represented by ten words. These particles were chosen from the eight used in Experiment 1 because they were comparable to the prefixes in length, ruling out the bisyllabic *efter-* ('after-') and *over-* ('over-'), and were found on enough bimorphemic items that fulfilled other requirements. Five prefixes were used, each of which also occurred on ten different words. Out of the seven prefixes used in Experiment 1, these were the ones for which it was possible to find enough appropriate items that had not been used in the previous experiment. The simple words were chosen to approximately match the complex groups in terms of morphological family size, frequencies and word class.

As in the previous experiments, each of the words was used as the basis for constructing a nonword by changing one to three phonemes in the stem. The prefixes and particle prefixes were retained on the nonwords. The nonwords that were based on the compounds preserved no real constituents in either position, because that was deemed to make the task too difficult and less comparable to the previous experiments. The experimental items are listed in Appendix C, the affixes in Appendix D.

There was some repetition of stems from the previous experiments, because many bimorphemic transparent words carry the same stems, but there were only three whole-word repeats from Experiment 1, all prefixed, namely *be-laste* ('to load' or 'to strain'), *mis-tolke* ('to mis-interpret') and *sam-drift* ('joint operation').

By accident, two of the complex words chosen contained allomorphs of the same stem. Both were presented in the experiment, in order to keep constant the number of times each affix occurred in the experiment, but one of them (the particle prefixed *om-valg*, 're-choice' or 'new election') always occurred after the other and was never included in the analyses. Table 7.1 and Appendix C also exclude this word. This leaves 199 items.

7.2.2 Recording and preparation of stimuli

The stimuli were recorded in a quiet room directly onto a hard disk at a sampling rate of 48 kHz and a bit depth of 16 bit, using an AKG microphone (C 414 B-XLS) and Final Cut Pro for Macintosh. As in previous experiments, words and nonwords were mixed in the reading lists, and read by the same female native speaker of Danish as the previous items.

7.2.3 Procedure

The procedure of Experiment 3a was identical to that of Experiment 2, that of Experiment 3b identical to Experiment 1a, see sections 6.2.3 and 5.2.4, respectively. The difference between 3a and 3b was the ISI: in the slower Experiment 3b, each trial had a fixed duration of 3000 ms (with time-out occurring two ms before the trial ended, for technical reasons), while in the faster Experiment 3a each trial ended when the subject responded or at a time-out of 3000 ms after the beginning of the trial. Thus, the time-out was in practice the same for the two experiments, but in Experiment 3a, a response initiated the next trial, creating a faster pace. This difference meant that Experiment 3a lasted 10 to 15 minutes, while Experiment 3b lasted approximately 25 minutes, in both cases including two breaks. The instructions and course of the experiments were otherwise identical and like those of the previous auditory Experiments 1a and 2.

7.2.4 Participants

40 volunteers were tested individually in a sound-attenuated room, 20 for each version. Seven males and 13 females participated in Experiment 3a; they were between the ages of 24 and 41 (mean 28.8 years). Five males and 15 females participated in Experiment 3b; their ages ranged between 21 and 39 with a mean of 29.5. All participants had grown up with Danish as their first language and reported normal hearing. The same background questionnaire as in the previous experiments was used (see Appendix F).

7.3 Results and discussion

As in previous experiments, the reaction time analyses excluded all error responses (3.9% of responses) as well as all responses to items where the error rate was above 30%, in this case two very low-frequency complex items *mis-klæde* ('to not suit (someone)', literally 'to mis-suit') and *mis-greb* ('mistake'). All in all, 4.4% of the responses were removed from the RT-analyses in this way. Measured from word onset, mean correct RT to the remaining words was 958 ms in Experiment 3a (with the variable ISI) and 993 ms in Experiment 3b (with the fixed ISI). Experiments 3a and 3b were analysed together, including ISI as a factor in the analysis, with the two levels variable (Experiment 3a) and fixed (Experiment 3b). The significance of this ISI-factor is discussed in section 7.3.1.

All responses to all 199 items were analysed in the All Items Error Analysis and all responses to all 149 complex items in the Complex Error Analysis.

Tables 7.2 to 7.5 summarise the final analyses which were reached in the same way as for the previous experiments: All items were analysed using variables relevant to both simple and complex words, while analyses of the complex items tested additional variables that were relevant specifically to those. The data were filtered by removing data points with large standardised residuals in the initial analyses, and then refitting the models to these trimmed data sets. For both All Items and Complex RT Analyses, 2.7% of correct RTs were removed by this filtering procedure. A large number of variables were investigated and non-significant variables were removed in a backwards stepwise procedure. The models all include adjustments to the intercept for the random variables participant and word; the All Items Analyses also include random slopes for whole-word frequency; these were not justified for the Complex Analyses. In addition to the models summarised in tables 7.2 to 7.5, the prefixed and particle prefixed words were analysed alone to investigate possible interactions between whole-word, stem and affix frequency. Additionally, the compounds and particle prefixed words were analysed to assess the potential contribution of the frequencies of the constituents as independent words. These subanalyses never revealed patterns that were not also present in the more general analyses.

Collinearity between the variables was reasonable, with κ-values below 15 for all models, when previous RTs were orthogonalised using principal components analysis (PCA), continuations were residualised from whole-word frequency, and form-variables were included as log distance from word onset to UP1, from UP1 to CUP and from CUP to offset, as they were for the previous experiments. Collinearity did become a problem for the analysis of complex RTs when other form-related variables were included, namely constituent durations and second constituent UP; therefore the secondary analysis that included these variables operated with principal components based on all UP- and duration-measures, see further section 7.3.3. As usual, measures of frequency, family size, UP and duration were logarithmically transformed to reduce skewness.

Figures 7.1 to 7.7 show partial effects of the different significant variables, adjusted to the medians for other continuous predictors and to the relevant fixed factor reference levels. Most of the figures use the same 600 ms range on the vertical axes as in the previous chapters, but figures 7.6 and 7.7 employ a reduced range to show more clearly some rather subtle differences.

The remainder of this section presents the significant effects found in the two experiments. Since the context effects serve as control variables, these are presented first (section 7.3.1). Then, the first main hypothesis of the experiments, namely that the pace of the experiment would affect the morphological family effects, is addressed in section 7.3.2. Section 7.3.3

Table 7.2: Summary of regression model for RTs to simple and complex words in Experiments 3a and 3b (the All Items Analysis), using contrast coding for Complexity Type with Simple as the reference level, for Previous Response with Correct as the reference level, for Previous Item with Nonword as the reference level and for ISI with Variable (i.e. Experiment 3a) as the reference level. In this and the following tables for the random effects, 'Groups' denote the main grouping factors (Word and Participant), and 'Name' specifies whether the standard deviation refers to random intercepts or to random slopes for some variable, in this case word frequency. Df = 7374.

Fixed Effects

			HPD 95% CI			
	Estimate	MCMC mean	Lower	Upper	p(MCMC)	p(t)
Intercept	11.4683	11.4802	8.9051	13.8901	0.0001	<0.0001
Previous PC1	0.1100	0.1100	0.0968	0.1232	0.0001	<0.0001
Previous PC3	0.0266	0.0266	0.0107	0.0415	0.0002	0.0007
Previous PC4	0.0284	0.0283	0.0130	0.0438	0.0004	0.0004
Previous response: Error	0.0245	0.0245	0.0073	0.0409	0.0046	0.0042
ISI: Fixed	0.0312	0.0314	-0.0224	0.0842	0.2396	0.2342
Previous item: Word	0.0295	0.0295	0.0218	0.0374	0.0001	<0.0001
ISI:Fixed×Prev item:Word	-0.0264	-0.0263	-0.0374	-0.0152	0.0001	<0.0001
Trial number	0.0000	0.0000	-0.0001	0.0000	0.5148	0.5135
ISI:Fixed×Trial	0.0001	0.0001	0.0000	0.0001	0.0032	0.0028
Type: Compound	-0.0280	-0.0274	-0.0922	0.0381	0.4084	0.3913
Type:Compound×Trial	-0.0001	-0.0001	-0.0002	0.0000	0.0014	0.0012
Type: Prefix	-0.0646	-0.0641	-0.1195	-0.0101	0.0224	0.0197
Type:Prefix×Trial	-0.0001	-0.0001	-0.0002	0.0000	0.0098	0.0078
Type: Particle	-0.0785	-0.0782	-0.1387	-0.0170	0.0120	0.0103
Type:Particle×Trial	0.0000	0.0000	-0.0001	0.0000	0.5798	0.5670
Continuations^r	-0.0196	-0.0197	-0.0292	-0.0105	0.0001	<0.0001
Log UP1^l	-1.8456	-1.8499	-2.7511	-0.9807	0.0002	<0.0001
Log UP1^q	0.1793	0.1797	0.1023	0.2582	0.0001	<0.0001
Log UP1 to CUP	0.0282	0.0281	0.0176	0.0384	0.0001	<0.0001
Log CUP to offset^l	-0.0292	-0.0291	-0.0565	-0.0017	0.0362	0.0337
Log CUP to offset^q	0.0074	0.0074	0.0030	0.0119	0.0008	0.0009
Log whole-word frequency	-0.0167	-0.0166	-0.0227	-0.0109	0.0001	<0.0001
ISI:Fixed×log word freq	-0.0041	-0.0041	-0.0077	-0.0005	0.0240	0.0256

Random Effects

Groups	Name	SD
Word	Intercept	0.0633
Participant	Intercept	0.0784
Participant	Log word frequency	0.0040
Residual		0.1175

^{l} Linear.

^{q} Quadratic.

^{r} Residualised.

Table 7.3: Summary of regression model for RT to complex words only (the Complex Analysis), using contrast coding for the factors Complexity Type with Compound as the reference level, Previous Response with Correct as the reference level, Previous Item with Nonword as the reference level, ISI with Variable (Experiment 3a) and Sex with Female as the reference level. Df = 5543

FIXED EFFECTS

	Estimate	MCMC mean	HPD 95% CI Lower	HPD 95% CI Upper	p(MCMC)	p(t)
Intercept	10.7211	10.7449	7.9079	13.5383	0.0001	<0.0001
Previous PC1	0.1101	0.1128	0.0982	0.1276	0.0001	<0.000
Previous PC3	0.0280	0.0280	0.0108	0.0454	0.0026	0.0015
Previous PC4	0.0409	0.0406	0.0227	0.0586	0.0001	<0.0001
Previous response: Error	0.0245	0.0243	0.0054	0.0428	0.0104	0.0103
ISI: Fixed	0.0639	0.0642	0.0151	0.1134	0.0110	0.0302
Previous item: Word	0.0333	0.0332	0.0244	0.0421	0.0001	<0.0001
ISI:Fixed×Prev item:Word	-0.0278	-0.0277	-0.0402	-0.0155	0.0002	<0.0001
Type: Prefix	0.0232	0.0230	-0.0122	0.0573	0.1972	0.2247
ISI:Fixed×Prefix	-0.0312	-0.0317	-0.0648	0.0002	0.0512	0.0619
Type: Particle	0.0068	0.0069	-0.0253	0.0392	0.6636	0.6938
ISI:Fixed×Particle	-0.0496	-0.0500	-0.0819	-0.0164	0.0032	0.0032
Trial number	-0.0001	-0.0001	-0.0002	0.0000	0.0072	0.0064
ISI:Fixed×Trial	0.0000	0.0000	-0.0001	0.0001	0.9150	0.8901
Prefix×Trial	0.0000	0.0000	-0.0001	0.0001	0.5908	0.5730
Particle×Trial	0.0000	0.0000	-0.0001	0.0001	0.8714	0.8599
ISI:Fixed×Prefix×Trial	0.0001	0.0001	0.0000	0.0002	0.2126	0.2301
ISI:Fixed×Particle×Trial	0.0001	0.0002	0.0000	0.0003	0.0198	0.0225
Continuations^r	-0.0130	-0.0129	-0.0234	-0.0032	0.0130	0.0240
Log UP1^l	-1.6760	-1.6842	-2.6823	-0.6692	0.0014	0.0035
Log UP1^q	0.1712	0.1719	0.0772	0.2568	0.0002	0.0008
Log UP1 to CUP^l	-0.0636	-0.0637	-0.0881	-0.0408	0.0001	<0.0001
Log UP1 to CUP^q	0.0173	0.0173	0.0130	0.0214	0.0001	<0.0001
Log CUP to offset^l	-0.0811	-0.0809	-0.1121	-0.0477	0.0001	<0.0001
Log CUP to offset^l	0.0161	0.0161	0.0112	0.0207	0.0001	<0.0001
Log whole-word frequency	-0.0143	-0.0143	-0.0196	-0.0094	0.0001	<0.0001
Sex: Male	-0.0086	-0.0088	-0.0574	0.0400	0.7240	0.7725
ISI:Fixed×Log word freq	-0.0065	-0.0065	-0.0098	-0.0031	0.0010	0.0002
Sex:Male×Log word freq	0.0051	0.0051	0.0015	0.0086	0.0070	0.0048

RANDOM EFFECTS

Groups	Name	SD
Word	Intercept	0.0550
Participant	Intercept	0.0827
Residual		0.1146

^{l} Linear.

^{q} Quadratic.

^{r} Residualised.

Table 7.4: Summary of logistic regression model for response correctness on all words (the All Items Error Analysis), using contrast coding for Complexity Type with Simple as the reference level, for Previous Item with Nonword as the reference level and for ISI with Variable (Experiment 3a) as the reference level. Df = 7945.

Fixed Effects

	Estimate	SD	z	p(z)
Intercept	12.83023	4.7488	2.702	0.0069
Previous PC1	-0.6604	0.2610	-2.446	0.0144
ISI: Fixed	0.6623	0.3090	2.143	0.0321
Previous item: Word	0.3851	0.1765	2.182	0.0291
ISI:Fixed×Prev item: Word	-0.5211	0.2577	-2.022	0.0432
Log first constituent family size	-0.2493	0.1001	-2.490	0.0128
Continuations^r	-0.4788	0.1230	-3.893	<0.0001
Log length	-2.2074	0.7151	-3.087	0.0020
Type: Compound	0.0987	0.8364	0.118	0.9061
Type: Prefix	0.4220	0.8288	0.509	0.6106
Type: Particle	-0.3394	0.8574	-0.396	0.6922
Log whole-word frequency	-0.2220	0.1153	-1.926	0.0541
ISI:Fixed×Log word frequency	-0.1320	0.0610	-2.164	0.0305
Compound×Log word frequency	-0.4366	0.1988	-2.196	0.0281
Prefix×Log word frequency	-0.2307	0.1449	-1.592	0.1115
Particle×Log word frequency	-0.1459	0.1545	-0.944	0.3452

Random Effects

Groups	Name	SD
Word	Intercept	0.8809
Participant	Intercept	0.5483
Participant	Log word frequency	0.0650

^{r} Residualised.

describes how effects of UP1 and CUP were observed both for compounds and for both kinds of prefixed words, confirming the second main hypothesis of the experiments. The two last sections confirm and extend findings of the previous experiments: the frequency-effects in section 7.3.4 and the complexity advantage in section 7.3.5.

7.3.1 Context variables

In this experiment, even more of the context-related control predictors were significant than in previous experiments. Three principal components based on reaction times to the four previous items were significant in both the All Items and the Complex Analysis; they are shown for all items in fig. 7.1. The first and strongest component is PC1; the factor loadings of the original

Table 7.5: Summary of logistic regression model for response correctness for complex items in Experiments 3a and 3b (the Complex Error Analysis), using contrast coding for Complexity Type with Compound as the reference level. Df = 5965.

FIXED EFFECTS

	Estimate	SD	z	p(z)
Intercept	-1.0989	0.8814	-1.247	0.2124
Continuations^r	-0.6937	0.1740	-3.988	<0.0001
Log whole-word frequency	-0.5857	0.0836	-7.007	<0.0001
Type: Prefix	1.1405	0.3152	3.618	0.0003
Type: Particle	0.2740	0.3242	0.845	0.3979
Log juncture bigram frequency	-0.1725	0.0687	-2.511	0.0120

RANDOM EFFECTS

Groups	Name	SD
Word	Intercept	0.9110
Participant	Intercept	0.7746

^{r} Residualised.

variables show that slower responses on previous trials are associated with slower response on the current trial. PC3 and PC4 are mixed with negative factor loadings of some of the four previous RTs and positive loadings of others, and are thus harder to interpret. All three components serve to control the effect of response context. Fig. 7.1 shows the effects based on the All Items Analysis, but the same pattern is observed when the complex items are analysed alone, see table 7.3. Of the Previous PCs, only the strongest, Previous PC1, was significant in the All Items Error Analysis, showing fewer errors when previous responses were slow. None of the Previous PCs were significant in the statistically less powerful Complex Items Error Analysis.

Further, there was an effect of the correctness of the previous response, with RT being longer when the previous response was incorrect. This effect is illustrated in the top left panel of fig. 7.2 and is in line with the findings in Experiments 2 and 1b. The absence of such an effect in Experiment 1a was understood in terms of this experiment being slower, as a consequence of the fixed ISI; however, the fact that an interaction between ISI and previous response correctness does not reach significance in the present analyses (although the tendency is in the predicted direction, with little effect of previous response correctness for the slower Experiment 3b) suggests that other factors may also be responsible for the difference between the experiments. The effect is straightforwardly interpretable in terms of more

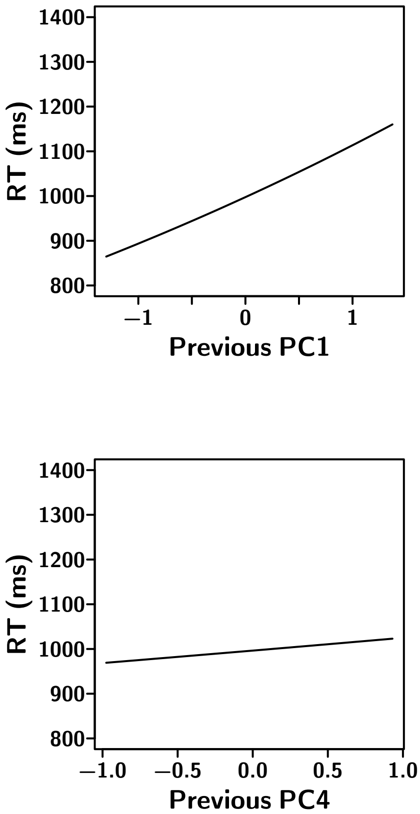

Figure 7.1: Partial effects of the principal components (PCs) based on RTs on the previous four trials. The plots are based on the All Items Analysis but the same pattern is observed in the Complex Analysis. The lines between the circles in the top panels are drawn to make the differences between the categories more clearly discernible.

careful and therefore slower responding immediately following an incorrect response. This slower responding did not result in significantly fewer errors: there was no effect of previous response correctness in either error analysis.

The next panel of fig. 7.2, in the top right, shows the effect of a "new" context-related predictor, namely the effect of the lexicality of the previous item, introduced by Diependaele et al. (2007). Responses were slower when the previous response was a word than when it was a nonword, suggesting that the activation of unrelated concepts in the mental lexicon by the previous item, which only occurs if the previous item was a word, slows down processing of the current item. This activation dies away quickly, as indicated by interaction between previous lexicality and ISI (see tables 7.2 to 7.3): The effect was only significant when ISI was variable and therefore generally shorter, as in Experiment 3a, not with the relatively long fixed ISI in Experiment 3b, as illustrated by the difference between the solid line representing Experiment 3a and the dashed line representing Experiment 3b in the top right panel

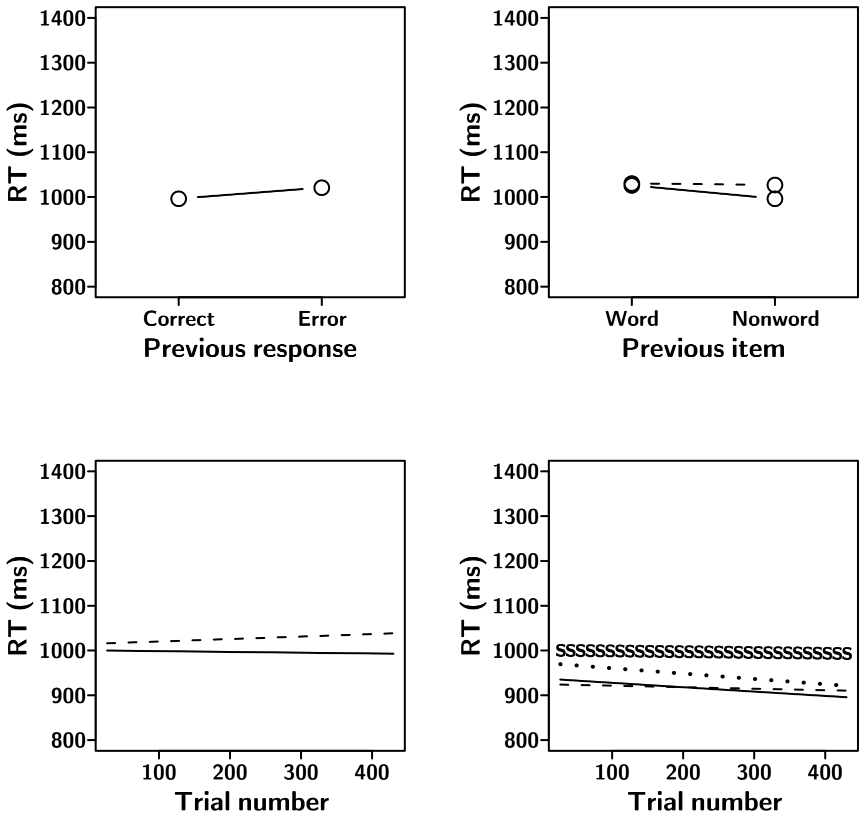

Figure 7.2: Partial effects of context-related control variables correctness on the previous trial, lexicality of the previous item and trial number. The plots are based on the All Items Analysis. In the top right and bottom left panels, the solid line shows effects in Experiment 3a with the variable ISI, and the dashed line effects in Experiment 3b with the fixed ISI. In the bottom right panel, s's indicate simple words, the dotted line compounds, the dashed line particle prefixed words and the solid line prefixed words.

of fig. 7.2. Diependaele et al. found that the lexicality of the previous item had the opposite effect: if the previous was a nonword, reaction times were slower, indicating a switch cost, since the reaction times that were analysed were, of course, all to words. The difference between the experiments could be due to list composition, to the maximum number of words or nonwords in a row, or to the fact that Diependaele et al.'s experiment was a priming study.

The All Items Error Analysis also showed an interaction between ISI and the lexicality of the previous item, similar to that observed for the RT-analyses: In Experiment 3a with the variable ISI, there were more errors when the previous item was a word. The interaction was not quite significant in the Complex Error Analysis and therefore not included in the final model summarised in table 7.5.

There were significant effects of trial number, in this experiment interacting with both ISI and complexity type to form a rather complex pattern which is illustrated in the bottom panels fig. 7.2. The difference between variable and fixed ISI is shown for the reference level simple words in the bottom left panel: there was no significant effect of trial number when ISI was variable, as shown by the solid line, but a small fatigue effect when ISI was fixed as shown by the dashed line. The fatigue effect when ISI was fixed is a natural consequence of Experiment 3b lasting longer.

Trial number also interacted with complexity type as shown in the bottom right panel of fig. 7.2, for the reference level variable ISI. The simple words (shown by the s's) and the particle prefixed words (shown by the dashed line) exhibited no significant effect of trial number, while there were significant training effects for compounds (dotted line) and prefixed words (solid line). This effect is not an artefact of affix priming: A variable indexing affix repetition was not significant and its inclusion in the analysis did not affect the other results (see further below). Apparently, the compounds and prefixed words became easier to process as the experiment progressed.

In the Complex Analysis, there was a three-way interaction between trial number, complexity type and ISI, which did not reach significance in the All Items Analysis. This interaction shows a significant training effect for the reference levels compound and variable ISI, the only factor level that differs significantly from this is the particle prefixed words in Experiment 3b with the fixed ISI, which showed no training effect. When the slopes for trial number were assessed individually for each of the six levels of ISI (variable/fixed) by complexity type (prefix/particle/compound), it emerged that the training effects were only significant for the prefixed and compound words in Experiment 3a. This pattern is in accordance with what is found in the All Items Analysis and illustrated in the bottom right panel of fig. 7.2. In the error analyses, there were no significant effects of trial number.

In Experiment 2, an effect of affix repetition was found. A similar variable, indexing the number of times an affix had been repeated, was also investigated for the present data. The secondary analysis of only the prefixed and particle prefixed words — i.e. those words for which the affix repetition variable was actually meaningful — showed an effect of affix repetition in interaction with Previous PC1 which was similar to the interaction observed in Experiment 2. However, in the Complex Analysis, where compounds were also included with a value of one for all items on this variable, this effect became non-significant, and there were also no main or other interaction effects involving the affix repetition variable. The same applied in the All Items Analysis which included even more items which had the same value of one on this variable, namely the simple words. Crucially, even if this non-significant variable *was* included in the All Items and Complex Analyses, all other effects remained stable. It can be concluded that affix repetition is not a confounding factor in the present analyses.

Finally, there is the difference between Experiment 3a with the fixed ISI and Experiment 3b with the variable ISI. This enters into so many interactions that it is problematic to conclude what the main effect of ISI is. The mean raw RT was longer in Experiment 3b with the fixed ISI (993 ms) than in Experiment 3a with the variable ISI (958 ms). In the Complex Analysis, the adjustment to the intercept for fixed relative to variable ISI is significant, but this significance is for the hypothetical situation where all co-variates are zero. The various interactions also show that RTs were longer in Experiment 3b, but more so when the previous item was a nonword, late rather than early in the experiment and for low rather than high-frequency words.

7.3.2 Morphological family

The experiment set out to investigate whether the difference between the inhibitory family size effect in Experiment 1a and the facilitatory effect in Experiment 2 was due to the difference in the pace of the experiments. However, the hypothesis that family size is facilitatory for fast responses, while inhibition emerges as a result of competition at a later stage of word recognition, was not confirmed by the present experiments: There was no effect of morphological family size or frequency in any of the RT-analyses, and no interactions indicating a difference between the two experiments.

In the All Items Analysis, the parallel between the morphological families of the complex and the simple words can be defined in different ways: The position-specific family of the first constituent (i.e. the words where the first constituent appears as a first constituent), the family of the second constituent, or the sum of these two were all tested as equivalents to the standard families of the simple words. Neither type or token counts of these measures had any effect in the RT-analysis, also not in interactions with sex, ISI or complexity type. In the All Items Error Analysis (see table 7.4), there was an effect of first constituent family size: Words with more first constituent family members were less error-prone than words with fewer, i.e. an expected facilitatory effect in both Experiments 3a and 3b. A large first constituent family does not make responses faster, as indicated by the absence of an effect in the RT-analyses. This may be because a relatively precise word identification is made, where a possible beneficial early effect of family activation could be cancelled out by later competition from the family members. For the correctness, however, the competition effect is irrelevant; instead the activation of many real words in the mental lexicon makes a correct yes-decision more likely. This is to some extent similar to Experiments 1a and 1b, where the inhibitory component of morphological family found in the RT-analyses disappeared in the error analyses.

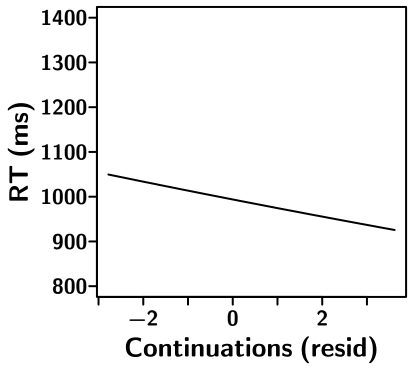

Figure 7.3: Partial effect of the number of morphologically related continuation forms in the analysis of All Items. The continuations measure is decorrelated from whole-word frequency.

For the complex words, both first and second constituent families, as type or token counts, were tested but found non-significant in both the RT and the error analyses. No family measure entered into any interaction. One reason why family size and frequency had no significant effects could be that these are raw measures that do not take into account onset-alignment of the family members with the target, thus disregarding the change of the cohorts over time which is central to auditory processing. In contrast, the set of morphologically related continuation forms only includes onset-aligned family members, and may therefore be a more appropriate measure for auditory processing. As fig. 7.3 shows, there was indeed a facilitatory effect of the continuations measure. This was significant in all analyses: the more continuation forms a word had, the faster and the more accurate the response. The continuations count correlated significantly with whole-word frequency, and the measure used was therefore the residuals of a regression model with continuations as a function of whole-word frequency. There were no significant interactions between continuations and sex, complexity type, ISI or whole-word frequency.

The continuations are a subset of the morphological family, in that they include only onset-aligned morphological family members, but they differ from the standard family size by also including inflectional forms. However, given the relatively small number of inflectional suffixes in Danish, the effect is unlikely to be driven exclusively by the inflectional forms of the target.

The differences between the family effects in the three experiments are further discussed in chapter 8.

7.3.3 Uniqueness points

The second main aim of the experiments was to further investigate the effects of two new UPs, attempting to replicate the effects for prefixed and particle prefixed words found in Experiment 1a, and to investigate the role of the two UPs in compound processing. Only in the Complex Analysis could both the two new UPs be tested, but duration and UPs are also essential variables to control in the analysis of all items, so, as in the previous experiments, both UP1 and CUP were included in the All Items Analysis, although they coincide for the simple words. UP1 is the point where the target becomes uniquely distinguishable from morphologically unrelated competitors, excluding continuation forms of the first constituent from the competition cohort. The CUP measures competition from morphologically related words, including the continuation forms of the first constituent in the competition cohort, but excluding continuation forms of the whole word. For the prefixed and particle prefixed words, the CUP corresponds to the standardly defined UP, while for the compound words, there is to my knowledge no standard definition of UP. This contrasts with the suffixed words, for which it is UP1 that corresponds to the standard UP. The two UPs are shown in fig. 2.2 on p. 33.

As in the previous experiments, the two uniqueness points and duration were recalculated as three non-overlapping parts of the signal to reduce collinearity (cf. sections 4.4.2, 5.3.5 and 6.3.5). For the complex words, there were effects of all three measures, illustrated in fig. 7.4. The non-linear effect of UP1 is shown in the top left panel; the initial flatness of this effect confirms the idea that a certain amount of acoustical information is necessary for a lexical decision to be made, therefore there is no benefit of a very early UP1. In Experiments 1a and 2, there were indications of a stronger non-linearity for simple than for complex words; this is not the case in the present experiments.

The Complex Analysis also confirms the role of the Complex UP: There was a significant effect of the distance from UP1 to CUP, as shown in the top right panel of fig. 7.4. As in Experiment 2, the effect was non-linear, with a smaller effect of a CUP occurring very shortly after UP1. Among the complex items, there are eight words where UP1 and CUP coincide, i.e. with a UP1 to CUP value of zero, which could be responsible for the non-linearity. However, the effect was also significant when these outliers were excluded from the analysis, and all other variables also remained significant without these outliers. In fig. 7.4, the horizontal axis does not include the value of the outliers, since that would suggest a stronger non-linearity than is actually justified.

A similar non-linearity was observed in Experiment 2 (see the bottom left panel of fig. 6.5 on p. 170). These non-linearities indicate that competing cohort members become more activated and therefore more inhibitory the

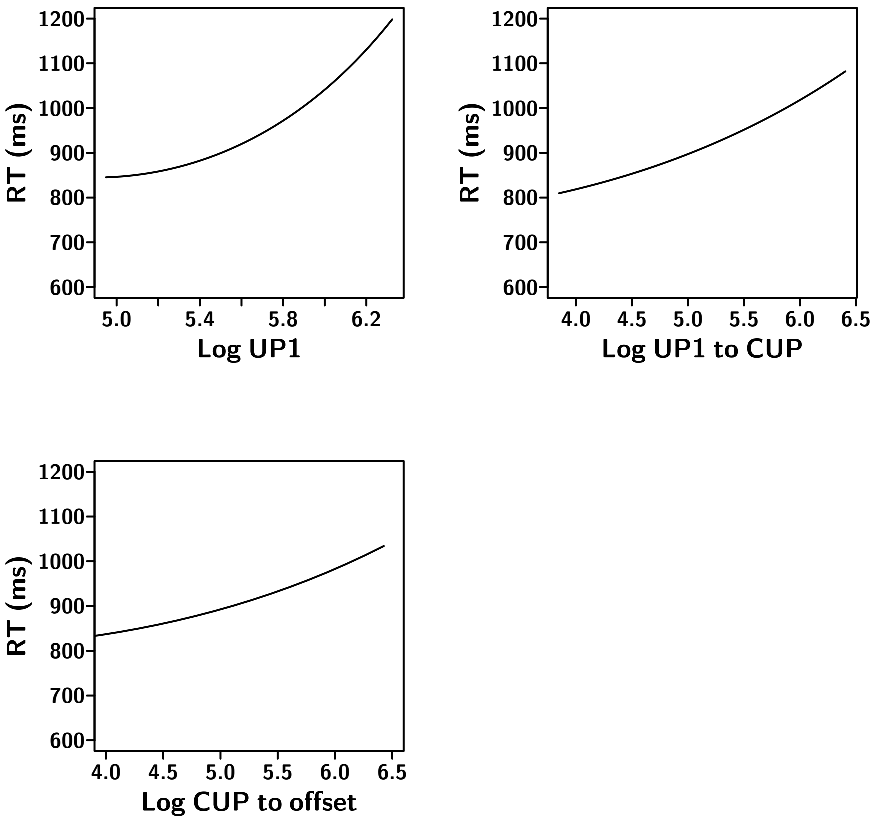

Figure 7.4: Partial effects of UP1, CUP and duration in the Complex Analysis. To reduce collinearity, CUP is measured as distance from UP1 to CUP and duration as distance from CUP to offset. The units on the horizontal axes are log ms. The horizontal axes in the top right and the bottom left panels are shortened to exclude eight and three outliers, respectively, with values of zero. Had these outliers been included, the non-linearity of the effects would appear misleadingly large.

longer they have been consistent with the input. A further factor that could be involved is that, although CUP strictly speaking measures the duration of the competition, it also correlates with the number of related words: the longer the target is ambiguous, indicated by a late CUP, the more competitors it generally has, and these competitors strengthen the inhibitory effect of a late CUP. The non-linearity was not quite significant for Experiment 1.

Importantly, there were no significant interactions between complexity type and either of the UP-variables: both were clearly significant for all three types of complex words. This strongly confirms that the new UPs are relevant for all types of complex words in Danish: inflectional, derivational and compound words, and both for prefixed and suffixed words. To my knowledge, this is the first study of UP-effects in compound words.

The effect of the distance from CUP to word offset was also non-linear, as shown in the bottom panel of fig. 7.4. The non-linearity was also found without three outliers with CUP to offset distance of 0, and the horizontal axis of the bottom right panel of fig. 7.4 was restricted to exclude these potentially misleading outliers. The non-linearity of this effect is not strong, but it suggests that only if a certain amount of material is encountered after the CUP — at which point it should be clear what the second morpheme of the word is — does it have a serious inhibitory effect. In Experiment 2, this effect was also non-linear, but with a clearer initial absence of effect, in a range of CUP to offset distances that was lower than those for the present experiment.

The All Items Analysis shows a similar pattern to that of the Complex Analysis, with effects of both UPs and duration. The inclusion of both UPs, although they coincide for the simple words, is not unproblematic, but it is the best way of accounting for this important source of variance among the complex words also in the All Items Analysis, as argued in section 5.3.5. For the simple words, the distance from UP1 to CUP is zero, resulting in a bimodal distribution of this variable for which it is inappropriate to use a quadratic parameter like that applied in the Complex Analysis.

The error analyses only showed two form effects: In the All Items Error analysis, there was a strong and significant effect of duration (which was not recalculated as distance from CUP to offset, since no UP-measures were significant), such that longer words were less error-prone. This is surprising at first sight, since none of the other experiments showed any effects of duration on correctness. However, the words in this experiment were longer, both in terms of mean and range, which may be responsible for this difference between the experiments. Intuitively, longer words should be more difficult to understand, and hence more error-prone, but once variables such as frequency are controlled, the longer words are more word-like, while longer nonwords are more nonword-like. This is so particularly because the longest nonwords were those based on the compounds which, in contrast to those based on prefixed and particle prefixed items, contained no real constituents. In other words, the contrast between words and nonwords was clearer for longer than for shorter words, resulting in more correct responses for longer words. This latter fact, which holds only for the present items, could also help explain why the effect is only significant for this experiment.

The Complex Error Analysis showed an effect of the frequency of the bigram straddling the juncture between the two constituents of the complex words. As described in section 5.3.6, this variable is inversely correlated with the salience of the juncture: The higher the frequency of the bigram, the more likely it is to occur inside a word, and the less likely it is to be perceived as a boundary. In the visual Experiment 1b, this visual measure

Table 7.6: Correlations between the form-related variables for the complex items: whole-word and constituent lengths, UP1, CUP and second constituent UP.

	Length	Length1	Length2	UP1	CUP	UP2
Whole-word length	1.00					
First constituent length	0.70	1.00				
Second constituent length	0.44	-0.33	1.00			
UP1	0.38	0.59	-0.24	1.00		
CUP	0.54	0.62	-0.07	0.23	1.00	
Second constituent UP	0.72	0.81	-0.06	0.52	0.61	1.00

had an inhibitory effect such that more frequent bigrams at the juncture resulted in slower responses. In the present experiment, words with higher bigram frequencies — and thus less salient boundaries — were less error-prone (significant effect in the Complex Error Analysis, see table 7.5), but not recognised faster (no significant effect in the Complex RT-analysis, see table 7.3). This is probably a result of bigrams with lower frequencies occurring often in the nonwords, particularly at the juncture between the real prefix and the nonsense stems in the "complex" nonwords when compared to the junctures in the real prefixed words. This means that real complex words with low juncture bigram frequencies are more like the nonwords in the experiment and therefore more error-prone. The effect of juncture bigram frequency is not one of the most robust effects, but there is occasional evidence (in Experiment 1b and in the present experiment) that it plays a role.

UP1, CUP and word duration are not the only possible UP- and duration measures for the complex words, but they have received strong support in the previous experiments and are therefore included in the final analyses reported in tables 7.2 and 7.3 above. The additional or alternative measures that could be of interest are the duration of the constituents, reported to have an effect on nonword rejection times in lexical decision by Laudanna and Burani (1995), and the UP of the second constituent as an independent word, corresponding to the root UP investigated for prefixed words by Tyler et al. (1988) and Schriefers et al. (1991). However, collinearity between the full set of UPs and durations is extremely large, with a κ of 103 between the logged values, and strong pairwise correlations, as shown in table 7.6. This collinearity made it impossible to reach a stable regression model that could determine which combination of variables provided the best analysis of the data. In order to investigate whether the inclusion of these additional variables substantially changed the other effects in the Complex Analysis while avoiding the harmful collinearity between them, the variables were entered into a Principal Components Analysis and the best principal components

(PCs) based on these form variables (henceforth FormPCs) were used in a separate analysis of the RTs to complex items. This analysis showed the same significant effects as the model summarised in table 7.3, with the exception of the interaction between ISI and complexity type, see section 7.3.5 below.

In addition to confirming the model summarised in table 7.3, the model involving the FormPCs could be informative in itself, though interpretation of PCs is not always straightforward. The FormPC-model showed effects of FormPC1 and FormPC2: FormPC1 reflects inhibitory effects of all the form variables summarised in table 7.6 except second constituent duration; the effect is non-linear, but because of the many contributing variables, this non-linearity — and the effect in general — is not readily interpretable. The non-linear effect of FormPC2 is clearer, reflecting effects of whole-word and second constituent duration, which are stronger in the lower than in the higher end of the range. In sum, the FormPC-model turns out not to be very informative in itself, but it still serves to confirm the significant effects of other variables in the Complex Analysis.

7.3.4 Frequency

There were normal facilitatory effects of whole-word frequency in all analyses (see tables 7.2 to 7.5), though the effect was only borderline significant for the reference levels variable ISI and complexity type simple in the All Items Error Analysis as discussed below. There were no effects of constituent frequencies in any of the analyses.

In these as in the analyses of the previous experiments, the preferred frequency measure for the whole words was the frequency of the base form of the whole word rather than of the lemma, mainly because this is the form actually encountered in the experiments. Generally, the two measures are very highly correlated and produce very similar results in the analyses.

Whole-word frequency entered into a number of interesting interactions: In both RT-analyses and in the analysis of correctness on all items, frequency interacted with ISI, such that there were significantly stronger frequency effects in the slower Experiment 3b than in the faster Experiment 3a. This interaction is illustrated for the complex items in the left panel of fig. 7.5; the same pattern was found in the All Items Analysis. It seems that the slower pace of Experiment 3b and consequent longer reaction times (though the difference in RT was not entirely clear, as discussed in section 7.3.1 above) result in better activation of lexical memory representations. The interaction could also be interpreted as evidence that the frequency effects observed in lexical decision tasks are partly caused by decision rather than recognition processes (Balota and Chumbley, 1984), such that longer decision times result in larger frequency effects.

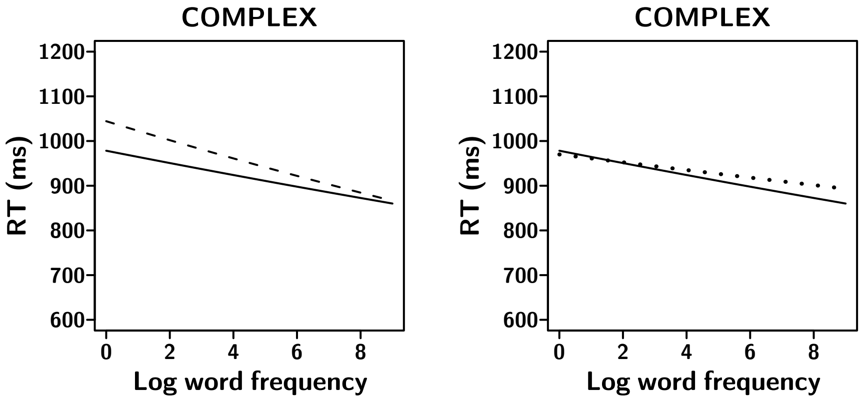

Figure 7.5: Partial effects of whole-word frequency for the complex items. The left panel shows the frequency effect for variable ISI (Experiment 3a) with the solid line and for fixed ISI (Experiment 3b) with the dashed line; this is based on the Complex Analysis, but the same pattern is found in the All Items Analysis. The right panel shows the effect of frequency for female participants (solid line) and for male participants (dotted line) for the complex items; this difference is not significant in the All Items Analysis which also includes the simple words.

In the Complex Analysis, whole-word frequency also interacted significantly with the sex of the participants, with females showing stronger frequency effects than males, as shown by the solid and the dotted lines, respectively, in the right panel of fig. 7.5. In the DP-Model of Ullman (2001, 2004), stronger frequency effects for females are taken as evidence that females are more likely to process complex words as wholes, due to their superior verbal memory (Kimura, 1999), while men are more likely to process them through decomposition. However, the whole-word frequency effects for the males were significant, just weaker than the effects for the females, so males also seem to rely to some extent on whole-word processing of complex words. For simple words, the difference between whole-word and decompositional processing is not relevant, and the inclusion of the simple words in the All Items Analysis is the likely reason why the interaction of sex and frequency was not significant in the All Items Analysis. The results are thus broadly speaking in line with the theory (Ullman, 2004, 2001) and results (Ullman et al., 2002) of Ullman, although this theory does not predict significant whole-word frequency effects for males. The sex by frequency interaction found in Experiment 2 shows that the verbal memory of females seems also to be involved in constituent frequency effects. The differences between the experiments are discussed in section 8.3.

In the All Items Error Analysis (see table 7.4), the main effect of whole-word frequency for the reference levels simple words and variable ISI was only marginally significant, but for Experiment 3b with the fixed ISI, the frequency effect was significantly more facilitatory (as in the RT-analyses). Additionally, the frequency effects were stronger for the complex words, though only significantly so for the compounds relative to the simple words. This is similar to the stronger frequency effects for complex items observed in Experiment 1a, but only occurs in the error analysis of the present experiment. These interactions did not reach significance in the Complex Error Analysis where there was a straightforward facilitatory main effect of frequency.

In addition to the questions that can be answered by the analyses of either all the items or the complex items, there are issues that can only be resolved by considering smaller subsets of the items. One such subanalysis considers only the compounds and particle prefixed words, to investigate whether the frequency of the first constituent as an independent word would affect word recognition. Such an effect is not likely to be significant since second constituent frequency was not significant in the analysis of the complex items and because the family of the first constituent was also non-significant, and indeed the subanalysis showed no constituent frequency effects, neither as main nor interaction effects.

The other potentially relevant subanalysis investigates whether there were effects of affix productivity measures which are only relevant for the prefixed and particle prefixed words. Since the productivity measures are based on type, token and hapax frequencies of the affixes (see 3.4.2) and since the Complex Analysis showed no effects of the type and token frequencies of the first constituents (discussed as first constituent family size and family frequency above, section 7.3.2), no effects of the productivity measures were expected in this subanalysis and none were found.

7.3.5 The complexity advantage

The processing advantage for complex words observed in Experiments 1 and 2 was not a main concern of the present experiments. However, since simple words were included in the experiments, they were chosen so that they could be analysed in an All Items Analysis, which allowed comparison of the different types. In this All Items Analysis, complexity interacted with trial number as described above. The main effect of complexity type was significant for prefixed and particle prefixed words relative to simple words but not for the compounds, when all co-variates including trial were set to zero. However, fig. 7.6 shows that the learning effect for the complex words increased the advantage for the complex relative to the simple words as the experiments progressed. The figure also suggests that the complexity advantage was significant for the compounds relative to the simple words once participants were habituated to the compounds; the advantage for the prefixed and particle prefixed words was significant from the first trial (indicated by the significance of the contrast coefficients for prefixes and particles in

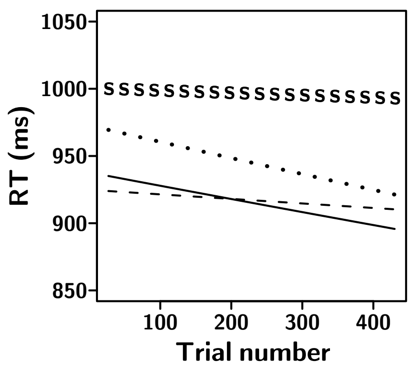

Figure 7.6: Interaction between complexity type and trial number in the All Items Analysis. The s's represent the simple words, the dotted line the compounds, the dashed line the particle prefixed words and the solid line the prefixed words. Note that scale of the vertical axis is reduced to 200 ms.

table 7.2) and, in the case of the prefixed words, became significantly more so as the experiments progressed. Fig. 7.6 is a repetition of the bottom right panel of fig. 7.2, here with a reduced vertical axis scale of 200 ms to make the effects more clearly discernible. The analysis confirms the advantage for complex over simple words found in the previous experiments.

Fig. 7.6 suggests that the compounds were recognised more slowly than the two types of prefixed words. This is not confirmed in the Complex Analysis where a three-way interaction between complexity type, ISI and trial was significant. When ISI was variable (Experiment 3a), there were no differences between the three types. When ISI was fixed (Experiment 3b), the particle prefixed words were recognised faster than the compounds early in the experiment, but unlike the compounds which showed a training effect, the particle prefixed words were recognised more slowly as the experiment progressed, and the difference disappeared. These gradually longer RTs to the particle prefixed words were not significant in the All Items Analysis. This pattern is illustrated in fig. 7.7, but even on the reduced vertical axis scale of 200 ms, the effects are small. When a different orthogonalisation of the UP- and duration variables was used, in the shape of FormPCs, the slopes for trial for the different types were slightly different. In sum, no strong conclusions can be drawn about the differences between the three types based on the present results; clearly, no general, categorical distinctions between prefixed, particle prefixed and compound words are supported.

In the All Items Error Analysis, there was no main effect of complexity type, but an interaction between complexity type and whole-word frequency, as discussed in the previous section (7.3.4). In the Complex Error Analysis, this complexity type by frequency interaction was only borderline significant,

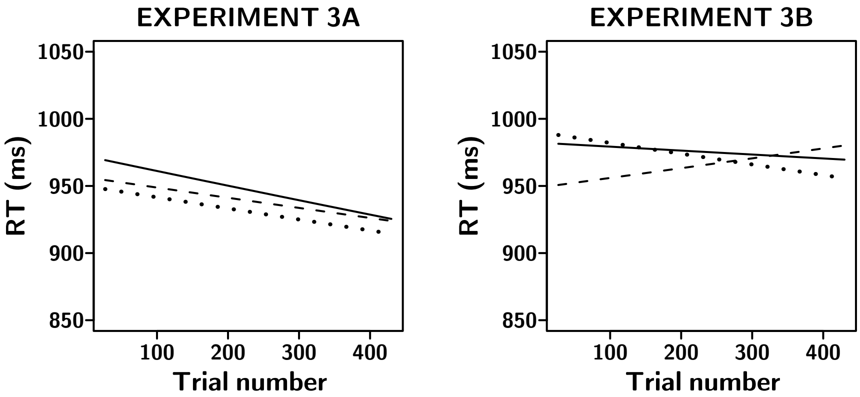

Figure 7.7: Interaction between affix type and trial number in the Complex Analysis. The dotted line represents the compounds, the dashed line the particle prefixed words and the solid line the prefixed words. Note that the scale of the vertical axis is reduced to 200 ms.

and therefore not included in the final analysis; instead, there was a main effect of complexity type (see table 7.3), with the prefixed words being more error-prone than the compounds which were the reference level.

7.4 Conclusion and summary

The present experiments set out to explore whether ISI could explain the differences in the morphological family effects between the previous Experiments 1a and 2. The hypothesis that ISI would interact with morphological family was not confirmed: there were no effects of morphological family size on reaction time at all, and the facilitatory effect of the number of continuations did not vary with ISI. The different family effects are further discussed in section 8.2.

Although the first main hypothesis was not confirmed, the experiments still are informative in a number of ways, of which two are central:

Firstly, the effects of the new UPs are confirmed for a new set of items and extended also to compound words. The distinction between UP1 and CUP gives a more detailed picture of the development of the competition cohorts over the course of word recognition and provides a way of drawing in the morphological structure of the words, making the UP-construct more relevant for these types of words.

Secondly, whole-word frequency interacted with both sex and ISI. The interaction with sex signifies stronger whole-word frequency effects for females; this is in line with the results of Ullman et al. (2002). In contrast, no effects of constituent frequencies were observed for either sex. Whole-word frequency also interacted with ISI, such that there were stronger frequency effects in the slower Experiment 3b with the fixed ISI. This can be understood both as a result of better access to lexical representations when the pace of the experiment was less forced, and as a result of elongated decision processes in the slower-paced task.

Finally, the experiments showed differences between the complexity types, with advantages for all three types of complex words, though in the case of the compounds only after some habituation with this type of item.

Chapter 8

General discussion

8.1 Outline

This chapter draws together the results of the experiments reported in the previous chapters for a coherent picture of morphological processing in Danish and more generally, discussing similarities and differences between the different experiments and their relation to the literature. Several of the present results demonstrate the functionality of morphology in lexical processing; these results are discussed in section 8.2. Section 8.3 focuses on the frequency effects observed in the experiments. The UP-effects and their implications for lexical processing are discussed in section 8.4. Section 8.5 outlines perspectives for future research, while section 8.6 provides a final summary. Throughout, the focus of the discussion is on those results that recur in several of the present experiments or have parallels in the experimental literature.

8.2 The functionality of morphology

The most convincing piece of evidence in favour of the functionality of morphology in word recognition is the processing advantage for complex over simple words that emerged across all three experiments. In Experiments 1 and 3, complexity type entered into various interactions, indicating that the advantage for complex words was modulated by other variables. In Experiment 1, the interaction between complexity type and UP1 suggested that the complexity advantage is stronger when competition from unrelated words, indexed by UP1, is resolved early. In Experiment 3, the complexity advantage for compounds was initially small, but it became larger as the experiment progressed. The overall pattern across the three experiments was an advantage for various types of complex words (inflected, derived and compound words) over simple words. In all the experiments, the mean RT for the group of simple words was lower than the mean RT for the group of complex words, but the analyses show that this difference is due to distributional factors such as length and frequency. Once these distributional factors are taken into account, complex words show a processing advantage.

The complexity advantage suggests that morphologically complex words benefit from the fact that both whole-word and morphemic information may be involved in the recognition process. Phrased in terms of a route metaphor, complex words may be recognised both via a whole-word and via a morphemic route. The advantage associated with the availability of two processing routes could arise in two ways: either because of statistical facilitation or because of interaction between the routes.

Statistical facilitation between recognition routes is the explanation proposed by Bertram et al. (1999) for the processing advantage they observe for certain complex words. Statistical facilitation between two processing routes arises when the distributions of the recognition times produced by the two routes overlap sufficiently: if one route is slow for a particular word, the other route may be fast, so that, considered across a group of words, recognition is faster when two routes are available. In contrast, if one route is consistently faster than the other — i.e. if the distributions of processing times for the individual routes do not overlap sufficiently — statistical facilitation will not arise.

Statistical facilitation between two processing routes requires that each route operates independently of the other. However, evidence is accumulating that the processing routes are not independent, indicating that the complexity advantage is not due to statistical facilitation. In the present experiments, the interaction between whole-word and second constituent frequency in Experiments 1a and 2 went in the opposite direction to what would be predicted based on statistical facilitation between independent processing routes. In these experiments, each frequency had a stronger facilitatory effect when the value of the other frequency was low. In contrast, words with high values on both frequencies were at a disadvantage. Under the assumption of statistical facilitation between independent processing routes, the prediction would be that high whole-word frequency and high constituent frequency would make for optimal processing, but the experiments did in fact show the opposite pattern. Similar interactions are also reported in the recent literature, as outlined in section 2.4. This means that the complexity advantage observed cannot be explained in terms of statistical facilitation.

This evidence of interacting routes makes the processing advantage for the complex words likely also to be an effect of interaction between processing routes: in the recognition of complex words, multiple types of information, both whole-word and morpheme-related, are activated in parallel, and these types of information interact to allow faster recognition of complex words. The complexity advantage itself does not necessitate an explanation in terms of route interaction, but other results indicate that route interaction does take place, and this in turn explains the complexity advantage. Importantly, although complex words as a group are recognised faster than simple words, complex words of high whole-word and high constituent frequency are at a disadvantage relative to other complex words. In other words, the benefit of having two recognition routes applies only up to a point. When both recognition routes are highly active, it seems that competition arises, as further discussed in section 8.3.

Such an interactive model architecture may be characterised as a dual-route model with interaction between a morphemic and a whole-word route, as proposed by Wurm (1997) for auditory results and by Pollatsek et al. (2003) for visual eye-movement results. It may also be called an interactive multiple-route model (Kuperman et al., 2008b) because there are several sources of morphemic information, for instance paradigms on different levels of generality (Kostic, Markovic and Baucal, 2003; Milin, Kuperman, Kostic and Baayen, 2008).

The complexity advantage is in accordance with eye-movement results from Dutch suggesting that derived words were recognised faster than simple words (in terms of gaze-duration, Kuperman p.c.) and with the advantage reported for Italian by Burani and Thornton (2003) for derived words with high-frequency roots compared to simple words. These processing advantages for complex words contrast with several studies of Finnish (e.g. Lehtonen et al., 2007; Soveri et al., 2007) that showed slower recognition of inflected than of simple words, as described in section 2.6.1. This difference is likely to be caused by differences between the two languages. For Danish, I argue that the complexity advantage is caused by the availability of two interacting recognition routes. For Finnish, whose morphology is much richer, the whole-word memory storage for complex words could be less pervasive than for Danish (Lehtonen and Laine, 2003; Soveri et al., 2007; though see also Vannest et al., 2002), with the result that this source of information is less available in the recognition of inflected words in Finnish than in Danish. The idea that whole-word storage is more or less extensive is problematic from an exemplar-based perspective, but the difference between Finnish and Danish may also be explained without explicit reference to the degree of whole-word storage, by arguing that Finnish readers may be more likely to rely on morphemic generalisations, because of the richness of Finnish morphology. Not only may recognition of morphologically complex words in Finnish be more dependent on a single recognition route, it may also be the case that this morphemic route is sometimes relatively slow, due to the extensive stem and affix allomorphy (see for instance Järvikivi et al., 2006) and the complex vowel harmony patterns (see e.g. Bertram, Pollatsek and Hyönä, 2004) that are characteristic of Finnish.

The complexity advantage in Danish was found for all types of affixed words, but two findings indicate that it is stronger for suffixed than for prefixed words: Firstly, when comparing the experiments, the complexity advantage was clearly the largest in Experiment 2 where the complex words were all suffixed (compare fig. 6.2 on p. 160 with fig. 5.2 on p. 122 and fig. 7.6 on p. 200). Secondly, in Experiment 1, secondary analyses of all items revealed that the complexity advantage was carried more by the suffixed than by the prefixed and particle prefixed words. The same pattern was evident in the analysis of the complex words alone in Experiment 1a, which showed that the suffixed words were recognised significantly faster than both types of prefixed words. That is, when the stem occurs first, the complexity advantage is stronger. This indicates that the complexity advantage is driven more by stem morphemes than by affixes, which is natural given that stems are generally longer, formally more salient and semantically richer than affixes. Stems also carry primary stress more systematically in the suffixed than in the prefixed words, which may contribute to the prominence of the stem in auditory processing and consequently to the advantage for suffixed words. The stronger complexity advantage for the suffixed words is in line with the general preference for suffixation across the languages of the world (Cutler et al., 1985) and with the experimental evidence of differences between the processing of prefixed and suffixed words reviewed in section 2.6.4.

An alternative but not inconsistent reason for the difference between suffixed and prefixed words in Experiment 1a could be that prefixes tend to activate larger and more diverse word-initial cohorts than the stems of suffixed words do. This could lead to a processing delay for prefixed relative to suffixed words, although the availability of the morphemic information along with the whole-word information still makes prefixed words easier to recognise than simple words.

The fact that suffixed words were the easiest to process leads to the prediction that compounds would also have an advantage relative to simple words (as reported for English by Inhoff et al., 1996; Ji et al., 2006; Fiorentino and Poeppel, 2007), because the first constituent of transparent compounds is relatively informative, parallel to the stem of transparent suffixed words. The results of Experiment 3 are not entirely clear-cut, but the general pattern is one of faster processing of compounds than of simple words, although only after an initial phase during which participants seemingly found the compounds relatively difficult.

The differences between compounds, prefixed and particle prefixed words were investigated in the separate analysis of the complex items of Experiment 3. However, the differences between the three types of complex words were small and varied with fixed vs. variable ISI, making it impossible to draw any firm conclusions with respect to the differences between compounds, prefixed and particle prefixed words. The disadvantage for particle prefixed words relative to other complex types observed in Experiment 1a (and discussed in section 5.3.2) was not replicated in Experiment 3.

In contrast to the difference between suffixed and prefixed words, there was no difference between inflected and derived words in terms of their advantage relative to simple words, a pattern which is similar to that observed by Baayen et al. (2007). The fact that the derived words in Experiment 2, which differed considerably in semantic transparency, showed a similar complexity advantage to the inflected words, which differed minimally in semantic transparency, indicates that the complexity advantage is unlikely to be the result of an efficient rule-mechanism or parsing route.

In addition to showing similar advantages relative to simple words, inflected and derived words did not differ significantly from each other. In the Complex Analysis of Experiment 2, where all relevant morphological variables could be drawn in, there was no categorical difference between inflected and derived words. In contrast, the graded variable affix frequency, which differs systematically between typical inflection and typical derivation, did show effects in Experiment 2. Semantic transparency, which also differs between inflection and derivation, was included in the analysis but was only marginally significant. Not only was there no main effect of the categorical difference between inflected and derived forms, the categorical complexity type factor also did not interact with other variables, notably frequency. This is problematic for the DP-Model of Ullman (2001), which operates with different memory systems for processing regular and irregular words and predicts stronger stem frequency effects for regular inflected words like those included in Experiment 2. The results of Experiment 2 favour a continuum-based rather than categorical view of different morphological processes: the continuous variable affix frequency was a significant predictor of processing time, while the categorical difference between inflection and derivation was not.

The facilitatory effects of morphological family size reported in the literature (see section 2.5) can also be interpreted as evidence of the functionality of morphology: the stronger the morphological connectivity of a word in the mental lexicon, the easier it is to recognise. However, these effects have primarily been found in visual word recognition. In auditory word recognition, the standard family size may be a less appropriate measure of facilitation from morphologically related words, because many members of the standard family are not part of the competition cohort and thus mismatch the target signal from word onset. This mismatch could result in inhibition from the standard family measure. In contrast, the morphological family members that are onset-aligned with the target may have a facilitatory effect in auditory word recognition (cf. the effects of position-specific family measures observed by De Jong et al., 2002 and Gagné and Spalding, 2008). The set of continuation forms consist only of such onset-aligned family members, and may therefore be a better measure of facilitation from the morphological family. Moreover, all continuations of a target word contain the whole target, and the measure is thus somewhat parallel to the dominated family of Moscoso del Prado Martín et al. (2004a) which showed stronger facilitatory effects in Finnish than the raw stem family, as discussed in section 5.3.4 and below.

The effect of the continuations was facilitatory in Experiments 1 and 3. Of these, Experiment 3 showed no effect of the standard family size, while Experiment 1a showed inhibition. The visual Experiment 1b showed facilitatory effects of both the standard family size (up to a point) and the morphologically related continuations. In both Experiments 1 and 3, the majority of the complex words were prefixed, though Experiment 1 also included both suffixed and simple words and Experiment 3 both compounds and simple words. In contrast, all the complex words in Experiment 2 were suffixed; for these suffixed words, the standard morphological family size effect was facilitatory, and there was no effect of the continuation forms.

It seems, then, that when the stem is encountered after a prefix, as it is for the prefixed words in the auditory Experiments 1a and 3, the family members that facilitate recognition are those that are onset-aligned with the target and contain both the prefix and the stem, i.e. the whole-word continuation forms. When the stem of a prefixed word is encountered, the competition cohort is already restricted to those words that share the onset, and the standard family size, which is based on the stem, does not facilitate. Although this effect was only found in the experiments that included prefixed words, it was not restricted to prefixed words: the number of continuations had a significantly facilitatory effect for both prefixed, suffixed and simple words in Experiment 1a and for prefixed, compound and simple words in Experiment 3.

In Experiment 1a, the standard family size was even inhibitory. Apparently, the mismatching members of the morphological family compete with the target for recognition; the more family members there are, the stronger this competition is, resulting in an inhibitory effect of morphological family size. This inhibitory effect was further explored by comparing the effects of the family of the stem with the family of the whole word. Moscoso del Prado Martín et al. (2004a) reason that the family members that contain the whole word rather than just the stem constitute a semantically more coherent subset of the family. Moreover, for the prefixed words, the stem family contains only those words that are not onset-aligned with the target. This semantically more distant, more diverse and less aligned stem family was responsible for most of the inhibition observed in Experiment 1a. This suggests that the onset-alignment does indeed play a role, along with the semantic diversity of the large families in Danish. The distinction between the raw family and the continuations may be particularly relevant in a language like Danish with its large morphological families, whereas Wurm et al. (2006) report effects of both family size and continuations in auditory lexical decision with prefixed words in Dutch.

In Experiment 2, all the complex words were suffixed; these showed a facilitatory effect of the standard family size and no effect of the continuations. In the visual Experiment 1b, both the standard family size and the continuations had facilitatory effects on recognition latency, though the standard family measure was non-linear and became inhibitory for the larger families. It seems that the morphological family size may have a facilitatory effect when the stem, on which the family size is based, is encountered early in processing, either because the word is suffixed or because it is presented visually. In auditory processing, this prominence of the stem in suffixed words may be further enhanced by the fact that the stems of suffixed words carry primary stress much more systematically than the stems of prefixed words. However, although there was a facilitatory family size effect for the suffixed words in Experiment 2, the presence of suffixed words is not sufficient to ensure a facilitatory family size effect: the inhibitory effect of this measure was found for all types of words, including suffixed words, in Experiment 1a. The facilitatory family effect would be the result of a general activation of the stem and its semantics which tends to be helpful when the stem is encountered early, i.e. for suffixed words and for visually presented words. This general activation is not helpful for prefixed words in auditory processing, because many of the activated family members are not members of the competition cohort.

One reason for the absence of a continuation effect in Experiment 2 could be that particularly the inflected but also the derived suffixed forms have fewer continuation forms than prefixed derived words do. Only very few affixes and no other constituents can follow an inflectional affix (except when the suffix doubles as a linking element, as *-er* in *blomst-er* ('flower-s') does in Experiment 2), and the restrictions on what can follow a derivational suffix are also stronger than the restrictions on continuations of the stems of prefixed words.

The possibility was also explored that the difference between the inhibitory effect of the standard family size in Experiment 1a and the facilitation observed in Experiments 1b and 2 was an effect of a difference in ISI. The fixed ISI in Experiment 1a is likely to prolong decision processes relative to Experiments 1b and 2 where ISI was variable and therefore generally shorter. These decision processes may in turn take more time if many family members are activated and compete with the target. ISI was manipulated for a single set of items in the auditory modality in Experiment 3, since Experiments 1a and 1b differed in modality and Experiments 1a and 2 in items, so that neither pairwise comparison could determine the isolated effect of ISI. However, Experiment 3 showed no effects of the standard family size, and no differences in family effects between fixed and variable ISI.

In sum, the experiments show that the most appropriate family measure for auditory processing of prefixed words is the set of morphologically related continuation forms, the family members that are part of the cohort at word offset. In contrast, the standard family size measure tends to show facilitatory effects for auditory recognition of suffixed words where the stem, on which the family measure is based, is encountered first. Both effects show that morphologically related words have a facilitatory effect when the development of the auditory input over time is taken into account, and the effects thus support the idea that morphological structure is functional. In auditory word recognition, this facilitatory effect of morphological family members is co-determined by cohort dynamics.

The effect of juncture bigram frequency in Experiment 1b may also be interpreted as evidence of the functionality of morphology: words with low juncture bigram frequency, which indicates high separability of the constituent morphemes, were recognised faster than words with high juncture bigram frequency. This suggests that the more available the constituent morphemes are, the easier is the recognition of the complex word. This facilitation was only observed in the visual Experiment 1b, but the non-replication of this effect is explicable in terms of modality: the bigram measures are visual, hence it makes sense that juncture bigram frequency was only significant in the single visual experiment. However, there was one additional effect of juncture bigram frequency, namely in the error analysis in Experiment 3, which went in the opposite direction: words with low juncture bigram frequencies, i.e. high morpheme separability, were more error-prone. This is presumably a result of the fact that nonwords contain low bigram frequencies, making words with low-frequency bigrams more nonword-like and hence more error-prone. Prefixes were present on both words and nonwords, so when participants encountered a relatively unfamiliar transition from a prefix to a stem, they were more likely to interpret the stem as a nonstem and categorise the word as a nonword, thus producing an error.

Facilitatory effects of semantic transparency (such as those observed by Wurm, 1997 and Baayen et al., 2007, see section 2.3.3) could also be understood as a reflex of the functionality of morphology. However, semantic transparency was only marginally significant in Experiment 2. In relation to the other experiments, where semantic transparency was not included as a co-variate, this is reassuring: the variable was only borderline significant in Experiment 2, although it was explicitly manipulated and varied much more than in the other experiments, for which only relatively transparent items were chosen. Any variations between the relatively transparent items of Experiments 1 and 3 are therefore unlikely to have an effect, given the small effect of the much larger variation of this variable in Experiment 2.

An important restriction on the complexity advantage is imposed by the interactions between whole-word and constituent frequencies observed in Experiments 1a and 2. Although complex words as a group were processed faster than simple words in both these experiments, complex words with high whole-word frequency and high second constituent frequency were at a disadvantage relative to those complex words which either had a high whole-word frequency or a high second constituent frequency. This implies that the advantage for the complex words applies only up to a certain point. The implications of these interactions are further discussed in the next section.

8.3 Interacting frequency effects and their implications

A first conclusion drawn from a joint consideration of the frequency effects in all three experiments is that stem frequency is not always the best measure of morphemic processing. This is contrary to the traditional interpretation but in accordance with the recent experimental literature reviewed in section 2.4. In the present experiments, stem frequency effects were only observed in Experiment 1, while other morphemic effects (e.g. the Complex UP) were found in all experiments. Affix frequency constitutes an alternative measure, as evidenced by the effect of this in Experiment 2. Across the auditory experiments, whole-word frequency had consistently strong and significant effects. Effects of the constituent frequencies were more unsystematic and, when the effects were significant, they entered into interactions with whole-word frequency.

Whole-word frequency may also be understood as a morphemic measure because it indexes the probability of the constituent morphemes occurring together, just like bigram frequency indexes the probability of two letters occurring together. Within an information-theoretical framework, the whole-word frequency of complex words can be simultaneously understood as the straightforward corpus probability of the whole complex word — corresponding to the traditional interpretation — and as the combinatorial probability of the constituent morphemes occurring together — i.e. a morphemic interpretation as proposed by Baayen et al. (2007; see also suggestions of a similar interpretation in Pollatsek et al., 2003). Such a dual interpretation is consistent with the results of Experiment 1a, where the complex words showed stronger whole-word frequency effects than the simple words (illustrated in the top right panel of fig. 5.3 on p. 126), and with the error analysis of Experiment 3. This suggests that facilitation from both the string and the combinatorial probabilities are expressed in the effect of whole-word frequency for complex words.

Experiment 1a showed an interaction for the complex words between the frequency of the whole word and the frequency of the stem. This interaction was significant in the analysis of all complex items but predominantly carried by the prefixed words. This indicates that the significant interaction was primarily between whole-word and second constituent frequency. The interaction between stem and whole-word frequency observed in Experiment 1a is therefore parallel to the interaction for the suffixed words in Experiment 2 between whole-word and suffix frequency: both are interactions between whole-word and second constituent frequency. The directions of the two interactions are also parallel (see 5.3 on p. 126 and fig. 6.3 on p. 162). The effect of each frequency was stronger when the other frequency assumed a low value. When whole-word frequency was low, second constituent frequency had a strong facilitatory effect; when second constituent frequency was low, whole-word frequency had a strong facilitatory effect. In contrast, when second constituent frequency was high, the effect of whole-word frequency was attenuated, and when whole-word frequency was high, the effect of second constituent frequency reversed to become inhibitory.

It seems that when one type of information is relatively weak, the other type of information becomes more important. In contrast, competition apparently arises when both routes are strong, leading to a processing delay for words with high whole-word and high second constituent frequency. There may be several different explanations of this apparent competition, none of which are directly supported by the present data and all of which are therefore rather speculative. The processing delay could arise because it takes time for the processing system to ascertain whether several very highly activated representations for the same word are compatible; this delay could be exacerbated if there is some semantic ambiguity between the different representations. Competition may also arise between the target word and other words containing the constituents of the target; this is more convincing when considering the affix frequency because this is based on the occurrences of the affix across many different words, but the high stem and whole-word frequencies may also activate other words containing the constituents in question.

The fact that the interaction is between whole-word and second constituent frequency indicates that the interpretation of the first constituent is modulated by the processing of the second constituent. In the visual experiments that show similar interactions, second constituent measures are sometimes involved (Baayen et al., 2007; Kuperman et al., 2008a), but the interactions between whole-word frequency and first constituent measures seem more systematic (Baayen et al., 2007; Kuperman et al., 2008a,b). These interactions between whole-word and first constituent frequency or family size reflect the processing of the second constituent given that the first has been processed, as do the results of Bertram, Pollatsek and Hyönä (2007). Bertram et al. found that the second constituents of Finnish compounds were read faster when the set of possible second constituent candidates was small given the first constituent (though overall reading time was not shorter for such words). The difference in the interactions in the visual vs. auditory word recognition is explicable in terms of differences in the rate of information uptake between the two modalities. In visual word recognition, the reader has more control over the information uptake than the listener does in auditory word recognition. The reader determines him/herself when to start processing a second constituent. Hypotheses about the second constituent may be generated while the first constituent is processed, before moving on to the second constituent, with the result that processing of the second constituent may be restricted by the processing of the first constituent. In auditory word recognition, the onset of the second constituent, and thus the time available to process the first constituent and its implications for the second, is not controlled by the listener. However, in single-word tasks like the present lexical decision tasks, the listener does control how much time is spent on processing the second constituent, and this processing may co-determine the interpretation of the first constituent and the whole combination, resulting in interactions between second constituent and whole-word frequency.

The interpretation of the interactions in both the auditory and the visual modality can follow from the understanding of whole-word frequency as the combinatorial probability of the two morphemes occurring together. In contrast, it is difficult to see why the interactions would arise if whole-word frequency is exclusively interpreted as the isolated long-term probability of the whole word in a corpus. Thus, the interactions are more consistent with the understanding of whole-word frequency as a combinatorial probability than with whole-word frequency as a long-term probability.

Whatever the precise interpretation of the different frequencies, the interactions between whole-word and second constituent frequencies demonstrate that the different sources of information that are involved in the recognition of morphologically complex words are not used independently. In other words, the interactions provide evidence that the processing routes interact rather than race. This in turn accounts for the advantage for complex over simple words which was found across the experiments, as argued above, though the availability of two processing routes seems to be an advantage only up to a certain point.

The different frequency interactions can be modelled within the information-theoretical framework of the Probabilistic Model of Information Structure (or PROMISE, Kuperman et al., 2008a, see section 2.4). This model accounts both for the dual function of whole-word frequency as an unconditional and a combinatorial probability, which is apparent in the interaction between whole-word frequency and complexity in Experiment 1a, and for the interactions between whole-word and constituent frequencies. However, although PROMISE models these interactions, it does not predict which interactions will be significant and what their direction will be for a particular experiment.

Turning to a different set of interactions, frequency also interacted with the sex of the participants. The investigation of interactions between frequency and the sex of the participants is inspired by the work of Ullman (e.g. Ullman et al., 2002), but the interactions actually observed are not entirely consistent with those reported by Ullman and colleagues. Ullman argues that women's superior verbal memory allows them to remember and process morphologically complex words as whole words, while men may have to rely more on combinatorial processing. According to Ullman et al. (2002), this results in whole-word frequency effects for regular words for females but not for males. In the present experiments, females did show significantly stronger whole-word frequency effects for complex words than males did, both in the error analysis in Experiment 2 and the RT analysis of Experiment 3. However, the effects of whole-word frequency were significant for the males, contrary to the predictions of Ullman. In the RT-analysis of Experiment 2, it was the interaction between whole-word and affix frequency that was stronger for females than for males. In other words, it seems that the superior verbal memory of females (Kimura, 1999) results not only in stronger whole-word frequency effects, but also in stronger affix frequency effects. This suggests, albeit indirectly, that the morphemic processing, for which the affix frequency effects provide evidence, is not the product of a separate processing mechanism, as the DP-Model of Ullman (2001) holds. The affix frequency effects are not predicted within the DP-Model, which focuses on stem frequency, but as indices of morphemic processing, the affix frequency effects in themselves are interpretable within the DP-Model. However, the interaction between whole-word and affix frequencies in Experiment 2 poses a challenge to the separate processing mechanisms of the DP-Model, a challenge which is strengthened by the parallel interaction in Experiment 1a between whole-word and stem frequency. The inhibition observed for words with high whole-word and high constituent frequency might be explained within the DP-Model in terms of problems with integrating information from different neural substrates, but this is not consistent with the complete absence of a categorical difference between derived and regular inflected words in Experiment 2. Moreover, it is unclear how the DP-model would account for the fact that such integration difficulties seem to be stronger for females than for males.

From a methodological point of view, the interactions involving sex observed here, as well as by Tabak et al. (2005), Kuperman et al. (2008b) and Ullman and colleagues, underline the importance of including this factor in experimental designs and analyses. It remains an empirical question whether the sex differences would generalise beyond the group of highly literate and highly educated individuals sampled for the present experiments, though the variety of sex differences in language processing summarised by Kimura (1999) suggests that they would.

The present experiments challenge the traditional view of stem frequency effects as crucial evidence of decompositional morphemic processing and whole-word frequency effects as evidence of continuous, whole-word-based processing. Instead, the different frequency effects and interactions observed are evidence of a flexible processing system in which conditional processing and interaction between different types of information are central. The finding of both whole-word and morphemic effects do not necessitate separate whole-word and morphemic representations; both effects may instead emerge from generalisations across word representations or across stored word exemplars

(Hay and Baayen, 2005). The conditional processing of one constituent given another is seen in the interactions between whole-word and second constituent frequency. This conditionality is one aspect of the functionality of morphology, though the functionality is only observed up to a point: words that both have very high whole-word and very high constituent frequencies are at a disadvantage. The flexibility and conditionality are inconsistent with sequential models, whether they consider morphological structure sublexical (e.g. Taft, 1994, 2004) or supralexical (Giraudo and Grainger, 2000, 2001, 2003b). The interactions observed also challenge the separate mechanisms posited within the DP-Model of Ullman (2001, 2004). In contrast, the interaction and conditionality evident in the frequency effects of the present experiments follow naturally from the PROMISE-model

8.4 Uniqueness points: conditionality and context in processing

All the auditory experiments showed strong and significant inhibitory effects of the two new UPs introduced in section 2.3.2, for all types of morphologically complex words. The fact that both uniqueness points are significant indicates that there are two key points during the winnowing of competitors from the cohort. UP1 is the point where early competition between morphologically unrelated lemmas in the mental lexicon is resolved. The CUP represents the later point in time at which morphologically related continuations of the first constituent are eliminated from the cohort.

UP1 is the point where a complex word deviates from all words that share its onset but not its first constituent. There were inhibitory effects of UP1 in all three experiments, with non-linearities in the All Items Analyses and in the Complex Analysis of Experiment 3 such that the inhibition became stronger for later UP1s. It seems that a certain amount of acoustical information may be necessary for a lexical decision to be made, therefore there is little or no difference between the early UP1s. In Experiment 1a, the benefit of an early UP1 for complex but not for simple words could also be understood in terms of the complexity advantage as outlined in section 5.3.2: the earlier the UP1 occurs, the earlier competition from non-related competitors is resolved, and the earlier morphemic information can contribute to recognition. This could also apply for the visual Experiment 1b where there was a small effect of UP1 (but not of CUP) for the complex but not the simple words. The complex words may also be more likely to be read in two fixations, triggering a UP1-effect for these words. In any case, the UP1-effect in Experiment 1b is evidence of the activation of acoustical memory representations also in reading.

The Complex UP or CUP is the point where words that share the first constituent of the complex target become inconsistent with the input. The CUP was indexed by the distance from UP1 to CUP to reduce collinearity; this measure showed consistently inhibitory effects: the later the ambiguity between the target and other words that share its first constituent, the longer the recognition time. The effect was linear in Experiment 1a and non-linear, with an initial flatness, in Experiments 2 and 3. The stronger effect later in the UP1 to CUP range in Experiments 2 and 3 could reflect both that competitors become stronger when they have been consistent with the input for a longer time, and that a late CUP is associated not only with late ambiguity but also often with more competitors.

The two UP-measures and word duration were included in the analyses as non-overlapping parts of the auditory signal in order to minimise collinearity. This means that part of the duration-effect may be picked up by the UP-measures, and conversely that the UP-effects may partly be simple effects of word duration. However, even when the full duration was added to the analyses instead of the distance from CUP to offset, both UP1 and CUP remained significant in all analyses, indicating that they do indeed measure two distinct kinds of competition.

A central characteristic of this UP-framework is that the processing of the second constituent is conditional on the first constituent having been accessed: the CUP is based on competition between those second constituent candidates that can co-occur with the first constituent. The fact that competition between unrelated lemmas is captured by the UP1 means that the CUP can be based exclusively on second constituent candidates, and thus index conditional processing of the second constituent given the first. This interpretation of the CUP is similar to the interpretation of whole-word frequency as a combinatorial probability: the CUP may reflect competition between whole-word representations, but only those whole-word representations that share the same first constituent. Seen in this way, the CUP-effect reflects morphemic relations between whole-word representations or exemplars in the mental lexicon.

The interpretation of the two UPs is unified across all types of words, whether their first and second constituents are free or bound. The competition between words sharing a first constituent seems to be the same whether that constituent is bound (the prefix of prefixed words) or not (the particle of particle prefixed words, the stem of suffixed words and the first constituent of compounds), as evidenced by the uniform effects of UP1 across these three types of complex words. Similarly, the competition indexed by the CUP does not seem to distinguish between free and bound constituents: targets which are suffixed continuations of a particular stem compete with other suffixed continuations of that stem as well as with compounds in which the stem is the first constituent and the second constituent is a free morpheme.

The effects of UP1 and CUP are remarkably consistent across the three auditory experiments and across the different types of complex words, in spite of the fact that neither UP corresponds clearly to the traditional single UP of Marslen-Wilson (1984). The traditional UP corresponds to UP1 for suffixed words and to CUP for prefixed and particle prefixed. The fact that the same effects of the two UPs are observed for the different types supports the argument that the new UPs provide a more uniform and more accurate account of lexical competition in complex words than the traditional UP does. The broad applicability of the new UPs is also an advantage relative to the combination of UP and CRUP of Wurm (1997) which only comes into play for a minority of prefixed words.

As evidence of conditional processing, the CUP is similar to the CRUP of Wurm (1997), and it can be interpreted within an interactive dual-route model, like Wurm's CRUP and like the frequency effects discussed above. Both early uniqueness from unrelated competitors and early uniqueness from related competitors make processing faster, indicating that whole-word and morphemic information interact. As in the case of the frequency effects, the two types of information need not be based on separate representations: both UP1 and CUP may be based on competition between whole-word representations or exemplars, but the CUP reflects competition which is restricted by morphemic information.

The logic of the CUP is similar to the logic underlying the continuations measure, although the way this logic is expressed differs between the two measures. Both measures are based on cohort members that are consistent with the input after an initial UP: the CUP is based on the stem continuations which are consistent with the target after UP1 but before word offset, and the continuations are those words which are consistent with the target at word offset. They differ in that the CUP measures uniqueness from the competing cohort members, whereas the continuations are a type count of the relevant cohort members. In the present experiments, the CUP had a consistently inhibitory effect, while facilitatory effects of continuations were found in two out of three experiments. Previous studies have shown different effects of continuation entropy: Wurm et al. (2006) found facilitation, Kemps et al. (2005a) inhibition. There may be two reasons for this discrepancy which are also related to the present results: Firstly, Wurm et al. include only morphologically related continuations, as is done here, whereas Kemps et al. include both related and unrelated continuation forms. Secondly, Wurm et al.'s stimuli are prefixed while Kemps et al.'s are simple. This means that the continuation cohorts of Kemps et al. consist of stem continuations which perhaps makes them more similar to the inhibitory CUP-cohort of the complex words in the present experiments, than to the continuations of the complex words.

The three cohorts on which the two UPs and the continuations measure are based represent a continuum of relatedness to the target word: the UP1-cohort includes only unrelated lemmas; the CUP-cohort contains words that are related to the target through its first constituent; and the continuations cohort consists of the words that are related to the target by containing the whole target word. All three are restricted to onset-aligned words. In contrast, standard morphological family measures include both words that are onset-aligned and words that are not. The distinction between dominated family members (those containing the whole word) and non-dominated family members (those that contain only the stem) is similar to the distinction between CUP and continuations, and the former did indeed carry most of the inhibitory effect of family size in Experiment 1a. In general, the effects of cohort-measures (two UPs and the continuations) were more robust in the experiments reported than effects of paradigms whose members are not onset-aligned. This contrasts with the visual word recognition literature where paradigm-effects are strong, as shown in section 2.5. This is probably mainly a result of the experiments being auditory, though UP1 and continuations also has significant effects on processing time in the visual Experiment 1b.

At the most general level, the UP-effects in the present experiments provide further evidence of the dynamic nature of lexical competition over time in auditory processing, demonstrated for instance by Wurm et al. (2006) and Magnuson et al. (2007). Like Wurm et al., the experiments demonstrate that morphological relatedness affects the competition process. More specifically, the effect of the Complex UP shows that processing is conditional, an insight that was implemented in the CRUP-measure of Wurm (1997), and which is extended to all types of complex words in the present CUP-measure.

8.5 Perspectives for future research

The two new UP-measures had clear effects in all experiments on Danish. A main objective of future research within the UP-framework would be to obtain documentation of these effects also for other languages. Judging from the strong effects of these measures in Danish, similar effects would be predicted for other languages with concatenative morphology. Given the possibilities of regression analyses, such investigations of UP1 and CUP may be based on existing datasets from various auditory tasks. Apart from attempting to replicate the UP-effects in other languages, further investigation of the relation between the two UPs, particularly the CUP, and the continuations measures is required; this could be done relatively straightforwardly by comparing stem-continuation cohorts for suffixed words with whole-word continuations cohorts for both simple and suffixed words. Finally, an investigation of the time-course of related vs. unrelated competition might be possible by investigating the two UPs in a visual-world paradigm, along the lines of Magnuson et al. (2007), using compounds as stimuli.

The visual-world paradigm might also be a way of studying the time-course of the different effects observed in the present experiments in more detail. Kuperman et al. (2008a) and Kuperman et al. (2008b) demonstrate for visual word recognition that different types of information are simultaneously active at a very early stage in word recognition and that they interact. Similar activation patterns and interactivity would be predicted for auditory processing, given the various interactions observed in the present experiments. However, the precise time-course of the different effects in auditory processing remain a topic for future investigation.

Different explanations of the inhibitory effect of morphological family size in Experiment 1a were discussed in sections 5.3.4 and 8.2. A way of further exploring the family effects in Danish could be to consider the relatedness and coherence of the morphological families of the three experiments through Latent Semantic Analysis (cf. Landauer, Foltz and Laham, 1998), in order to assess whether the items in Experiment 1a may differ in a way that contributes to the inhibition. Given that the same items were used in Experiments 1a and 1b, but showed different effects, this is unlikely to be the sole explanation of the phenomenon, but may nonetheless contribute to the understanding of this effect.

For practical reasons, the experiments reported all used the lexical decision task, but several of the effects could also be explored using other experimental paradigms. The visual-world paradigm mentioned above would be one option, naming another. The UP-effects may be investigated in auditory naming, and continuation effects in both visual and auditory naming. Visual naming and eye-tracking of reading might also be used to investigate to what extent the complexity advantage generalises to more naturalistic tasks. The tools established here for determining lexical variables for Danish words should be useful also for further experimental work on Danish.

8.6 Conclusions

The findings of the experiments reported in the previous chapters provide evidence of a flexible and interactive lexical processing system. Both whole-word and morphemic information is involved in the recognition of complex words, leading to a general processing advantage for complex relative to simple words. In the experiments, this functionality of morphology is also evident in the facilitatory effects of continuation forms and of junctural probability. In everyday language processing, the availability of several sources of information may be particularly important in allowing more efficient processing of low-frequency complex words; morphological structure and morphological relations thus contribute to the naming of rare concepts. Morphology also contributes to the creation and recognition of those ultimate low-frequency items, new words. In all cases, whether a word is novel or of high or low frequency, both morphemic and whole-word resources are used to maximise the opportunities for word recognition (Libben, 2006). The relative prominence of the different types of information varies, as indicated for instance by the frequency interactions. The functionality demonstrated in the present experiments is a local functionality, in that the availability of both morphemic and whole-word information is functional for the individual act of word recognition. The idea that morphological structure is globally functional because it partially alleviates the arbitrariness of the sign (outlined in section 1.4) is also in accordance with the empirical observations, but it remains more speculative.

Importantly, the processing of morphemes depends on the context in which they occur: No independent effects of either stem or affix frequency were observed. Instead, the processing of first constituents is conditional on the processing of second constituents, as evidenced by the interaction between whole-word and second constituent frequency in two of the three auditory experiments. Processing of the second constituent is conditional on the first constituent having been processed, as shown by the robust effects of the Complex UP across all types of complex words in all auditory experiments. This conditional processing may contribute to the functionality of morphology: morphology may aid processing partly because the internal structure of a particular word restricts the part of lexical memory that must be searched in the process of recognising that word. These effects also show how sensitive the auditory processing system is to the relative probabilities of different constituents and to the development of these probabilities over time.

The fact that so many different variables and interactions between them could be investigated in the same experiments is an advantage of the present work which comes with the use of regression designs. Particularly, the interactions between different frequencies demonstrate the advantages of a regression approach in which different interactions can relatively straightforwardly be investigated. The regression technique also allowed the control of considerable amounts of variation through the inclusion of random intercepts and random slopes, and through the possibility of including the various effects of experimental context. Moreover, the items are more representative of the language in general than those that can normally be included in factorial designs with their strict control criteria. Along with the other advantages of regression approaches outlined in chapter 4, this arguably makes the results more reliable. The inclusion of random intercepts and random slopes maximises the chance of generalisation of these results. However, two caveats are in order: Firstly, the participants in the experiments were all highly literate and highly educated, though not all of them were students. The group of participants are therefore not claimed to be representative of the whole population of speakers of Danish, but only a subset of this population. Secondly, the selection of items from the vocabulary was not entirely random, being limited to the words for which the relevant lexical statistics could be extracted.

The main results of the experiments reported above pose problems for both single- and dual-mechanism models. There is no evidence of the separate processing mechanisms for regular and irregular words that are central in the DP-Model, and the absence of any categorical differences between regular inflected words and more irregular derived words in Experiment 2 directly challenges the assumption of separate mechanisms. Also the interactions between whole-word and morphemic measures, which indicate interactivity in processing between whole-word and morphemic information, are problematic to the DP-Model. The interaction between sex and frequency observed in Experiment 3 is more in line with the DP-Model, but may also be interpreted in terms of males and females relying to a different degree on different but simultaneous and interactive processing strategies.

The results are more consistent with the single-mechanism models that constitute the opposite position to the dual-mechanism view implemented in the DP-Model. The effects observed in the experiments may be based on a lexicon without specific morphemic representations, and are thus consistent with one central aspect of the single-mechanism position, namely that whole-word and morphemic effects arise as a result of generalisations across the same set of distributed whole-word representations. However, morphological structure plays a much more prominent role in the present results — in the complexity advantage and the effects of frequencies and UPs — than it is usually awarded within single-mechanism models. Therefore, the specific claim that morphological structure is epiphenomenal of semantic and phonological relatedness, which is central to models such as the Convergence Account of Seidenberg and Gonnerman (2000), is not supported: the effects of semantic similarity were small, while morphemic effects were robust.

The problems are more severe for the dual-mechanism approach, but this is arguably because the predictions of the DP-Model are more precise and easier to test. The repeated claim in single-mechanism approaches (e.g. Seidenberg and Gonnerman, 2000) that morphological effects are caused by interacting effects of form and semantics is much vaguer and therefore more difficult to falsify. The current experiments do not directly falsify this hypothesis, but they do indicate a stronger role for morphological structure per se than is posited within the single-mechanism framework.

The results are not only relevant to the binary distinction between single- and dual-mechanism models but also to the continuum of models that are more narrowly concerned with morphological processing in word recognition.

The interactions between whole-word and constituent frequency effects challenge the idea of separate levels of morphemic and whole-word representation, whether the morphemic representations are sublexical (Taft, 1994, 2004) or supralexical (Giraudo and Grainger, 2000, 2001, 2003a). The decompositional model of Taft moreover suffers from problems with accounting for the conditional processing shown both by the present CUP-effects and by the effects of Wurm's (1997) CRUP. In contrast, the evidence of conditional processing in the present experiments adds to a growing body of experimental work that shows extensive interaction between whole-word information and different types of morphemic information. As discussed in the previous sections and in section 2.4, information-theoretical approaches constitute a promising way of modelling these effects and interactions, both on a behavioural (Kuperman et al., 2008a) and a neurological level (Moscoso del Prado Martín, 2007).

The results say little about the precise nature of the representations that are the basis of this flexible processing system. It is clear that both morphemic and whole-word information is involved in processing but, as argued by Hay and Baayen (2005), this does not necessitate morphemic representations that are separate from whole-word representations. The morphemic effects may be based on paradigmatic generalisations across the same whole-word representations that are the basis of the whole-word effects, i.e. a psycholinguistic version of the linguistic Word-and-Paradigm approach discussed in section 1.3. In fact, the extensive interactions observed arguably makes it more likely that the different effects do stem from generalisations across the same representations. Such whole-word representations may be relatively abstract or they may be exemplars. The latter position is favoured by the amount of subphonemic detail that has been shown elsewhere to be available to the listener (see section 2.3.2).

The present experiments provide a first investigation of on-line lexical processing in Danish and thus adds to our knowledge about that language. At the same time, the new data improve our understanding of morphological processing in auditory word recognition, which has been studied less than visual word recognition. The knowledge about morphological processing gained from these results contributes to our understanding of the impressively fast and accurate process of word recognition. The recognition of complex words draws on all available information resources, both morphemic and whole-word related, and morphological structure contributes to the recognition of rare words and the coining of new words. The study of morphologically complex words, particularly when considering the new UPs introduced here, also provides a way of documenting conditional processing on the word level and bears witness to the probabilistic and dynamic nature of auditory word recognition.

English Summary

The experimental work reported in the thesis is motivated by a fascination with the speed and accuracy of word recognition, on which the study of morphological processing provides a window. The experiments investigate the auditory modality, because this is the primary modality, and because it was a main object to study the changes in lexical competition over time which are characteristic of auditory processing. The language under investigation is Danish; this is a "new" language in the field and has various properties that are interesting for the study of morphological processing.

Chapter 1 describes the models that have been formulated to account for morphological processing. The contrast between storage of whole complex forms and computation based on constituent morphemes is central. Many models of morphological processing fall on a continuum ranging from complete computation to full storage of complex words. Between these two extremes, there are different hybrid positions which operate with a mixture of whole-word and morphemic processing. Among the hybrid models are dual-route models which include both a whole-word and a morphemic recognition route for morphologically complex words; these routes may operate consecutively, in parallel or in interaction. A partially different research tradition uses the study of morphological processing as a testing ground for general claims about the nature of human language. This tradition focuses on the distinction between regular and irregular inflection. Dual-mechanism models operate with distinct mental mechanisms for the processing of regular and irregular inflectional forms: Regularly inflected forms, like syntactic structures, are processed by the application of a symbolic rule in the procedural memory system. Irregularly inflected forms are stored in declarative memory, along with simple words. In contrast, connectionist single-mechanism models argue that both regular and irregular complex forms are processed by the same subsymbolic, associative memory system. Graded morphological effects are taken by proponents of single-mechanism models as evidence that morphological structure is epiphenomenal of semantics and phonology/orthography, but it is argued in section 1.3 that graded morphological effects may also arise as a consequence of paradigmatic relations between morphologically related whole-word exemplars. The information-theoretical PROMISE-model is a possible model of such graded and probabilistic morphological processing. Further, it is argued in section 1.4 that morphology may be functional to lexical processing, because the recognition of morphologically complex words may rely simultaneously on storage and computation, making processing more efficient and robust.

Chapter 2 discusses central effects in experiments on morphological processing. Effects of morphological similarity are often compared with effects of formal and semantic similarity. The effects of neighbourhoods, which include words that are formally similar to a target word at different points in the word, are generally inhibitory. In auditory word recognition, the onset-based similarity captured by the cohort concept plays a larger role. The cohort consists of the words that are consistent with the target word from its onset up to a certain point in the word. The uniqueness point (UP) occurs when the cohort contains only the target word; the later this UP occurs, the longer it takes to recognise the given word. The UP does not take morphological structure into account. Cohort-based measures that do take morphological structure into account are discussed in section 2.3.2. These include the Conditional Root Uniqueness Point (CRUP) which demonstrates that the processing of the stem of a prefixed word may be conditional on the prefix having been processed. The CRUP is relevant only to a minority of prefixed words, and two new UPs are therefore introduced which allow the investigation of similar conditional processing for all types of morphologically complex words. Effects of formal similarity may also be studied using the priming methodology. Morphological priming effects are generally larger than formal and semantic priming effects, and they have a different time course, but there is also evidence that morphological priming effects vary with the degree of semantic relatedness between prime and target, implying that morphology is not independent of semantics.

The comparison of stem and whole-word frequency effects is central to the study of morphological processing. Whole-word frequency effects are generally seen as the hallmark of non-morphemic processing of complex words, and stem frequency effects as the hallmark of morphemic processing. However, it is argued in section 2.4 that stem frequency is not always the best measure of morphemic processing. An alternative measure may be affix frequency. Moreover, whole-word frequency may be understood as morphological because it indexes the probability of constituent morphemes co-occurring.

Morphological family size is a type count of the derivations and compounds that share the stem of a target word. The more family members a word has, the faster it can generally be recognised, as described in section 2.5. The morphological family is not a morphological paradigm in the traditional sense but may reflect paradigmatic relations in the mental lexicon which are similar to those revealed by effects of inflectional paradigms.

Section 2.6 discusses differences between different types of words and different types of morphological operations: between simple and complex words; between inflection, derivation and compounding; between prefixed and suffixed words; and between regular and irregular inflection. The three first contrasts are relevant to the experiments on Danish reported in chapters 5 to 7.

Chapter 3 briefly reviews relevant aspects of the Danish language in section 3.2. *Korpus90/2000*, the Danish text corpus from which the lexical variables were extracted, is described in section 3.3, along with the substantial modifications to the corpus that were necessary. The extraction of different frequencies, morphological families, uniqueness points and lexical neighbourhoods is described in section 3.4.

Chapter 4 first motivates the choice of the lexical decision task for the experiments on Danish. Then, the reasons for the choice of regression designs and linear mixed-effects models are outlined in sections 4.3 and 4.4. Regression designs and analyses are more informative and statistically more powerful than the factorial designs traditionally used in the field.

Three lexical decision experiments on Danish are reported in chapters 5 to 7. Experiment 1 was run in both an auditory (1a) and a visual version (1b), the other experiments were auditory.

A main finding across all three experiments was an advantage for all types of morphologically complex words (inflected, derived and compounds) over simple words: when other relevant variables were taken into account, complex words were recognised faster than simple words. This is evidence that the morphological structure of complex words is functional in processing. It is hypothesised that this advantage arises because of the use of both morphemic and whole-word related information in the recognition of complex words. The advantage was stronger for suffixed words in which the more informative constituent occurs first. In Experiment 2, no categorical differences were observed between regular inflectional forms and more irregular derivational forms; this finding challenges the categorical distinction which is central to dual-mechanism models.

All three experiments also investigated frequency effects. In Experiment 1a, there were significantly stronger whole-word frequency effects for complex than for simple words, suggesting that whole-word frequency for complex words has both a non-morphological and a combinatorial component. In Experiments 1a and 2, whole-word and second constituent frequency interacted. When either frequency was low, the other frequency had a strong facilitatory effect on recognition time. In contrast, words for which both frequencies were high were at a disadvantage. When both types of information are highly active, competition seems to arise; potential reasons for this are outlined in chapters 6 and 8. The interactions indicate that the different types of information are not used independently, suggesting that the complexity advantage is also best understood in terms of interaction between recognition routes, rather than in terms of statistical facilitation. The fact that it was second constituent frequency that interacted with whole-word frequency suggests that the processing of the first constituent is to some extent conditional on the processing of the second constituent in auditory word recognition. In Experiments 2 and 3, whole-word frequency effects were significant for both sexes, but significantly stronger for females than for males. This result is probably a consequence of females' superior verbal memory and in line with the dual-mechanism model, but in Experiment 2, females also showed stronger affix frequency effects and a more pronounced frequency interaction than males, which is problematic for the dual-mechanism model. A final significant frequency effect was the effect of juncture bigram frequency in the visual Experiment 1b where words with more salient morpheme boundaries were recognised faster.

A main purpose of the three experiments was to seek experimental support for the two new UPs introduced in chapter 2. UP1 is the phoneme where competition from unrelated words is resolved, while the Complex UP (CUP) is the phoneme where words that share the first constituent of the target become incompatible with the input. These measures had strong and significant effects in all auditory experiments: the later the UP1 and CUP, the longer a word took to recognise. Even in the visual Experiment 1b, there was a significant effect of the auditory UP1-measure. The effects were similar for all types of complex words. The significance of both effects shows that both unrelated and related words enter into the competition process independently of each other. The CUP-effect is evidence of conditional processing: the competition between second constituents is conditional on the first constituent having been recognised.

Different effects of morphological family measures were found: In the auditory experiments that included prefixed words (1a and 3), the facilitatory effect of morphological family was restricted to the set of continuation forms which were onset-aligned and contained the whole target word. In Experiment 1a, the standard family size was even inhibitory. In contrast, in the visual Experiment 1b and for the suffixed words in Experiment 2, there was a facilitatory effect of morphological family size, at least up to a point; when the stem is encountered early in processing, the stem-based morphological family size measure seems to aid the recognition of the target. The different cohort dynamics for different types of words co-determine which family members may aid recognition. Experiments 3a and 3b investigated whether Inter-Stimulus Interval (fixed vs. variable) would affect the family effects, but this was not the case.

Finally, effects of experimental context were included in all experiments. There were significant effects of reaction time and correctness on previous trials, of affix repetition, trial number and the lexicality of the previous item. These variables account for variance that would otherwise be noise in the data.

Together, the results indicate that morphological structure is functional to lexical processing. They also show a flexible and interactive processing system, in which the conditional processing of one constituent given another plays a major role. The results suggest that an interactive dual-route model is the most plausible model of morphological processing.

Dansk resumé

De eksperimenter der beskrives i denne afhandling, er motiveret af en fascination af ordgenkendelsens hurtighed og præcision og belyser disse to aspekter ved at undersøge genkendelsen af morfologisk komplekse ord. Eksperimenterne fokuserer på lyttegenkendelse, fordi den auditive modalitet er sprogets primære, og fordi et hovedsigte var at undersøge ændringerne i den leksikalske konkurrence over tid som er karakteristisk for auditiv ordgenkendelse. Det undersøgte sprog er dansk; dansk er et "nyt" sprog indenfor feltet og har forskellige karakteristika der er relevante for studiet af morfologisk processering.

Kapitel 1 beskriver de modeller der er blevet formuleret for processering af morfologisk komplekse ord. En central kontrast er mellem lagring af hele komplekse former og processering baseret på morfemer. Mange modeller falder på et kontinuum fra ren morfemprocessering til ren lagring af hele ord. Mellem disse to yderpositioner er der forskellige hybridmodeller som opererer med en blanding af helords- og morfemprocessering. Blandt hybridmodellerne er dobbeltrute-modeller som omfatter både en helordsbaseret og en morfembaseret genkendelsesrute for komplekse ord; disse ruter kan operere efter hinanden eller samtidig i konkurrence eller interaktion. En anden forskningstradition, der delvist overlapper med det nævnte kontinuum, fokuserer på processeringen af regelmæssige og uregelmæssige bøjningsformer som en måde at teste vidtrækkende påstande om sprog på. Dobbeltmekanisme-modeller antager at regelmæssige og uregelmæssige bøjningsformer processeres via to forskellige mentale mekanismer: regelmæssige former processeres ved hjælp af en symbolsk regel i den procedurale hukommelse, ligesom syntaktisk struktur. Uregelmæssige former lagres i den deklarative hukommelse, lige som simple ord. I modsætning hertil argumenterer enkeltmekanisme-modeller for at både regelmæssige og uregelmæssige former processeres associativt i den samme del af hukommelsen. Graduerede morfologiske effekter ses af fortalere for enkeltmekanisme-modeller som bevis for at morfologisk struktur i ordhukommelsen er et epifænomen der skyldes overlap i semantik og udtale/stavemåde mellem morfologisk relaterede ord, men der argumenteres i sektion 1.3 for at sådanne graduerede effekter også kan opstå som et resultat af paradigmatiske forbindelser mellem morfologisk relaterede helordseksemplarer i hukommelsen. Den informationsteoretiske PROMISE-model er en mulig model for gradueret og sandsynlighedsbaseret morfologisk processering. Videre argumenteres det i sektion 1.4 at morfologien spiller en gavnlig rolle for ordgenkendelsen, fordi genkendelsen af komplekse ord kan baseres både på hele ord og på morfemer, hvilket skulle gøre processen mere effektiv og robust.

Kapitel 2 beskriver centrale effekter i den eksperimentelle litteratur. Effekter af morfologiske relationer mellem ord sammenlignes ofte med semantiske og formelle ligheder. Nabolagseffekter viser at et ord der skal genkendes (et målord), generelt bliver genkendt langsommere hvis der er mange andre ord hvis form overlapper med målordets. I auditiv ordgenkendelse er det de ord hvis begyndelse overlapper med målordets, såkaldte kohorter, der er vigtigst. Kohorter består af ord der overlapper med målordet op til et vist punkt. Det fonem i målordet hvor kohorten er reduceret til kun at indeholde målordet, kaldes UP ('Uniqueness Point'). Jo senere dette UP falder, des længere tager det at genkende ordet. Det traditionelle UP tager ikke højde for morfologisk struktur. Andre kohortebaserede variabler der indregner morfologisk kompleksitet, er beskrevet i sektion 2.3.2, bl.a. 'Conditional Root Uniqueness Point' (CRUP). CRUP-effekter viser at forståelsen af præfigerede ords stammer afhænger af det præfiks der er blevet genkendt. CRUP er kun relevant for et fåtal af præfigerede ord, og derfor introduceres to nye UP'er som gør det muligt at undersøge den samme slags afhængigheder mellem de forskellige morfemer i genkendelsen af alle typer komplekse ord. Effekter af formel lighed kan også studeres ved hjælp af priming. Morfologiske priming-effekter er som regel større end formelle og semantiske priming-effekter, og de udvikler sig anderledes over tid, men graden af semantisk relation mellem ordpar i priming påvirker de morfologiske effekter, hvilket tyder på at morfologiens indflydelse på ordgenkendelsen ikke er uafhængig af semantik.

Sammenligningen af helords- og stammefrekvenseffekter er central i studiet af morfologisk processering. Helordsfrekvenseffekter ses almindeligvis som tegn på genkendelse på basis af hele ord, mens effekter af stammefrekvens ses som tegn på morfembaseret genkendelse. I sektion 2.4 argumenteres det derimod at stammefrekvens ikke altid er den bedste markør for morfembaseret genkendelse. Et alternativ er affiksfrekvens. Desuden kan helordsfrekvens forstås som morfologisk, fordi den udgør sandsynligheden for at morfemerne i et komplekst ord optræder sammen.

Et ords morfologiske familie består af de afledte og sammensatte ord der indeholder ordets stamme. Jo flere familiemedlemmer et ord har, des hurtigere genkendes det. Familien er ikke et morfologisk paradigme i traditionel forstand, men familieeffekterne kan siges at afspejle de samme paradigmatiske relationer i det mentale leksikon som effekter af bøjningsparadigmer.

Sektion 2.6 beskriver forskelle mellem typer af ord og morfologiske processer: mellem simple og komplekse ord, mellem bøjning, afledning og sammensætning, mellem præfigerede og suffigerede ord, og mellem regelmæssig og uregelmæssig bøjning. De tre første kontraster er relevante for de eksperimenter om genkendelse af danske ord der beskrives i kapitel 5 til 7.

Kapitel 3 beskriver relevante aspekter af det danske sprog i sektion 3.2. Det danske tekstkorpus Korpus90/2000 beskrives i sektion 3.3, som også gør rede for de betydelige modifikationer af dette korpus der var nødvendige. Udtrækningen fra Korpus90/2000 af forskellige frekvenser, morfologiske familier, UP'er og nabolag beskrives i sektion 3.4.

Kapitel 4 forklarer først brugen af leksikalsk beslutning i eksperimenterne. Deltagernes opgave var at afgøre om de genkendte en række ord og vrøvleord som ord på dansk eller ej. I sektion 4.3 beskrives fordelene ved at bruge regressionsdesigns i stedet for de faktorielle designs der tradionelt bruges indenfor feltet: Regressionsdesigns og -analyser giver mere information og er statistisk stærkere. Sektion 4.4 forklarer de lineære "mixed-effects" regressionsmodeller der blev brugt til de statistiske analyser.

I kapitel 5 til 7 beskrives tre eksperimenter om dansk ordgenkendelse. Eksperiment 1 havde både en auditiv (1a) og en visuel version (1b); de øvrige eksperimenter var auditive.

Et gennemgående resultat af alle eksperimenterne var at alle typer morfologisk komplekse ord (bøjninger, afledninger og sammensætninger) blev genkendt hurtigere end simple ord, når der i analyserne blev taget højde for andre relevante variabler. Det argumenteres at de komplekse ords fordel skyldes at der kan gøres brug af både helords- og morfeminformation i genkendelsen. Fordelen var størst for suffigerede ord, hvor det mest informative morfem, nemlig stammen, optræder først. Eksperiment 2 viste ingen kategorisk forskel på genkendelsen af regelmæssigt bøjede ord og mindre regelmæssige afledte ord, hvilket er problematisk for dobbeltmekanisme-modellen.

Alle tre eksperimenter undersøgte effekter af forskellige frekvenser. I Eksperiment 1a var der signifikant stærkere effekter af helordsfrekvens for komplekse end for simple ord hvilket tyder på at helordsfrekvensen både har en ikke-morfologisk og en kombinatorisk komponent for komplekse ord. I Eksperiment 1a og 2 var der en interaktion mellem helordsfrekvens og frekvensen af de komplekse ords andet morfem. Når den ene frekvens var lav, havde den anden en større effekt på genkendelseshastigheden. Derimod tog ord med høj helords- og høj morfemfrekvens relativt lang tid at genkende. Det lader altså til at der opstår konkurrence mellem forskellige mentale repræsentationer når de er meget aktive; mulige grunde til det diskuteres i kapitel 6 og 8. Interaktionerne viser at forskellige typer information ikke anvendes uafhængigt af hinanden, hvilket betyder at fordelen for de komplekse ord formentlig også skyldes interaktion mellem genkendelsesruter, snarere end såkaldt statistisk facilitering. At det var frekvensen af det andet morfem der interagerede med helordsfrekvensen, antyder at processeringen af det første morfem i nogen grad afhænger af processeringen af det andet i auditiv ordgenkendelse. I Eksperiment 2 og 3 var effekterne af helordsfrekvens signifikante for begge køn, men stærkere for kvinder end for mænd. Dette er formentlig en konsekvens af kvinders bedre sproglige hukommelse og passer med dobbeltmekanisme-modellen, men i Eksperiment 2 viste kvinderne også stærkere effekter af affiksfrekvens og en mere udtalt interaktion mellem helords- og affiksfrekvens, hvilket er problematisk for dobbeltmekanisme-modellen. Endelig var der i det visuelle Eksperiment 1b en effekt af frekvensen af det bogstavpar der optrådte hen over grænsen mellem de komplekse ords to morfemer; denne effekt viser at ord med mere tilgængelige morfemer genkendes hurtigere.

Et hovedformål med alle tre eksperimenter var at undersøge effekterne af de to nye UP'er der blev introduceret i kapitel 2. I komplekse ord er UP1 det fonem hvor konkurrence fra urelaterede ord ophører, mens det komplekse UP (CUP) er det fonem hvor ord der starter med målordets første morfem, ikke længere findes i kohorten. Begge variabler havde stærke og signifikante effekter i alle de auditive eksperimenter: jo senere UP1 og CUP falder i et ord, des længere tager det at genkende ordet. Selv i det visuelle Eksperiment 1b var der en signifikant effekt af UP1, selvom variablen grundlæggende er auditiv. Effekterne var de samme for alle typer komplekse ord. At begge effekter var signifikante viser at både urelaterede og relaterede ord spiller en rolle i den leksikalske konkurrence uafhængigt af hinanden. CUP-effekten viser betinget processering: konkurrencen mellem de relaterede ord er betinget af at det første morfem er blevet genkendt.

De forskellige eksperimenter viste forskellige effekter af morfologisk familie: I de auditive eksperimenter der omfattede præfigerede ord (1a og 3), var der kun gavnlige effekter af fortsættelsesformer, de familiemedlemmer som indeholder hele målordet, og hvis begyndelse overlapper med målordets. I Eksperiment 1a var der endda en hæmmende effekt af standardfamilien. Derimod viste det visuelle Eksperiment 1b og de suffigerede ord i Eksperiment 2 gavnlige effekter af standardfamilien, i det mindste op til en vis familiestørrelse. Når en stamme findes tidligt i ordet, har den stammebaserede morfologiske familie en gavnlig effekt på genkendelsen. Den forskellige kohortedynamik for de forskellige ordtyper påvirker altså hvilke familievariabler der er gavnlige, og hvilke der virker hæmmende på genkendelsen. Eksperiment 3 undersøgte om tidsintervallet mellem præsentationen af ordene påvirkede familieeffekterne; dette var ikke tilfældet.

Endelig viste alle analyser effekter af eksperimentel kontekst. Der var signifikante effekter af reaktionstiden og korrektheden for de foregående ord og vrøvleord, af hvor mange gange et affiks var blevet præsenteret, af hvornår i eksperimentet et ord optrådte, og af om det foregående var et ord eller et vrøvleord. Regressionsanalysens inddragelse af disse variabler betyder at varians som ellers ville være støj i analysen, bliver kontrolleret.

Tilsammen viser resultaterne at morfologien er gavnlig for ordgenkendelsen. Resultaterne viser også et fleksibelt og interaktivt processeringssystem hvor processering af hvert morfem er betinget af de andre morfemer det optræder med. Resultaterne tyder på at en interaktiv dobbeltrute-model er den bedste model for morfologisk processering.

Appendix A

Items in Experiments 1a & 1b

Table A.1: Prefixed words of Experiments 1a and 1b and associated lexical variables. The columns are defined as follows: Under 'Frequency', 'Word' is the base form frequency of the whole word; 'Stem' is the lemma frequency of the stem as an independent word; 'Affix' is the token frequency of the affix; 'Fam.' is the type count of morphological family members; 'Cont.' is the type count of morphologically related continuation forms; 'Juncture' is the frequency of the bigram crossing the juncture between the constituent morphemes. All these measures are extracted from *Korpus90/2000*, the values are for the 43.6 million tokens in that corpus. 'UP1' and 'CUP' are UP-measures in ms for the auditory stimulus, 'Dur.' the duration of the same. The same variables are listed under the same column names for the remaining words in Experiments 1a and 1b in the following tables.

		Frequency					ms			
Word	Class	Word	Stem	Affix	Fam.	Cont.	UP1	CUP	Dur.	Juncture
be-koste	V	31	9 501	162 009	258	9	149	375	590	372 745
be-laste	V	148	243	162 009	334	7	86	340	566	1 872 068
be-nægte	V	151	3 119	162 009	68	11	93	361	591	6 207 423
be-slutte	V	598	6 687	162 009	772	11	90	327	552	1 389 144
be-snakke	V	6	6 966	162 009	261	3	75	337	578	1 389 144
gen-digte	V	3	481	23 287	296	4	277	447	641	2 597 157
gen-kende	V	568	19 480	23 287	620	29	269	375	588	237 708
gen-klang	N	106	375	23 287	167	3	230	378	504	237 708
gen-lære	V	0	12 328	23 287	1 060	1	242	335	644	83 338
gen-spejle	V	5	473	23 287	181	7	173	327	654	1 060 206
mis-kredit	N	42	254	8 155	276	7	283	321	688	1 670 176
mis-lyd	N	16	4 817	8 155	678	4	269	409	508	285 419
mis-mod	N	70	1 108	8 155	18	8	271	386	614	240 566
mis-tolke	V	0	874	8 155	225	4	300	423	722	2 364 484
mis-unde	V	83	264	8 155	26	30	320	384	603	136 244
sam-drift	N	18	2 022	34 391	692	11	429	545	799	5 900
sam-ordne	V	44	1 471	34 391	995	6	405	633	818	324 890
sam-råd	N	520	7 075	34 391	1 126	16	360	559	687	53 354
sam-sende	V	2	16 866	34 391	463	2	354	542	781	108 709
sam-tænke	V	3	23 007	34 391	761	1	365	513	790	148 927
u-fiks	A	1	306	107 527	1	3	125	398	619	47 044
u-klar	A	216	16 910	107 527	312	9	177	344	486	109 961
u-lykke	N	904	2 213	107 527	354	114	158	402	539	360 262
u-stabil	A	111	1 258	107 527	119	7	117	398	632	325 660
u-vejr	N	143	3 643	107 527	284	14	191	406	527	18 907
und-gå	V	4 369	125 016	12 222	190	11	260	335	472	76 027
und-skylde	V	316	851	12 222	141	8	198	327	593	675 237
und-slippe	V	143	5 066	12 222	65	2	184	386	614	675 237
und-være	V	901	1 413 049	12 222	477	4	232	426	674	132 119
und-vige	V	72	681	12 222	56	22	227	398	606	132 119
van-ære	N	35	1 859	949	990	9	275	355	647	128 921
van-held	N	1	2 221	949	110	1	283	595	595	78 596
van-hellig	N	4	1 589	949	86	8	235	512	600	78 596
van-røgte	V	1	27	949	13	6	229	471	658	34 128
van-skæbne	N	10	2 325	949	110	1	216	579	725	1 060 206

Table A.2: Particle prefixed words of Experiments 1a and 1b and associated lexical variables. The column names are the same as in the previous table.

		Frequency					ms			
Word	Class	Word	Stem	Affix	Fam.	Cont.	UP1	CUP	Dur.	Juncture
af-hente	V	43	7 715	91 290	41	6	242	501	620	15 166
af-magt	N	326	6 136	91 290	524	19	183	585	612	3 128
af-savn	N	83	332	91 290	5	5	193	451	574	49 004
af-skaffe	V	551	4 862	91 290	63	10	153	368	608	49 004
af-sløre	V	942	351	91 290	72	8	197	443	663	49 004
bag-vende	V	0	14 004	13 825	292	1	270	350	576	28 840
bag-klog	A	36	3 003	13 825	39	6	180	375	527	7 699
bag-sæde	N	53	1 222	13 825	142	29	197	475	644	432 252
bag-side	N	192	22 311	13 825	1 129	41	215	425	626	432 252
bag-vagt	N	6	1 263	13 825	247	2	237	420	602	28 840
efter-gilde	N	0	154	27 305	54	2	248	492	734	261 088
efter-løn	N	655	3 694	27 305	894	91	206	418	562	222 510
efter-mæle	N	93	81	27 305	8	2	231	460	699	243 634
efter-smag	N	67	2 465	27 305	320	3	227	518	667	686 866
efter-skælv	N	20	23	27 305	29	1	205	486	700	686 866
med-bringe	V	183	9 744	41 481	150	8	194	394	649	84 968
med-fange	N	13	1 580	41 481	381	6	202	536	660	59 066
med-vind	N	157	2 847	41 481	283	7	235	403	598	132 119
med-virke	V	922	11 154	41 481	1 507	11	184	486	663	132 119
med-ynk	N	32	34	41 481	12	3	282	377	593	61 965
om-egn	N	301	698	79 498	163	33	161	347	591	2 201 994
om-favne	V	79	156	79 498	22	12	129	460	662	55 043
om-rids	N	75	37	79 498	16	4	180	503	582	53 354
om-serv	N	1	254	79 498	45	1	139	360	571	108 709
om-verden	N	259	25 218	79 498	1 085	33	175	371	630	12 099
op-fatte	V	669	1 445	129 876	622	20	102	385	591	31 913
op-finde	V	169	46 652	129 876	134	51	98	470	633	31 913
op-nå	V	2 432	20 978	129 876	5	14	98	359	458	8 735
op-sving	N	351	677	129 876	271	7	114	372	492	59 516
op-vask	N	108	223	129 876	339	45	105	326	597	5 344
over-moden	A	7	988	75 218	95	2	250	557	719	243 634
over-skue	V	559	357	75 218	460	16	217	575	630	686 866
over-skygge	V	68	400	75 218	174	7	273	621	744	686 866
over-tale	V	438	21 163	75 218	1 417	27	247	677	722	627 528
over-tone	N	7	2 424	75 218	324	10	257	490	698	627 528
til-kalde	V	157	18 164	120 225	233	19	309	590	644	145 361
til-navn	N	58	13 575	120 225	480	5	257	387	472	6 962
til-passe	V	574	10 105	120 225	236	8	320	529	647	51 774
til-snit	N	87	690	120 225	488	1	229	510	613	586 982
til-trække	V	590	16 210	120 225	546	6	269	553	682	425 663

Table A.3: Suffixed words of Experiments 1a and 1b and associated lexical variables. The column names are the same as in the previous tables.

		Frequency					ms			
Word	Class	Word	Stem	Affix	Fam.	Cont.	UP1	CUP	Dur.	Juncture
fabel-agtig	A	73	75	6 320	29	3	347	542	860	889 241
fejl-agtig	A	77	4 919	6 320	442	4	197	396	652	889 241
fløjls-agtig	A	4	68	6 320	70	3	394	607	940	631 291
løgn-agtig	A	29	1 134	6 320	40	5	228	393	726	350 152
tumult-agtig	A	5	165	6 320	13	2	286	533	904	775 963
brug-bar	A	158	37 703	8 653	2 103	6	300	300	578	15 064
bær-bar	A	70	7 187	8 653	292	5	224	266	520	255 494
læs-bar	A	4	13 149	8 653	726	5	312	421	678	69 638
mærk-bar	A	177	8 112	8 653	997	3	267	404	651	4 528
print-bar	A	0	134	8 653	88	1	291	500	620	17 275
accept-ere	V	2 376	642	260 308	24	10	321	609	782	2 766 352
billett-ere	V	3	2 155	260 308	246	1	206	464	720	2 766 352
march-ere	V	52	242	260 308	115	6	184	509	583	785 484
process-ere	V	1	3 640	260 308	687	1	395	550	801	1 681 882
respekt-ere	V	509	2 927	260 308	27	9	302	651	818	2 766 352
bygg-eri	N	860	10 289	19 965	1 731	10	155	382	484	2 809 478
fjoll-eri	N	2	99	19 965	28	7	140	347	535	2 311 504
gætt-eri	N	25	1 058	19 965	29	5	334	334	557	2 766 352
ras-eri	N	366	984	19 965	19	17	321	427	515	1 681 882
tyv-eri	N	428	545	19 965	197	49	342	349	583	1 806 919
frisk-hed	N	114	4 078	193 618	118	3	453	548	769	8 129
gerrig-hed	N	6	8	193 618	16	2	402	402	623	150 881
korrekt-hed	N	118	2 345	193 618	43	5	248	425	687	95 815
tæt-hed	N	64	11 196	193 618	119	6	219	232	470	95 815
tavs-hed	N	838	1 039	193 618	15	21	297	346	529	64 789
asket-isk	A	31	15	222 805	3	1	313	732	785	2 208 593
film-isk	A	83	13 909	222 805	1 297	2	366	544	679	625 569
gigant-isk	A	420	300	222 805	285	2	383	638	777	2 208 593
idyll-isk	A	105	359	222 805	90	4	291	658	724	1 691 894
kult-isk	A	8	96	222 805	124	2	296	499	656	2 208 593
arve-lig	A	97	621	664 759	165	23	277	486	559	1 872 068
kede-lig	A	630	621	664 759	28	9	373	496	559	1 872 068
natur-lig	A	1 458	6 570	664 759	724	20	463	706	706	222 510
pynte-lig	A	8	902	664 759	121	3	332	488	557	1 872 068
sommer-lig	A	10	5 910	664 759	864	4	431	675	675	222 510
miljø-mæssig	A	93	5 723	13 165	1 200	3	306	495	803	51 682
motiv-mæssig	A	1	2 186	13 165	244	1	348	582	851	11 190
regel-mæssig	A	139	9 917	13 165	640	6	171	513	688	80 144
rutine-mæssig	A	33	631	13 165	87	4	311	646	912	761 522
vækst-mæssig	A	1	3 816	13 165	411	3	332	487	768	16 730
arrig-skab	N	39	170	43 076	3	2	254	300	593	432 252
doven-skab	N	84	414	43 076	23	3	175	349	708	1 670 176
mester-skab	N	631	2 264	43 076	989	95	255	587	680	686 866
svanger-skab	N	77	37	43 076	26	25	394	480	745	686 866
ven-skab	N	541	9 591	43 076	547	111	319	319	581	1 670 176
føl-som	A	292	16 703	25 152	493	15	291	352	632	586 982
glem-som	A	26	6 631	25 152	32	6	190	400	565	108 709
gru-som	A	142	66	25 152	12	14	266	437	567	325 660
skån-som	A	57	348	25 152	38	6	367	566	709	1 060 206
stræb-som	A	10	406	25 152	99	5	302	498	745	35 743

Table A.4: Simple words of Experiments 1a and 1b and associated lexical variables. The column names are the same as in the previous tables.

		Frequency					ms			
Word	Class	Word	Stem	Affix	Fam.	Cont.	UP1	CUP	Dur.	Juncture
albue	N	105	–	–	24	34	382	382	581	–
bitter	A	466	–	–	25	28	255	255	329	–
bjælke	N	51	–	–	51	27	301	301	454	–
bleg	A	455	–	–	85	92	266	266	294	–
blød	A	1 048	–	–	87	131	239	239	239	–
blomst	N	408	–	–	531	492	283	283	431	–
bonus	N	366	–	–	92	113	300	300	586	–
brække	V	199	–	–	69	9	333	333	383	–
brat	A	550	–	–	1	36	283	283	371	–
briste	V	101	–	–	14	12	385	385	438	–
brøk	N	19	–	–	11	15	286	286	382	–
civil	A	254	–	–	138	233	370	370	548	–
dæmpe	V	380	–	–	67	15	427	427	427	–
datter	N	3 337	–	–	141	64	238	238	311	–
diskret	A	605	–	–	9	10	469	469	516	–
dølge	V	14	–	–	2	5	328	328	487	–
domæne	N	128	–	–	32	25	233	233	594	–
droppe	V	447	–	–	23	5	349	349	398	–
drysse	V	46	–	–	33	10	415	415	481	–
dufte	V	68	–	–	165	21	447	447	447	–
dyrke	V	1 016	–	–	286	16	310	310	441	–
ekstra	A	6 431	–	–	205	363	384	384	462	–
elefant	N	135	–	–	100	125	389	389	625	–
elite	N	368	–	–	288	353	432	432	546	–
emne	N	2 016	–	–	195	98	282	282	441	–
fejre	V	801	–	–	6	10	484	484	581	–
fersk	A	174	–	–	59	125	541	541	541	–
fjende	N	485	–	–	169	37	287	287	437	–
flette	V	54	–	–	86	15	391	391	444	–
fordre	V	28	–	–	111	21	430	430	582	–
frakke	N	361	–	–	83	36	424	424	495	–
fremmed	A	1 443	–	–	142	204	429	429	473	–
from	A	178	–	–	11	30	489	489	489	–
frygt	N	2 068	–	–	54	51	473	473	581	–
galakse	N	54	–	–	16	16	411	411	591	–
gardin	N	66	–	–	72	56	394	394	446	–
glad	A	5 453	–	–	340	82	347	347	347	–
god	A	20 104	–	–	356	517	329	329	329	–
hæmme	V	121	–	–	65	10	297	297	337	–
haste	V	46	–	–	312	98	394	394	443	–
hjørne	N	1 259	–	–	148	110	506	506	541	–
human	A	202	–	–	83	99	380	380	556	–
huske	V	3 877	–	–	37	52	417	417	469	–
kaffe	N	2 451	–	–	259	335	392	392	392	–
kaptajn	N	783	–	–	65	28	243	243	638	–
kartoffel	N	184	–	–	219	253	411	411	574	–
kigge	V	2 118	–	–	85	20	356	356	414	–
klukke	V	7	–	–	9	8	375	375	441	–
kold	A	1 411	–	–	136	217	261	261	261	–
krog	N	311	–	–	110	56	329	329	384	–
kujon	N	23	–	–	4	15	237	237	615	–
lampe	N	251	–	–	171	71	370	370	506	–
lokal	A	1 016	–	–	583	673	291	291	465	–
lukke	V	2 263	–	–	186	94	396	396	448	–
lyng	N	101	–	–	38	216	252	252	252	–
mæt	A	243	–	–	111	37	304	304	304	–
maskine	N	732	–	–	972	22	401	401	670	–
menneske	N	5 723	–	–	762	912	216	216	579	–
mild	A	453	–	–	27	29	394	394	394	–
mølle	N	158	–	–	237	201	458	458	458	–
mørk	A	888	–	–	210	200	479	479	479	–
morgen	N	7 511	–	–	430	587	288	288	504	–
mund	N	2 083	–	–	223	229	280	280	280	–
munter	A	264	–	–	10	10	335	335	398	–
naiv	A	309	–	–	22	22	229	229	385	–
nerve	N	170	–	–	153	198	463	463	603	–
nikke	V	234	–	–	8	53	427	427	427	–

Table A.4: Simple words of Experiments 1a and 1b and associated lexical variables. (continued)

		Frequency					ms			
Word	Class	Word	Stem	Affix	Fam.	Cont.	UP1	CUP	Dur.	Juncture
papir	N	1 923	–	–	611	494	375	375	413	–
penge	N	11 929	–	–	714	599	343	343	411	–
plads	N	13 394	–	–	861	119	456	456	456	–
præge	V	500	–	–	345	10	448	448	599	–
problem	N	7 225	–	–	1 175	208	255	255	543	–
puste	V	284	–	–	39	20	432	432	488	–
rasle	V	77	–	–	30	27	323	323	438	–
robust	A	180	–	–	3	11	247	247	538	–
ryste	V	579	–	–	52	43	359	359	420	–
saga	N	177	–	–	48	42	508	508	576	–
sekund	N	1 295	–	–	25	98	671	671	671	–
sikker	A	5 053	–	–	1 401	742	462	462	542	–
skærpe	V	279	–	–	73	12	502	502	632	–
skjule	V	1 176	–	–	40	13	352	352	585	–
skrabe	V	181	–	–	75	44	456	456	590	–
smæld	N	135	–	–	14	14	379	379	418	–
smal	A	1 682	–	–	34	70	306	306	432	–
smile	V	439	–	–	96	39	553	553	605	–
sofa	N	333	–	–	120	124	515	515	588	–
solid	A	629	–	–	110	115	456	456	478	–
spise	V	3 700	–	–	328	287	476	476	634	–
splejse	V	48	–	–	100	9	658	658	717	–
stærk	A	3 477	–	–	564	61	486	486	547	–
starte	V	2 397	–	–	345	26	613	613	672	–
stolt	A	1 398	–	–	48	27	622	622	622	–
straffe	V	338	–	–	252	219	452	452	600	–
stål	N	487	–	–	344	472	493	493	533	–
svække	V	226	–	–	39	11	386	386	523	–
svømme	V	459	–	–	203	193	320	320	586	–
symbol	N	675	–	–	205	133	494	494	693	–
sysle	V	7	–	–	18	6	431	431	566	–
tænde	V	413	–	–	117	19	294	294	450	–
tam	A	94	–	–	27	241	256	256	256	–
tomat	N	180	–	–	101	127	485	485	532	–
træt	A	2 233	–	–	103	83	382	382	382	–
trist	A	995	–	–	23	27	429	429	488	–
tvinge	V	846	–	–	43	6	322	322	466	–
tvivle	V	211	–	–	39	15	306	306	475	–
ussel	A	101	–	–	5	6	266	266	376	–
vræle	V	18	–	–	6	7	362	362	496	–
vride	V	175	–	–	57	11	351	351	539	–
ytre	V	110	–	–	24	5	270	270	437	–
ændre	V	3 592	–	–	194	7	291	291	450	–

Table A.5: Nonwords of Experiments 1a and 1b.

afløre	efterglag	hiljemæssig	lalmbar	moksvisp	pjandelmern	sofe	tærpshed
afmarf	efterknilde	homt	lamfe	movlen	pjultisk	solf	tærrigskab
afsalf	efterkrøn	humen	lartre	mufte	plaft	splense	tævse
afsmaffe	efterpræle	ilbe	latajn	mulfer	plist	sprise	ubjar
afstænte	efterskæs	illeptere	lekstre	mulkagtig	plivebolf	stramfe	ufroks
ajlsagtig	fegn	issel	likre	mulp	plyrke	stravle	uglakke
apselig	fevlagtig	jersk	likse	mutenemæssig	prable	svemmehås	undplippe
baglæde	fjærk	jugtrilge	lintbar	mække	prolletere	svukke	undskyvse
bagtvide	flatter	jurferke	lutse	mæltbar	prælge	svusom	undsvå
bagvajf	flibelagtig	juske	lylg	mæmbefølf	prølsom	svække	undvimle
bagvekse	folme	jækkehuft	lymsom	mæne	putse	svættegøve	undvure
bagvråg	frample	jængerskab	maber	mølke	rakshed	svød	ustibel
behuste	frybt	kalse	marlere	nasterskab	rektere	symbel	uvoj
belavse	garitshed	kegnelig	martiffel	netelmæssig	robelm	talmhed	vanfreld
beløtte	genbænne	kilmisk	masle	nysle	rolte	tarden	vanhølje
benævle	genfjang	kirfe	medbrilke	omkvegn	rylte	terve	vankubne
berfbar	genlælfe	kloppe	meddange	omplerv	rølke	tilnarp	vanrøtse
bevlakke	genmigte	klærpe	medslind	omstids	rålgnast	tilnække	vanærle
bindre	genpevle	kløgnagtig	medtvarke	omtravne	samdrit	tilpalge	viskre
birfe	gilperi	klørk	medyrf	omvurden	samrång	tilpilte	vrende
bleps	glage	knile	melenent	opfilge	samsejde	tilstralle	vrine
blimst	glank	knæt	mespese	opklå	samskardne	timmerlig	vræte
brart	glette	korfhed	mibantisk	opknatte	samtæpse	tjyste	vrånsom
briske	glivenskab	kremne	mikker	opspingst	sivel	tokit	vuktmæssig
brømme	glundbemple	kribse	milp	opvamf	skaseri	tribsom	vulbar
byrberi	gnoholje	krivelåbe	miskrevel	overfygge	skjælf	trærp	vyndskab
dollisk	gnøk	krold	mismovn	overkniden	skotisk	trål	ylke
drynke	govn	krong	misplyd	overskage	skrunt	tvam	
dvintelig	grones	krårtonge	mistolje	oversvone	smang	tvimle	
dvolleri	halletere	kujn	misurpse	overtapse	smastne	tvorslats	
dvørne	halpe	kvorfrip	mofitmæssig	pejfe	smægl	tyngsteri	
dæmfe	hanorlig	kvæmme	mokel	pjakke	sniv	tælke	

Table A.6: Filler words and training and warm-up items of Experiments 1a and 1b.

Filler Words		Training words			Training nonwords		
ajourføre	grundlægge	beskrive	inddrage	rulle	bauskal	holk	prølm
blåstemple	håbefuld	bibliotek	lunken	sejle	behyrpe	javn	rynt
boghandel	kongerige	ejendom	måde	skrap	bleng	kirmning	skramst
bosætte	kæmpestor	eventyr	pause	streng	erslimpe	klersel	smæp
fastholde	rivejern	fjernsyn	pragtfuld	styrte	fjægelse	kving	stryps
feriehus	skrivebord	fremstille	rejse	vigtig	gjokst	murst	trebs
frigøre	stemmeboks	gentage	ring		glår	narfdom	
frugtskål		hus	rose		henplaute	pjavl	

Appendix B

Items in Experiment 2

Table B.1: Derived words of Experiment 2 and associated lexical variables. The columns are defined as follows: Under 'Frequency', 'Word' is the base form frequency of the whole word; 'Stem' is the lemma frequency of the stem; 'Affix' is the token frequency of the affix; 'Fam.' is the type count of morphological family members; 'Cont.' is the type count of morphologically related continuation forms; 'Juncture' is the frequency of the bigram crossing the juncture between the constituent morphemes. All these measures are extracted from *Korpus90/2000*, the values are for the 43.6 million tokens in that corpus. 'Transp' is the mean of the semantic transparency ratings obtained in a rating task. 'UP1' and 'CUP' are UP-measures in ms for the auditory stimulus, 'Dur.' the duration of the same. The same variables are listed under the same column names for the remaining words in Experiment 2 in the following tables.

		Frequency						ms			
Word	Class	Word	Stem	Affix	Fam.	Cont.	Transp.	UP1	CUP	Dur.	Juncture
flyt-bar	A	7	9 073	8 653	200	2	4.36	241	340	616	17 275
frugt-bar	A	173	1 798	8 653	318	37	3.36	255	635	635	17 275
gang-bar	A	29	48 261	8 653	1 524	2	2.03	205	359	517	15 064
husk-bar	A	1	8 604	8 653	35	2	3.86	309	394	619	4 528
klik-bar	A	0	340	8 653	22	1	4.33	175	415	576	4 528
kost-bar	A	263	9 501	8 653	265	11	3.33	269	400	548	17 275
sang-bar	A	8	4 153	8 653	675	4	3.75	365	475	637	15 064
smør-bar	A	16	1 206	8 653	222	3	4.44	340	529	719	255 494
strid-bar	A	18	1 948	8 653	194	3	3.89	433	522	789	84 968
åben-bar	A	93	8 250	8 653	318	25	2.75	306	398	640	33 758
bas-ere	V	100	447	260 308	671	5	3.17	240	318	583	1 681 882
blok-ere	V	308	627	260 308	306	7	2.78	198	356	607	1 887 007
hus-ere	V	14	15 451	260 308	1 982	6	2.14	297	362	628	1 681 882
motiv-ere	V	197	2 186	260 308	240	6	3.19	291	490	754	1 806 919
not-ere	V	397	135	260 308	204	6	5.19	212	381	631	2 766 352
orient-ere	V	440	124	260 308	406	8	1.25	220	527	820	2 766 352
park-ere	V	159	1 686	260 308	332	4	1.86	296	366	630	1 887 007
punkt-ere	V	26	4 371	260 308	675	4	1.81	249	392	660	2 766 352
ruin-ere	V	24	645	260 308	37	6	2.97	205	449	689	2 129 275
ventil-ere	V	3	234	260 308	131	4	4.36	331	518	775	2 311 504
brygg-eri	N	147	200	19 965	125	87	5.28	223	477	558	2 809 478
drill-eri	N	61	785	19 965	28	3	5.97	185	334	515	2 311 504
føl-eri	N	24	16 703	19 965	492	4	4.28	260	424	624	2 311 504
heks-eri	N	26	593	19 965	120	2	5.58	214	451	550	1 681 882
hyl-eri	N	3	517	19 965	58	1	5.58	284	405	587	2 311 504
kammerat-eri	N	22	2 235	19 965	132	1	3.86	365	503	763	2 766 352
scen-eri	N	48	5 204	19 965	616	4	4.36	301	550	638	2 129 275
smøl-eri	N	8	13	19 965	5	1	5.72	378	510	656	2 311 504
svin-eri	N	125	833	19 965	276	3	3.81	290	440	630	2 129 275
tøj-eri	N	10	5 469	19 965	1 178	2	1.44	222	308	568	729 888
grov-hed	N	44	2 652	193 618	97	3	5.08	206	381	534	3 096
høj-hed	N	34	29 275	193 618	1 327	4	2.69	198	373	515	1 027
klar-hed	N	451	20 516	193 618	293	6	4.39	335	469	658	157 002
kæk-hed	N	6	158	193 618	6	0	5.81	207	278	445	8 129
nær-hed	N	490	18 879	193 618	230	19	5.11	260	448	590	157 002
skæv-hed	N	73	1 715	193 618	49	3	5.25	336	485	629	3 096
stolt-hed	N	679	2 211	193 618	44	7	5.83	389	452	663	95 815
streng-hed	N	40	1 242	193 618	24	2	5.14	373	452	687	150 881
sur-hed	N	58	2 146	193 618	59	5	4.83	280	376	620	157 002
sær-hed	N	16	1 250	193 618	193	2	4.86	304	477	631	157 002
dyr-isk	A	50	9 746	222 805	859	2	4.00	222	312	523	1 142 263
dæmon-isk	A	46	275	222 805	23	1	5.64	253	426	635	720 906
iron-isk	A	508	412	222 805	12	1	6.11	266	392	723	720 906
jord-isk	A	135	10 929	222 805	625	1	3.64	216	444	617	544 746
lyr-isk	A	134	15	222 805	72	1	2.19	302	479	612	1 142 263
metall-isk	A	59	827	222 805	449	1	4.92	396	567	770	1 691 894
organ-isk	A	320	1 217	222 805	855	1	2.75	288	679	730	720 906
rytm-isk	A	359	986	222 805	169	1	5.22	269	407	569	625 569
takt-isk	A	258	1 988	222 805	482	1	1.64	274	443	602	2 208 593
typ-isk	A	3 126	4 569	222 805	824	1	3.25	230	304	434	201 745

Table B.1: Derived words of Experiment 2 and associated lexical variables. (continued)

		Frequency						ms			
Word	Class	Word	Stem	Affix	Fam.	Cont.	Transp.	UP1	CUP	Dur.	Juncture
ark-iv	A	212	418	60 131	224	110	1.92	272	368	418	172 180
distinkt-iv	A	0	24	60 131	3	1	4.00	264	400	616	2 208 593
effekt-iv	A	1 445	2 619	60 131	432	89	3.67	244	450	571	2 208 593
instinkt-iv	A	27	248	60 131	55	2	5.28	282	550	694	2 208 593
intens-iv	A	276	1 524	60 131	61	31	4.31	391	492	654	1 047 406
kurs-iv	A	10	5 069	60 131	762	6	1.33	303	406	454	1 047 406
mass-iv	A	688	8 822	60 131	491	8	3.06	300	330	431	1 047 406
produkt-iv	A	77	5 536	60 131	1 504	58	3.44	233	534	681	2 208 593
refleks-iv	A	7	245	60 131	111	3	2.72	267	583	641	1 047 406
subjekt-iv	A	70	214	60 131	21	12	3.47	401	682	809	2 208 593
appetit-lig	A	18	481	664 759	15	2	4.25	205	641	714	90 771
gemyt-lig	A	23	232	664 759	9	4	3.83	300	353	542	90 771
hæder-lig	A	96	212	664 759	39	7	3.97	276	446	614	222 510
jævn-lig	A	78	1 365	664 759	59	2	2.42	222	436	519	83 338
konge-lig	A	234	1 658	664 759	432	4	5.25	215	350	499	1 872 068
liv-lig	A	300	29 871	664 759	1 333	7	3.86	248	527	527	34 858
nem-lig	A	16 909	5 450	664 759	142	3	1.42	247	332	496	89 528
nyde-lig	A	167	4 411	664 759	61	4	2.72	452	482	655	1 872 068
tyde-lig	A	621	3 243	664 759	176	21	3.31	220	369	556	1 872 068
virke-lig	A	7 526	11	664 759	1 507	127	2.03	280	358	582	1 872 068
bøj-ning	N	40	2 223	246 624	161	16	4.56	220	261	459	22 583
fyld-ning	N	11	8 081	246 624	416	11	3.56	306	388	532	113 058
fæst-ning	N	88	163	246 624	136	52	2.58	414	477	732	123 199
hæld-ning	N	48	2 668	246 624	25	6	4.42	299	334	523	113 058
mørk-ning	N	3	4 465	246 624	200	4	4.19	304	416	550	112 188
nævn-ing	N	15	8 877	246 624	234	43	–	357	440	591	125 766
ord-ning	N	1 208	1 471	246 624	946	12	3.72	274	329	473	113 058
sæt-ning	N	574	41 113	246 624	1 799	49	1.83	377	401	583	123 199
tolk-ning	N	308	874	246 624	201	29	5.31	328	373	560	112 188
træf-ning	N	12	3 761	246 624	95	3	3.42	285	355	513	7 660
frygt-som	A	22	2 836	25 152	49	4	4.75	264	446	590	268 666
lang-som	A	451	47 270	25 152	428	28	1.58	253	453	621	432 252
næn-som	A	68	108	25 152	3	5	3.28	226	335	577	1 060 206
slid-som	A	14	386	25 152	92	2	4.50	326	551	705	675 237
smit-som	A	106	1 818	25 152	91	3	5.47	340	505	668	268 666
spar-som	A	80	4 344	25 152	412	11	4.44	393	454	711	686 866
tvivl-som	A	256	7 371	25 152	35	5	4.64	260	504	653	586 982
tænk-som	A	64	23	25 152	739	3	4.86	163	277	521	250 148
vagt-som	A	24	1 264	25 152	247	5	3.81	267	430	587	268 666
vold-som	A	1 564	3 531	25 152	326	12	4.36	191	440	491	675 237

Table B.2: Inflected words of Experiment 2 and associated lexical variables. The column names are the same as in the previous table.

		Frequency						ms			
Word	Class	Word	Stem	Affix	Fam.	Cont.	Transp.	UP1	CUP	Dur.	Juncture
bytt-ede	V	85	822	300 081	205	0	5.78	194	424	432	2 766 352
duft-ede	V	143	619	300 081	170	0	6.31	165	513	565	2 766 352
dyrk-ede	V	316	3 074	300 081	269	1	6.75	241	439	502	1 887 007
dyst-ede	V	25	214	300 081	42	1	6.44	265	458	515	2 766 352
lign-ede	V	1 846	8 319	300 081	777	0	6.31	347	586	643	2 129 275
lytt-ede	V	928	4 346	300 081	115	1	6.58	208	470	522	2 766 352
løft-ede	V	1 245	3 927	300 081	179	1	5.97	228	539	607	2 766 352
mangl-ede	V	1 317	8 823	300 081	128	0	6.42	315	559	626	2 311 504
spill-ede	V	4 090	23 184	300 081	2 165	1	6.50	388	656	730	2 311 504
vælt-ede	V	660	2 527	300 081	27	1	6.56	291	519	570	2 766 352
bukett-en	N	68	524	1 277 604	41	1	6.92	274	476	536	2 766 352
fugl-en	N	318	3 158	1 277 604	519	2	6.89	438	643	643	2 311 504
grupp-en	N	2 547	13 337	1 277 604	1 441	1	6.94	222	331	412	474 088
kamp-en	N	5 929	16 787	1 277 604	1 184	1	6.92	327	479	479	474 088
musikk-en	N	2 475	8 932	1 277 604	1 073	1	6.83	296	459	522	1 887 007
pejs-en	N	75	140	1 277 604	15	1	6.78	289	485	485	1 681 882
pligt-en	N	94	2 080	1 277 604	356	1	6.72	154	383	436	2 766 352
prins-en	N	413	1 601	1 277 604	107	1	6.94	397	471	551	1 681 882
saks-en	N	108	296	1 277 604	57	1	6.67	453	603	603	1 681 882
fløjt-ende	V	31	376	123 443	111	0	5.94	453	527	702	2 766 352
lov-ende	V	574	5 566	123 443	1 498	0	3.47	330	419	590	1 806 919
rull-ende	V	259	2 508	123 443	274	0	5.58	340	463	529	2 311 504
sejl-ende	V	72	2 936	123 443	454	1	6.14	420	528	714	2 311 504
skubb-ende	V	12	2 187	123 443	29	0	6.03	401	627	681	1 069 157
smil-ende	V	802	5 014	123 443	86	0	6.42	415	459	684	2 311 504
strejk-ende	V	183	413	123 443	101	1	5.44	401	640	698	1 887 007
vandr-ende	V	83	1 288	123 443	444	0	6.33	166	501	554	3 359 664
vræng-ende	V	25	119	123 443	24	0	6.19	220	430	635	2 809 478
baron-er	N	17	201	1 172 208	37	2	6.89	398	507	562	2 129 275
blomst-er	N	1 865	2 751	1 172 208	531	458	6.97	266	430	495	2 766 352
brikk-er	N	233	638	1 172 208	54	2	6.69	215	306	364	1 887 007
hval-er	N	157	399	1 172 208	107	3	6.89	327	434	504	2 311 504
legat-er	N	109	332	1 172 208	62	1	6.86	414	502	581	2 766 352
person-er	N	6 378	12 803	1 172 208	863	2	6.81	396	631	702	2 129 275
problem-er	N	13 540	28 419	1 172 208	1 140	3	6.92	260	575	659	2 201 994
reol-er	N	114	444	1 172 208	46	2	6.97	401	500	576	2 311 504
sekund-er	N	1 875	3 585	1 172 208	64	3	6.75	355	593	664	6 990 569
visitt-er	N	14	188	1 172 208	70	1	6.47	348	444	545	2 766 352
dumm-est	A	2	2 004	49 684	39	1	6.56	210	452	518	2 201 994
dyb-est	A	456	9 180	49 684	240	2	6.00	236	477	563	1 069 157
grimm-est	A	1	1 251	49 684	24	1	6.53	209	382	535	2 201 994
lys-est	A	4	3 504	49 684	1 244	1	5.44	320	642	642	1 681 882
mild-est	A	255	1 796	49 684	25	1	6.11	331	428	566	6 990 569
oft-est	A	1 838	17 750	49 684	1	0	6.50	280	532	599	2 766 352
stærk-est	A	256	17 100	49 684	508	2	6.69	288	649	710	1 887 007
tung-est	A	60	4 980	49 684	171	2	6.42	270	586	586	2 809 478
tætt-est	A	201	11 196	49 684	118	1	6.31	205	456	521	2 766 352
varm-est	A	6	7 529	49 684	668	1	6.47	295	612	612	2 201 994
dans-et	V	628	2 893	656 990	611	0	6.56	347	457	457	1 681 882
fjern-et	V	1 678	6 407	656 990	371	0	6.44	383	567	567	2 129 275
frist-et	V	241	1 252	656 990	59	0	6.08	376	523	523	2 766 352
hent-et	V	1 751	7 716	656 990	40	0	6.50	287	488	488	2 766 352
knytt-et	V	1 472	2 967	656 990	60	0	6.17	270	449	449	2 766 352
løb-et	V	1 155	12 113	656 990	1 531	0	5.97	336	488	515	1 069 157
plant-et	V	371	1 398	656 990	658	0	5.11	372	495	495	2 766 352
snakk-et	V	740	6 970	656 990	251	0	6.47	478	598	598	1 887 007
tygg-et	V	44	396	656 990	60	0	6.47	285	424	424	2 809 478
ændr-et	V	3 070	10 845	656 990	189	0	6.61	263	478	478	3 359 664
dann-es	V	877	6 046	254 132	1 463	0	6.22	286	409	518	2 129 275
knus-es	V	73	1 252	254 132	66	0	5.78	397	617	739	1 681 882
kræv-es	V	946	12 529	254 132	516	0	6.31	360	543	640	1 806 919
lav-es	V	1 526	26 074	254 132	39	0	5.92	468	658	658	1 806 919
nægt-es	V	93	3 119	254 132	65	0	5.92	214	527	625	2 766 352
ryst-es	V	73	4 047	254 132	51	0	6.03	318	549	549	2 766 352
sats-es	V	251	3 453	254 132	425	0	6.11	336	588	682	1 681 882
start-es	V	144	9 844	254 132	312	0	6.39	477	675	765	2 766 352
træn-es	V	128	2 736	254 132	862	0	6.22	439	610	735	2 129 275
vent-es	V	2 680	15 572	254 132	157	0	5.78	239	537	537	2 766 352

Table B.3: Simple words of Experiment 2 and associated lexical variables. The column names are the same as in the previous tables.

		Frequency						ms			
Word	Class	Word	Stem	Affix	Fam.	Cont.	Transp.	UP1	CUP	Dur.	Juncture
akut	A	610	–	–	37	40	NA	0	398	494	–
allergi	N	330	–	–	123	84	NA	0	430	507	–
apparat	N	405	–	–	311	45	NA	0	402	602	–
avis	N	1 898	–	–	382	376	NA	0	459	553	–
avle	V	68	–	–	234	23	NA	0	365	413	–
blind	A	610	–	–	117	115	NA	0	193	240	–
blæse	V	305	–	–	139	76	NA	0	333	492	–
blå	A	3 642	–	–	279	238	NA	0	177	266	–
bur	N	190	–	–	54	43	NA	0	232	298	–
cykle	V	409	–	–	648	37	NA	0	370	489	–
dreje	V	1 022	–	–	138	80	NA	0	329	465	–
dræbe	V	455	–	–	244	80	NA	0	343	440	–
drøj	A	37	–	–	6	9	NA	0	265	295	–
dække	V	1 377	–	–	536	27	NA	0	351	351	–
feber	N	432	–	–	150	86	NA	0	381	543	–
felt	N	947	–	–	353	169	NA	0	457	555	–
flod	N	340	–	–	151	146	NA	0	453	479	–
flov	A	296	–	–	10	11	NA	0	355	425	–
folk	N	21 557	–	–	1 682	1 390	NA	0	442	534	–
gammel	A	6 003	–	–	139	185	NA	0	331	398	–
glat	A	649	–	–	70	70	NA	0	304	421	–
gravid	A	921	–	–	40	53	NA	0	293	434	–
hjerte	N	2 605	–	–	427	530	NA	0	381	485	–
horn	N	293	–	–	168	110	NA	0	309	358	–
hugge	V	245	–	–	195	18	NA	0	259	380	–
høre	V	8 568	–	–	392	161	NA	0	434	502	–
kabine	N	78	–	–	133	112	NA	0	294	581	–
kiosk	N	118	–	–	75	66	NA	0	248	553	–
klappe	V	223	–	–	158	33	NA	0	330	434	–
klog	A	702	–	–	34	25	NA	0	200	390	–
klub	N	1 798	–	–	710	336	NA	0	273	299	–
koge	V	648	–	–	204	128	NA	0	436	480	–
kuppel	N	72	–	–	60	24	NA	0	357	386	–
kær	A	155	–	–	415	494	NA	0	254	319	–
lyd	N	2 238	–	–	668	560	NA	0	269	290	–
lygte	N	113	–	–	92	50	NA	0	303	421	–
løs	A	2 567	–	–	1 436	278	NA	0	362	452	–
låne	V	1 066	–	–	456	269	NA	0	426	515	–
male	V	732	–	–	592	331	NA	0	351	508	–
mappe	N	161	–	–	70	18	NA	0	314	433	–
marmor	N	161	–	–	120	163	NA	0	414	596	–
milits	N	53	–	–	389	41	NA	0	367	601	–
mør	A	104	–	–	26	35	NA	0	362	362	–
plat	A	123	–	–	138	108	NA	0	411	411	–
pulver	N	135	–	–	132	58	NA	0	206	439	–
pæl	N	87	–	–	102	48	NA	0	414	414	–
rank	A	146	–	–	31	22	NA	0	432	432	–
rejse	V	3 460	–	–	832	654	NA	0	303	463	–
skarp	A	988	–	–	72	86	NA	0	425	434	–
skære	V	1 257	–	–	394	93	NA	0	633	708	–
sløv	A	122	–	–	19	15	NA	0	305	339	–
snitte	V	61	–	–	492	19	NA	0	339	483	–
snu	A	106	–	–	7	11	NA	0	396	396	–
sove	V	1 742	–	–	146	223	NA	0	550	576	–
springe	V	980	–	–	297	22	NA	0	436	583	–
spæd	A	80	–	–	47	73	NA	0	557	557	–
sted	N	14 264	–	–	954	192	NA	0	434	434	–
stikke	V	981	–	–	363	55	NA	0	372	506	–
støbe	V	56	–	–	152	84	NA	0	445	573	–
stå	V	9 436	–	–	264	40	NA	0	401	401	–
sval	A	30	–	–	11	37	NA	0	521	521	–
syg	A	1 895	–	–	1 031	962	NA	0	377	404	–
sød	A	1 341	–	–	81	81	NA	0	348	380	–
tekst	N	1 800	–	–	560	526	NA	0	254	429	–
teori	N	1 256	–	–	410	84	NA	0	345	523	–
tælle	V	454	–	–	403	51	NA	0	336	378	–
ung	A	4 487	–	–	778	1 112	NA	0	160	186	–
villa	N	511	–	–	96	76	NA	0	327	418	–
vokse	V	1 373	–	–	470	377	NA	0	302	465	–
vælge	V	4 548	–	–	1 591	153	NA	0	353	510	–
væve	V	52	–	–	246	63	NA	0	369	507	–
våd	A	593	–	–	66	57	NA	0	277	329	–

Table B.4: Pairs of complex words and their stems whose relatedness was rated in the semantic transparency questionnaire. The words that were not included in the analyses are not listed (*film-en–film*, *nævn-ing–nævne*, *storm-ende–storme*).

Derived		Inflected	
Whole word	Stem	Whole word	Stem
appetitlig	appetit	baroner	baron
arkiv	ark	blomster	blomst
basere	base	brikker	brik
blokere	blok	buketten	buket
bryggeri	brygge	byttede	bytte
bøjning	bøje	dannes	danne
distinktiv	distinkt	danset	danse
drilleri	drille	duftede	dufte
dyrisk	dyr	dummest	dum
dæmonisk	dæmon	dybest	dyb
effektiv	effekt	dyrkede	dyrke
flytbar	flytte	dystede	dyste
frugtbar	frugt	fjernet	fjerne
frygtsom	frygt	fløjtende	fløjte
fyldning	fylde	fristet	friste
fæstning	fæste	fuglen	fugl
føleri	føle	grimmest	grim
gangbar	gang	gruppen	gruppe
gemytlig	gemyt	hentet	hente
grovhed	grov	hvaler	hval
hekseri	heks	kampen	kamp
husere	hus	knuses	knuse
huskbar	huske	knyttet	knytte
hyleri	hyl	kræves	kræve
hæderlig	hæder	laves	lave
hældning	hælde	legater	legat
højhed	høj	lignede	ligne
instinktiv	instinkt	lovende	love
intensiv	intens	lysest	lys
ironisk	ironi	lyttede	lytte
jordisk	jord	løbet	løbe
jævnlig	jævn	løftede	løfte
kammerateri	kammerat	manglede	mangle
klarhed	klar	mildest	mild
klikbar	klikke	musikken	musik
kongelig	konge	nægtes	nægte
kostbar	koste	oftest	ofte
kursiv	kurs	pejsen	pejs
kækhed	kæk	personer	person
langsom	lang	plantet	plante
livlig	liv	pligten	pligt
lyrisk	lyre	prinsen	prins
massiv	masse	problemer	problem
metallisk	metal	reoler	reol
motivere	motiv	rullende	rulle
mørkning	mørk	rystes	ryste
nemlig	nem	saksen	saks
notere	note	satses	satse
nydelig	nyde	sejlende	sejle
nænsom	nænne	sekunder	sekund
nærhed	nær	skubbende	skubbe
ordning	ordne	smilende	smile
organisk	organ	snakket	snakke
orientere	orient	spillede	spille

Table B.4: Word pairs of the semantic transparency questionnaire (continued)

Derived		Inflected	
Whole word	Stem	Whole word	Stem
parkere	park	startes	starte
produktiv	produkt	strejkende	strejke
punktere	punkt	stærkest	stærk
refleksiv	refleks	trænes	træne
ruinere	ruin	tungest	tung
rytmisk	rytme	tygget	tygge
sangbar	sang	tættest	tæt
sceneri	scene	vandrende	vandre
skævhed	skæv	varmest	varm
slidsom	slid	ventes	vente
smitsom	smitte	visitter	visit
smøleri	smøle	vrængende	vrænge
smørbar	smøre	væltede	vælte
sparsom	spare	ændret	ændre
stolthed	stolt		
strenghed	streng		
stridbar	strid		
subjektiv	subjekt		
surhed	sur		
svineri	svin		
særhed	sær		
sætning	sætte		
taktisk	takt		
tolkning	tolke		
træfning	træffe		
tvivlsom	tvivl		
tydelig	tyde		
typisk	type		
tænksom	tænke		
tøjeri	tøj		
vagtsom	vagt		
ventilere	ventil		
virkelig	virke		
voldsom	vold		
åbenbar	åben		

Table B.5: Nonwords of Experiment 2.

akle	fjekseri	hubbende	knelt	mammel	olkning	rå'nere	stonghed	tval
astetitlig	fjyrnet	huftbar	knultet	mampen	orsen'tere	salfes	strantet	tvirfsom
avirf	flirkelig	hun'kere	knuvles	mapsiv	osk	sarmir	strappe	tvæstning
blenghed	flyngbar	hup'tere	knælge	mavle	panstenktiv	sjarisk	strinktiv	tvøjning
blirp	fovl	hurbe	knøbe	mekst	pa'rener	skajf	strykte	tylfelig
bly'kere	frasket	hypseri	kopselig	mote'sere	pavid	skappe	svildest	tæmsest
bløreri	fyngsom	hælkning	kovte	mugtbar	pjygget	ska'sere	svonisk	tøvleri
blål	fyntes	hænksom	kvangbar	mulver	plavl	skilende	svørkning	vabsker
boftest	følskeri	hæplerlig	kyrisk	muppel	plindsom	skunner	svåd	vaftsom
brejlende	gingbar	hærkest	kyvhed	myldning	plyg	skæfe	sytning	vense'lere
brimmest	glarp	hørphed	kælp	mytlig	premlig	sløIk	sæpshed	vikster
brur	glurhed	høtse	kætshed	mæbe	proktiv	smigten	søvl	vojdsom
brælle	glystes	irganisk	køjtende	mændret	pronsen	smilksom	tajlisk	vritmisk
daftede	gnillende	jerdisk	lalf	mæse	præfning	smyttede	teflyksiv	vryggeri
dajret	gnod	jostbar	larkiv	mætellisk	pusken	smøvbar	tekri	vræskende
di'menisk	golk	jænglig	larsksom	mø'ketten	pyr'kere	snuk	timpisk	vuktiv
drannes	gorphed	kalserateri	le'bine	mølf	pæjsen	snæplet	tirtes	vursiv
drejf	gryppen	kaves	lejse	mølt	rakut	spalsom	tjikke	vylmen
drersener	haksen	kivlbar	lilmelig	nar'tere	regøter	sparat	torbning	vænks
drilferi	handrende	kjur	lirfede	nejkende	remektiv	spineri	tormende	vølle
dryd	hanglede	kleber	lomfende	nove	rentansiv	spingede	trank	vørmest
durlest	hellegi	klerte	luglen	nudelig	ro'belmer	spittede	trestede	ybelbar
dvidbar	hemfet	klosk	lypsest	nuglsom	rokse	spripse	trung	æves
dvikker	hiske	klyrkede	lørket	nugtes	roler	spæjl	trælfes	ølteri
dymfest	hog	klømhed	låkse	næfthed	rældtede	spæl	trå	
dæfle	homster	kneje	malje	nærdning	røftede	steft	tuvlest	

Table B.6: Training and warm-up items of Experiment 2.

Words				Nonwords			
barndom	fatte	komfortabel	strejfe	bjasset	knamle	mevlagtig	skurme
bryllup	fejlagtig	rødme	tykkelse	brissen	knurpe	mimler	stejfe
bøtte	glasset	samle	vissen	bøtse	krimfe'rabel	mirndom	tyrfelse
fabrikant	knude	skule		favle	mefri'kant	plødme	

Appendix C

Items in Experiments 3a & 3b

Table C.1: Compound Words of Experiments 3a and 3b and associated lexical variables. The columns are defined as follows: Under 'Frequency', 'Word' is the base form frequency of the whole word; '2nd Const' is the lemma frequency of the second constituent as an independent word; 'Constituent Fam.' are the type counts of morphological family members for the first and second constituents; 'Cont.' is the type count of morphologically related continuation forms; 'Juncture' is the frequency of the bigram crossing the juncture between the constituent morphemes. All these measures are extracted from *Korpus90/2000*, the values are for the 43.6 million tokens in that corpus. 'UP1' and 'CUP' are UP-measures in ms for the auditory stimulus, 'Dur.' the duration of the same. The same variables are listed under the same column names for the remaining words in Experiments 3a and 3b in the following tables.

		Frequency		Constituent Fam.			ms			
Word	Class	Word	2nd Const.	1st	2nd	Cont.	UP1	CUP	Dur.	Juncture
adresse-bog	N	21	16 462	36	1 441	1	406	661	847	142 151
blod-rød	A	27	7 383	340	388	2	212	406	564	433 910
blød-gøre	V	30	91 696	35	940	4	300	300	705	76 027
bort-føre	V	39	16 512	175	1 201	15	237	459	745	28 153
brev-veksle	V	3	581	122	239	3	229	542	773	2 743
damp-koge	V	0	3 099	103	203	1	362	495	777	6 668
fast-spænde	V	2	3 687	194	377	2	387	643	966	268 666
fod-bold	N	3 487	2 891	762	1 129	924	287	391	607	84 968
frost-kold	A	2	5 462	97	135	1	345	657	728	13 834
gade-dreng	N	23	10 047	255	460	16	283	487	674	2 174 402
gen-splejse	V	49	2 958	59	99	10	230	522	890	1 060 206
guld-medalje	N	169	532	624	114	9	269	385	880	33 935
hals-hugge	V	9	926	53	184	6	308	561	741	64 789
hjerne-vaske	V	3	1 760	175	339	7	391	538	844	432 034
hyle-tone	N	23	2 424	13	324	3	303	459	922	3 644 600
kontor-chef	N	951	3 867	189	790	5	373	562	939	23 857
kunst-historie	N	85	15 139	658	928	9	279	488	985	95 815
lager-plads	N	21	16 951	113	861	2	400	688	817	34 958
lokal-radio	N	77	2 715	353	549	27	320	777	1 055	12 793
luft-tørre	V	4	1 934	472	96	5	310	569	872	513 799
luksus-vare	N	1	4 544	228	726	4	267	751	907	193 127
lys-stråle	N	31	608	413	272	4	222	585	894	360 592
mange-doble	V	12	87	48	489	5	298	572	889	2 174 402
menneske-masse	N	10	8 822	507	491	5	243	847	931	761 522
møbel-fabrik	N	24	2 250	155	740	10	267	590	873	62 897
mål-rette	V	83	5 712	334	1 339	6	252	538	729	12 793
navn-give	V	21	62 986	134	1 343	6	345	440	844	1 612 525
pande-hår	N	31	5 024	50	390	1	281	424	655	97 687
pant-sætte	V	17	41 121	48	1 798	7	261	466	699	268 666
plan-lægge	V	556	30 496	157	840	11	183	459	707	83 338
prøve-smage	V	18	3 035	277	322	4	346	658	954	1 389 144
rask-melde	V	1	4 982	8	364	2	323	397	810	10 188
risiko-gruppe	N	29	13 337	119	1 440	3	274	647	886	1 926 764
selv-sikker	A	101	14 657	1 044	1 400	3	307	550	747	50 214
silke-glat	A	7	1 121	201	69	1	332	745	821	790 672
slæde-hund	N	8	5 049	34	479	11	486	655	879	97 687
små-snakke	V	9	6 970	979	260	6	260	676	810	57 188
sne-blind	A	0	1 368	148	118	2	371	433	681	142 151
sol-bade	V	14	629	544	404	9	484	561	920	92 736
spare-penge	N	95	15 375	165	713	1	558	782	980	100 543
strand-vejr	N	1	3 643	174	284	0	459	574	803	132 119
struktur-problem	N	5	28 422	104	1 174	3	177	780	1 210	34 958
studie-lån	N	88	2 115	241	455	3	399	684	866	1 872 068
styrke-træne	V	7	2 772	53	861	4	427	626	1 005	3 644 600
succes-rig	A	86	3 987	142	661	2	271	780	844	42 504
vakuum-pakke	V	0	1 824	39	426	4	348	579	844	146 956
videre-sælge	V	21	12 300	5	835	2	417	650	1 009	1 389 144
vild-lede	V	52	6 713	5	1 432	5	335	406	823	214 928
vind-mølle	N	66	557	203	236	121	337	383	742	33 935
æble-most	N	21	76	130	9	0	242	470	778	761 522

Table C.2: Prefixed words of Experiments 3a and 3b and associated lexical variables. The column names are the same as in the previous table.

		Frequency		Constituent Fam.			ms			
Word	Class	Word	2nd Const.	1st	2nd	Cont.	UP1	CUP	Dur.	Juncture
be-fri	V	247	12 017	235	1 696	82	173	444	566	314 674
be-grave	V	129	1 409	235	517	109	145	466	727	790 672
be-klage	V	299	1 779	235	198	18	151	600	688	372 745
be-laste	V	148	243	235	333	5	161	390	619	1 872 068
be-mærke	V	374	8 112	235	997	15	161	294	741	761 522
be-ordre	V	64	1 689	235	135	5	280	487	726	40 539
be-sejre	V	259	629	235	262	7	169	541	688	1 389 144
be-skærme	V	3	168	235	285	3	141	321	715	1 389 144
be-spotte	V	0	115	235	20	7	141	385	640	1 389 144
be-vogte	V	38	423	235	67	4	157	460	643	432 034
gen-bruge	V	59	37 703	253	2 103	10	266	330	783	33 758
gen-danne	V	18	6 057	253	1 433	7	299	369	672	2 597 157
gen-finde	V	115	46 652	253	133	2	307	614	723	49 118
gen-lyd	N	159	4 817	253	677	5	274	503	580	83 338
gen-læse	V	20	13 148	253	654	2	269	539	682	83 338
gen-rejse	V	40	12 544	253	831	4	303	489	756	34 128
gen-skin	N	30	67	253	180	1	323	395	571	1 060 206
gen-starte	V	9	9 844	253	344	4	401	401	761	1 060 206
gen-syn	N	269	2 665	253	1 134	21	451	451	678	1 060 206
gen-vælge	V	18	18 471	253	1 590	2	273	439	682	52 127
mis-farve	V	4	1 142	50	880	4	310	427	814	74 038
mis-greb	N	2	762	50	972	0	361	401	672	32 777
mis-klang	N	106	375	50	165	0	378	457	652	1 670 176
mis-klæde	V	0	2 142	50	586	1	489	582	930	1 670 176
mis-røgt	N	18	10	50	13	5	357	417	689	42 504
mis-tanke	N	1 212	6 152	50	738	8	402	402	762	2 364 484
mis-tolke	V	0	874	50	202	4	422	422	744	2 364 484
mis-tro	N	171	3 674	50	229	8	404	524	658	2 364 484
mis-vise	V	0	33 993	50	1 299	3	359	429	752	193 127
mis-vækst	N	31	3 816	50	410	3	324	404	734	193 127
sam-arbejde	N	8 118	22 929	74	3 476	10	387	387	902	828 162
sam-drift	N	18	2 022	74	692	10	394	498	788	5 900
sam-handel	N	114	2 782	74	2 522	23	408	498	864	37 470
sam-køre	V	12	19 750	74	740	2	440	562	902	39 647
sam-liv	N	255	29 888	74	1 488	57	393	461	683	89 528
sam-råd	N	520	7 159	74	1 143	14	420	494	790	53 354
sam-spil	N	493	7 280	74	2 134	28	415	583	808	108 709
sam-tale	N	1 762	12 540	74	1 360	139	379	557	860	148 927
sam-tid	N	157	50 411	74	3 215	101	412	525	731	148 927
sam-vær	N	703	1 413 049	74	478	49	404	469	688	12 099
u-fred	N	58	3 278	1 476	496	10	286	345	583	47 044
u-held	N	1 086	2 221	1 476	110	46	210	450	450	12 485
u-jævn	A	81	1 365	1 476	59	6	215	290	619	1 322
u-klog	A	22	3 003	1 476	39	2	189	335	623	109 961
u-skarp	A	9	988	1 476	71	3	162	410	532	325 660
u-sund	A	84	2 888	1 476	374	5	199	445	524	325 660
u-tryg	A	121	1 585	1 476	58	6	155	509	509	165 503
u-tæt	A	80	11 196	1 476	119	5	191	415	494	165 503
u-ven	N	14	9 591	1 476	546	20	201	552	552	18 907
u-ægte	A	115	2 217	1 476	211	3	297	423	646	1 016

Table C.3: Particle prefixed words of Experiments 3a and 3b and associated lexical variables. The column names are the same as in the previous tables.

		Frequency		Constituent Fam.			ms			
Word	Class	Word	2nd Const.	1st	2nd	Cont.	UP1	CUP	Dur.	Juncture
af-blege	V	0	47	362	85	3	254	386	746	6 831
af-dæmpe	V	4	1 108	362	66	6	291	518	682	16 745
af-grænse	V	77	491	362	629	9	238	443	741	28 456
af-høre	V	144	27 589	362	391	3	283	655	750	15 166
af-klare	V	141	9 244	362	311	7	237	584	696	6 025
af-kræve	V	56	12 529	362	515	4	252	642	788	6 025
af-sende	V	22	16 866	362	463	28	268	577	729	49 004
af-spejle	V	198	473	362	181	6	244	424	773	49 004
af-vente	V	287	15 272	362	155	7	284	533	748	12 823
af-vige	V	40	681	362	56	21	302	596	766	12 823
bag-binde	V	3	3 236	149	620	1	294	468	625	15 064
bag-gård	N	130	3 172	149	568	42	271	357	560	170 608
bag-hjul	N	97	1 069	149	264	23	243	400	603	150 881
bag-hoved	N	17	12 725	149	1 359	5	287	462	747	150 881
bag-krop	N	19	8 977	149	354	6	252	386	515	7 699
bag-linje	N	2	4 708	149	545	2	360	360	728	188 813
bag-lomme	N	14	1 028	149	165	3	372	476	642	188 813
bag-pote	N	1	287	149	4	2	266	537	666	6 071
bag-tæppe	N	30	1 669	149	226	4	418	418	624	608 674
bag-vægt	N	1	4 106	149	451	0	353	551	632	28 840
om-bejle	V	0	100	226	10	3	300	535	731	57 416
om-døbe	V	18	697	226	90	2	321	435	748	5 900
om-eksamen	N	1	1 101	226	235	0	254	307	924	2 201 994
om-kranse	V	7	32	226	115	5	255	395	792	39 647
om-kreds	N	86	1 831	226	515	8	239	395	681	39 647
om-ryste	V	0	4 452	226	51	0	293	375	735	53 354
om-skole	V	24	633	226	1 650	5	257	457	719	108 709
om-sværme	V	0	105	226	28	2	256	451	789	108 709
om-vej	N	117	25 070	226	988	4	230	442	542	12 099
op-brud	N	204	1 192	359	798	10	310	572	672	10 052
op-digte	V	8	481	359	296	3	329	456	769	12 600
op-dyrke	V	65	3 073	359	285	10	329	378	720	12 600
op-fange	V	144	3 146	359	380	5	239	450	617	31 913
op-fordre	V	603	228	359	108	6	226	505	720	31 913
op-hæve	V	421	2 989	359	81	8	242	551	710	23 291
op-høje	V	48	29 275	359	1 326	9	257	340	710	23 291
op-kalde	V	25	18 164	359	232	5	284	495	683	6 668
op-løse	V	169	6 147	359	1 435	12	288	552	734	224 985
op-takt	N	251	1 988	359	481	18	399	524	674	54 843
til-bede	V	38	8 914	212	70	12	282	432	812	92 736
til-flugt	N	227	1 306	212	451	13	202	413	723	62 897
til-knytte	V	14	2 967	212	59	5	310	516	780	145 361
til-kæmpe	V	51	3 560	212	1 048	5	262	379	743	145 361
til-lære	V	1	12 328	212	1 059	0	242	692	793	1 049 739
til-løb	N	281	4 179	212	1 529	20	235	379	567	1 049 739
til-råb	N	105	495	212	93	6	221	588	672	12 793
til-stræbe	V	84	406	212	98	2	189	511	780	586 982
til-støde	V	6	2 114	212	234	3	191	496	765	586 982
til-træde	V	114	5 911	212	237	29	192	639	807	425 663

Table C.4: Simple words of Experiments 3a and 3b and associated lexical variables. The column names are the same as in the previous tables.

		Frequency		Constituent Fam.			ms			
Word	Class	Word	2nd Const.	1st	2nd	Cont.	UP1	CUP	Dur.	Juncture
almen	A	335	–	54	–	103	393	393	688	–
anonym	A	445	–	5	–	14	320	320	745	–
atom	N	106	–	338	–	600	511	511	574	–
base	N	403	–	13	–	59	313	313	574	–
bil	N	5 075	–	471	–	841	360	360	360	–
billede	N	5 416	–	490	–	27	260	260	537	–
blank	A	255	–	20	–	65	263	263	419	–
blok	N	258	–	95	–	161	394	394	394	–
bonde	N	305	–	192	–	354	478	478	478	–
byde	V	1 101	–	5	–	15	362	362	564	–
bygge	V	3 029	–	503	–	579	306	306	390	–
debat	N	4 552	–	113	–	217	293	293	504	–
famle	V	30	–	1	–	6	413	413	779	–
fjern	A	621	–	276	–	443	374	374	456	–
forme	V	401	–	241	–	91	591	591	689	–
gigant	N	85	–	131	–	178	282	282	638	–
hulke	V	30	–	6	–	12	320	320	570	–
hus	N	6 656	–	540	–	1 066	398	398	499	–
kede	V	179	–	3	–	15	568	568	568	–
klikke	V	166	–	13	–	11	321	321	571	–
kredit	N	142	–	147	–	255	264	264	573	–
krænge	V	55	–	3	–	6	266	266	537	–
mineral	N	47	–	67	–	104	148	148	654	–
nænne	V	21	–	2	–	4	345	345	626	–
nær	A	3 870	–	154	–	284	419	419	419	–
nøgle	N	313	–	135	–	228	430	430	632	–
printe	V	41	–	31	–	27	338	338	557	–
produkt	N	1 380	–	505	–	825	327	327	655	–
ride	V	345	–	150	–	166	371	371	553	–
sitre	V	42	–	1	–	5	437	437	657	–
skylde	V	40	–	45	–	5	440	440	682	–
slutte	V	982	–	223	–	20	429	429	657	–
sløre	V	65	–	20	–	8	670	670	800	–
snit	N	615	–	32	–	48	487	487	568	–
splitte	V	184	–	30	–	40	456	456	695	–
spytte	V	131	–	32	–	12	455	455	628	–
stof	N	2 667	–	181	–	277	470	470	563	–
storm	N	481	–	96	–	162	529	529	607	–
stram	A	549	–	11	–	36	595	595	595	–
stykke	N	6 442	–	27	–	18	410	410	625	–
styrte	V	258	–	37	–	10	548	548	682	–
sving	N	530	–	71	–	124	456	456	564	–
synge	V	1 522	–	0	–	45	361	361	619	–
trafik	N	1 274	–	354	–	610	274	274	633	–
tysse	V	9	–	2	–	8	417	417	530	–
tøj	N	4 190	–	186	–	318	368	368	437	–
ulme	V	19	–	1	–	6	313	313	510	–
vand	N	8 419	–	756	–	1 246	354	354	354	–
vende	V	3 351	–	24	–	41	432	432	432	–
vitrine	N	6	–	3	–	9	417	417	725	–

Table C.5: Nonwords of Experiments 3a and 3b.

blavn-gøve	mys-håle	begrafte	miskrang	afblevle	ommyste	bafte	mern
blosk-tød	mæde-tvund	behogte	mispranke	afdætse	omsvælke	bløk	metom
blud-frøre	nast-spælke	behurdre	mismise	afflase	omtrole	bovne	mihant
borm-løre	palk-sælte	behærke	missmøgt	afhøtse	omtrøbe	bramle	mitre
bruld-dalje	plajf-mægge	beklige	mistrolf	afmæve	omverf	bymfe	næske
dalk-komfe	ravl-mevte	belaspe	sambannel	afpojle	opbruft	flygge	plerat
drelse-boft	rol-knade	berirf	sam-dehejde	afsevse	opdivte	fofte	prylle
dvade-jeng	sipse-galf	beskyrme	samhift	afspinse	opfarske	gilefte	rallik
dvare-pefte	skuf-tøske	besvotte	samkøtse	afvunte	opfomle	glær	rirme
frast-kovs	smy-trakke	be-'særne	samlilk	afvøge	ophyrke	gulke	slede
hukses-varfe	snajn-blirk	gendakte	samråft	bagbynde	ophøpse	hetrit	sløpse
hukter-mobelm	stralk-vir	genmarte	samspilg	baggårf	ophøve	hinemal	smænne
hæple-jost	strenske-marfe	genmæse	samtase	bagjuts	opkapse	hodukt	snivt
høbel-brifte	stylje-fån	genremfe	samtuft	bagknate	oplitse	huts	spirfte
klen-splemfe	svod-brogn	genskirf	samvør	bagkroved	opvrakt	hutte	spyvle
kronter-smef	syftes-tvig	genslinde	u-bigte	bagkrovn	tilflurf	hynge	stolk
kyngs-histre	triseke-luppe	genstyd	ufrevs	baglønje	tilgride	hænenym	strum
lerne-varke	trøve-smajte	gensylk	uglyg	bagstromme	tilhæmpe	kinte	stårte
løger-plang	trål-bætte	genvuge	uhejt	bagtælge	tilknøb	kivle	til'men
mals-spugge	tvande-glår	genvæfte	u-jæks	bagvæmf	tilkrib	knil	tøps
mev-hoksle	vaktem-parske	misfreb	ukløg	ombævle	tillore	knof	umle
minge-hoble	vipsere-sølge	mis-gnarve	urarp	omhoksemen	tilmytte	knøgle	vasp
mokel-fradie	vond-mønge	misholke	usumf	omjalm	tilsjøde	krælke	vode
myle-tofe	vrel-simler	mishæde	u-verm	omkibsk	tilstrybe	kvank	vo'tvine
myrke-trune	vyl-spede	mishækst	uvæt	omkramfe	tiltævse	kving	ysse

Table C.6: Training and warm-up items of Experiments 3a and 3b.

Training Words				Training Nonwords			
efterløn	kendetegne	rulle	subkultur	efterprøn	hensprik	lause	submuflur
ernære	mobiltelefon	sele	undslippe	ernærle	hobrilmelekon	overplygge	undmippe
fremstille	overskygge	sideordne	vanskæbne	fremknille	knejle	sitseurdne	vankubne
henblik	pause	skrap		frulle	kængsetøgne	strap	

Appendix D

Affixes

Table D.1: Affixes used in the experiments. "Der" are derivational and "infl" inflectional affixes. Class refers to the word class(es) of the experimental items with the given affix.

PREFIXES

Affix	Type	Category	Experiments 1	2	3	Function/translation
be-	der	V	+		+	repeated/stronger X, transitive X, supply with X
gen-	der	N/V	+		+	to X again
mis-	der	N/V	+		+	negative X, change X for the worse
sam-	der	N/V	+		+	X together
u-	der	A/N	+		+	not X
und-	der	V	+			avoid X
van-	der	N/V	+			not X, negative X

PARTICLES

Affix	Type	Category	Experiments 1	2	3	Function/translation
af-	der	N/V	+		+	off, away
bag-	der	A/N/V	+		+	behind, back
efter-	der	N	+			after
med-	der	N/V	+			with
om-	der	N/V	+		+	about, re-
op-	der	N/V	+		+	up
over-	der	A/N/V	+			over
til-	der	N/V	+		+	to/towards

SUFFIXES

Affix	Type	Category	Experiments 1	2	3	Function/translation
-agtig	der	A	+			like X
-bar	der	A	+	+		able to be X'ed
-ere	der	V	+	+		to make X
-eri	der	N	+	+		place for/state of/case of X
-hed	der	N	+	+		quality of being X
-isk	der	A	+	+		like X/characterised by property X
-iv	der	A		+		characterised by property X
-lig	der	A	+	+		characterised by property X
-ning	der	N		+		result of X
-mæssig	der	A	+			relating to X
-skab	der	N	+			state of X
-som	der	A	+	+		characterised by property X
-ede	infl	V		+		verb past tense
-en	infl	N		+		noun definite, common gender
-ende	infl	V		+		verb present particle
-er	infl	N		+		noun plural
-es	infl	V		+		verb passive
-est	infl	A		+		adjective superlative
-et	infl	V		+		verb past particle

Appendix E

Productivity questionaire

Læs hver sætning i venstre kolonne omhyggeligt. Nogle af sætningerne indeholder ord der ikke findes på dansk, men godt kunne gøre det. Læg særligt mærke til det fremhævede ord i hver sætning og vurder om det ville lyde naturligt på dansk, i den sammenhæng det står. Angiv din vurdering ved at sætte ring om et tal mellem 1 og 7, hvor 1 betyder at ordet virker helt unaturligt på dansk, og 7 betyder at det virker helt naturligt. Det er vigtigt at vurdere selve ordene, ikke sætningerne, selvom sætningerne er nødvendige for at forstå ordene.

	HELT UNATURLIGT					HELT NATURLIGT
1) Det går godt med Davids bausk. Det er en **bauskmæssig** succes.	1	2	3	4	5	6	7
2) Karla elsker at vylfe. Især når hun kan komme til at **opvylfe** sine omgivelser.	1	2	3	4	5	6	7
3) En plaut er en vanskelig ting. Bo er nu god til at **plautere.**	1	2	3	4	5	6	7
4) Det er fint når en mand er prølm. Men Johans **prølmskab** går nu lidt for vidt.	1	2	3	4	5	6	7
5) Hugo er meget glad for at brøngste. Han vil gerne **genbrøngste** sit hus.	1	2	3	4	5	6	7
6) Tanja kan godt lide at vylfe. Men hun må ikke **bevylfe** sin lillebror.	1	2	3	4	5	6	7
7) Rebekka er ved at lære at narfe. Men hun har ikke lært at **tilnarfe** endnu.	1	2	3	4	5	6	7
8) Mange kan godt lide at bauske. Berta er nu mere **bausksom** end de fleste.	1	2	3	4	5	6	7
9) Mad med vylf smager dejligt. Men det kan nu godt komme til at smage for **vylfisk**.	1	2	3	4	5	6	7
10) Nogle er gode til at kirme. Stefan er bedst til at **undkirme**.	1	2	3	4	5	6	7
11) Kim plejer at have mere end en brøngst. Men hans ene **bagbrøngst** er gået i stykker.	1	2	3	4	5	6	7
12) Man skal være omhyggelig med at plaute. Det er farligt at komme til at **vanplaute**.	1	2	3	4	5	6	7
13) Et tvorft er et svært instrument at spille. Men blokfløjten har en ret **tvorftlig** lyd.	1	2	3	4	5	6	7
14) At dilpe træ er sjovt. Kristine kan især godt lide at **samdilpe** forskellige træsorter.	1	2	3	4	5	6	7
15) Man skal passe på med at narfe. Man kan nemt komme til at **misnarfe** noget**.**	1	2	3	4	5	6	7
16) Hver by har en tvorft. Derfor er en regional **intertvorft** vigtig.	1	2	3	4	5	6	7
17) Hun er dygtig til at plaute. Men derfor behøver hun vel ikke **overplaute** det hele.	1	2	3	4	5	6	7
18) Den nye statue er meget narf. Faktisk er dens **narfhed** blevet verdensberømt.	1	2	3	4	5	6	7
19) Karoline er en god tvorft. Men hendes **medtvorft** er bedre.	1	2	3	4	5	6	7
20) Man kan godt kirme derhjemme. Men det virker nu bedst i et **kirmeri**.	1	2	3	4	5	6	7
21) Det er ikke nok at dilpe sit gulv. Man skal også huske at **efterdilpe** det.	1	2	3	4	5	6	7
22) Henrik er sur på sin nye brøngst. Den er slet ikke **brøngstagtig** nok.	1	2	3	4	5	6	7
23) Frederik er en bausk type. Hans kone er til gengæld helt **ubausk**.	1	2	3	4	5	6	7
24) At prølme sko er kedeligt. Især er det træls når man skal **omprølme** dem.	1	2	3	4	5	6	7
25) Man skal passe på med at kirme køkkenet. Man kan hurtigt komme til at **afkirme** det.	1	2	3	4	5	6	7
26) Trine kan normalt godt lide at dilpe sit tøj. Men hendes nye trøje er alt for **dilpbar**.	1	2	3	4	5	6	7
27) Jeg har en rigtig flot prølm. Men jeg er ikke særlig god til at **prølme**.	1	2	3	4	5	6	7

1

The questionnaire used to assess the productivity of the affixes used in Experiment 1. Combinations of the affixes with non-stems are on this page, novel combinations with real stems on the next page.

	HELT UNATURLIGT						HELT NATURLIGT
28) Min første tamagotchi er sød, men dens **medtamagotchi** er sødere.	1	2	3	4	5	6	7
29) Hendes **retrohed** er imponerende.	1	2	3	4	5	6	7
30) Når hullet først er lavet, kan man ikke **upierce** folk igen.	1	2	3	4	5	6	7
31) En **browsermæssig** beslutning er nødvendig.	1	2	3	4	5	6	7
32) Indimellem skal man **omboote** sin computer.	1	2	3	4	5	6	7
33) Designerne er ved at **retroere** moden.	1	2	3	4	5	6	7
34) Lise er en nørd, men hendes **nørdskab** er ikke så ekstrem.	1	2	3	4	5	6	7
35) Han prøver vist nærmest at **overbrowse** hele nettet.	1	2	3	4	5	6	7
36) Han er ikke **chattelig** i dag.	1	2	3	4	5	6	7
37) Man kunne jo prøve at **genspamme**.	1	2	3	4	5	6	7
38) Man bør ikke **bagchatte** andre.	1	2	3	4	5	6	7
39) Man kan nemt komme til at lade sig **bechatte** af fremmede over internettet.	1	2	3	4	5	6	7
40) Musikere er slemme til at **tinnitusse** deres publikum med høj musik.	1	2	3	4	5	6	7
41) De unge kan godt lide at **samchatte** over nettet.	1	2	3	4	5	6	7
42) Han er generelt ikke så **mailsom**.	1	2	3	4	5	6	7
43) Sådan et vækkeur medfører en vis risiko for **snoozeri**.	1	2	3	4	5	6	7
44) Visse firmaer bruger mails til at **opspamme** folk.	1	2	3	4	5	6	7
45) Han skal nå at **afspamme** sin mailboks.	1	2	3	4	5	6	7
46) Nogen mener at man skal passe på med at **tilchatte** fremmede på internettet.	1	2	3	4	5	6	7
47) Man kan nemt få en lidt **tinnitusisk** fornemmelse af høj musik.	1	2	3	4	5	6	7
48) Det er svært at **undchatte**.	1	2	3	4	5	6	7
49) Man kan nemt komme til at **mismaile**.	1	2	3	4	5	6	7
50) Trine kan godt lide at **interchatte**.	1	2	3	4	5	6	7
51) Det er muligt at **vanmaile**.	1	2	3	4	5	6	7
52) Serien er lidt **sitcomagtig**.	1	2	3	4	5	6	7
53) Hun har ingen meninger selv, hun kommer altid bare til at **efterchatte** andre.	1	2	3	4	5	6	7
54) Min e-mail adresse er heldigvis ikke **spambar**.	1	2	3	4	5	6	7

Alder: ________________ Køn: M / K

Modersmål: ________________

Afleveres til Sten Vikner eller Ken Christensen i Syntax & Morphology seminarerne i uge 47 eller i Laura Ballings dueslag i løbet af ugen. Mange tak for hjælpen!

Appendix F

Consent form and background questionnaire

Samtykkeerklæring

Jeg er blevet informeret om eksperimentet, og jeg har forstået de forklaringer, som jeg har fået om eksperimentet. Jeg er klar over, at jeg kan trække mig ud af eksperimentet på ethvert tidspunkt, hvis jeg ønsker det. Jeg er klar over, at mine oplysninger bliver behandlet fortroligt.

Med min underskrift tilkendegiver jeg, at jeg deltager frivilligt i forsøget.

Dato: __________

Underskrift: ____

ID-nummer: __________

Informed consent form used in Experiments 1 to 3.

Spørgsmål

1. Navn: ________________________ 2. Alder: _______

3. Mail: ________________________ 4. Køn: M / K

5. Er du højre- eller venstrehåndet? Højre / Venstre

6. Har du problemer med din hørelse? Ja / Nej

Hvis ja, hvilke problemer? ____

7. Har du nogensinde haft tale- eller læseproblemer (ordblindhed, stammen etc.)? Ja / Nej

Hvis ja, hvad var/er problemets art? ____

8. Hvad er dit modersmål? ____________________

Din mors modersmål? ____________________ Din fars? ____________________

9. Har du boet i udlandet mere end et år ad gangen? Ja / Nej

Hvis ja, hvor og hvor længe? ____

____

10. Angiv dit kendskab til andre sprog end dansk, 1 = kun lidt, 5 = meget godt

Sproget	Kendskab
________________	1 2 3 4 5
________________	1 2 3 4 5
________________	1 2 3 4 5
________________	1 2 3 4 5
________________	1 2 3 4 5

Background questionnaire used in Experiments 1 to 3.

Bibliography

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

Alegre, M. & Gordon, P. (1996) Red rats eater exposes recursion in children's word formation. *Cognition*, 60, 65–82.

Alegre, M. & Gordon, P. (1999) Frequency Effects and the Representational Status of Regular Inflections. *Journal of Memory and Language*, 40, 41–61.

Allan, R., Holmes, P. & Lundskær-Nielsen, T. (1995) *Danish: a comprehensive grammar*. London: Routledge.

Alvarez, C.J., Carreiras, M. & Taft, M. (2001) Syllables and Morphemes: Contrasting Frequency Effects in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 545–555.

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

Andrews, S. (1989) Frequency and neighborhood size effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 802–814.

Andrews, S. (1992) Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 234–254.

Arndt, H. (1997) Grammatisk Analyse. Manuscript, Dept. of Linguistics, University of Aarhus.

Baayen, R.H. (1992) Quantitative aspects of morphological productivity. In G. Booij. & J. van Marle, editors, *Yearbook of Morphology 1991*, pages 109–149. Dordrecht: Kluwer Academic Publishers.

Baayen, R.H. (1993) On frequency, transparency and productivity. In G. Booij & J. van Marle, editors, *Yearbook of Morphology 1992*, pages 181–208. Dordrecht: Kluwer Academic Publishers.

Baayen, R.H. (2007) Storage and computation in the mental lexicon. In G. Libben & G. Jarema, editors, *The Mental Lexicon: Core Perspectives*, pages 81–104. Amsterdam: Elsevier.

Baayen, R.H. (2008a) *Exploratory data analysis: An introduction to R for the langage sciences*. Cambridge, UK: Cambridge University Press.

Baayen, R.H. (2008b) languageR, version 0.95. Available from the Internet: http://cran.r-project.org/web/packages/languageR/index.html (Accessed October 18, 2008).

Baayen, R.H., Davidson, D. & Bates, D.H. (2008a) Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, in press.

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., Feldman, L.B. & Schreuder, R. (2006) Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55, 290–313.

Baayen, R.H., Levelt, W.J.M., Schreuder, R. & Ernestus, M. (2008b) Paradigmatic structure in speech production. *Proceedings of the Chicago Linguistics Society 43*, 1, 1–29.

Baayen, R.H., McQueen, J.M., Dijkstra, T. & Schreuder, R. (2003) Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R.H. Baayen & R. Schreuder, editors, *Morphological Structure in Language Processing*, pages 463–484. Berlin: Mouton de Gruyter.

Baayen, R.H. & Moscoso del Prado Martín, F. (2005) Semantic density and past-tense formation in three Germanic languages. *Language*, 81, 666–698.

Baayen, R.H., Piepenbrock, P. & Gulikers, L. (1995) The CELEX lexical database. CD-ROM. University of Pennsylvania, Philadelphia, PA: Linguistic Data Consortium.

Baayen, R.H., Tweedie, F.J. & Schreuder, R. (2002) The Subjects as a Simple Random Effect Fallacy: Subject Variability and Morphological Family Effects in the Mental Lexicon. *Brain and Language*, 81, 55–65.

Baayen, R.H., Wurm, L.H. & Aycock, J. (2007) Lexical Dynamics for Low-Frequency Complex Words: A Regression Study Across Tasks and Modalities. *The Mental Lexicon*, 2, 419–463.

Balling, L. (2002) *Storage and rule-application in Danish noun-plural formation: A psycholinguistic investigation.* Master's thesis, University of Essex.

Balling, L.W. (2008) A brief introduction to regression designs and mixed-effects modelling by a recent convert. *Copenhagen Studies in Language*, 36, in press.

Balling, L.W. & Baayen, R.H. (2008) Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, in press.

Balota, D. (1994) Visual Word Recognition: The Journey From Features to Meaning. In M.A. Gernsbacher, editor, *Handbook of Psycholinguistics*, pages 303–358. San Diego, CA: Academic Press.

Balota, D.A. (1990) The role of meaning in word recognition. In D.A. Balota, G.B. Flores d'Arcais & K. Rayner, editors, *Comprehension Processes in Reading*, pages 9–32. Hillsdale, NJ: Lawrence Erlbaum Associates.

Balota, D.A. & Chumbley, J.I. (1984) Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 340–357.

Balota, D.A., Cortese, M.J., Hutchison, K.A., Neely, J.H., Nelson, D., Simpson, G.B. & Treiman, R. (2002) The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. Web: http://elexicon.wustl.edu/. Accessed September 7, 2007.

Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.H. & Yap, M.J. (2004) Visual Word Recognition of Single-Syllable Words. *Journal of Experimental Psychology: General*, 133, 283–316.

Basbøll, H. (1998) Nyt om stødet i rigsdansk - om samspillet mellem lydstruktur og ordgrammatik. *Danske Studier*, 1998, 33–86.

Basbøll, H. (2001) Fuldproduktive bøjningsendelser og stød: nogle konsekvenser af en ny model for ordstruktur. In P. Jarvad, F. Gregersen, L. Heltoft, J. Lund & O. Togeby, editors, *Sproglige åbninger. E som Erik, H som 70. Festskrift til Erik Hansen 18. september 2001*. København: Hans Reitzels Forlag.

Basbøll, H. (2003) Prosody, productivity and word structure: the stød pattern of Modern Danish. *Nordic Journal of Linguistics*, 26, 5–44.

Basbøll, H. (2005) *The Phonology of Danish*. The Phonology of the World's Languages Series. Oxford: Oxford University Press.

Bates, D. M. & Sarkar, D. (2006) *lme4: Linear mixed effects models using S4 classes*. R package version 0.995-2. Available from: http://lib.stat.cmu.edu/R/CRAN/.

Bates, D. M. & Sarkar, D. (2007) *The lme4 library*. R package version 0.9975-11. Available from: http://lib.stat.cmu.edu/R/CRAN/.

Bauer, L. (1983) *English Word-Formation*. Cambridge Textbooks in Linguistics. Cambridge, UK: Cambridge University Press.

Bauer, L. (1999) Is the morpheme dead? *Acta Linguistica Hafniensa*, 31, 7–26.

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

Beard, R. (1998) Derivation. In A. Spencer & A.M. Zwicky, editors, *The Handbook of Morphology.* Oxford: Blackwell.

Becker-Christensen, C., Appel, V., Katlev, J., Rasmussen, I. & Troelsgaard, T. (2005) *Politikens Nudansk Ordbog med etymologi.* København: JP/Politikens Forlagshus.

Bentin, S. & Feldman, L.B. (1990) The contribution of morphological and semantic relatedness to repetition priming at short and long lags: Evidence from Hebrew. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 42, 693–711.

Bergen, B.J. (2004) The Psychological Reality of Phonaesthemes. *Language*, 80, 290–311.

Berlingske Aviskorpus (1999) Berlingske Tidende & Weekendavisen 1999. Reference in Nimb, S. Danske adverbier mellem leksikon og syntaks. Ph.d.-thesis, University of Copenhagen. Available from http://cst.dk/sanni/phd/. Accessed December 18, 2007.

Bertram, R., Baayen, R.H. & Schreuder, R. (2000a) Effects of Family Size for Complex Words. *Journal of Memory and Language*, 42, 390–405.

Bertram, R. & Hyönä, J. (2003) The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language*, 48, 615–634.

Bertram, R., Laine, M. & Karvinen, K. (1999) The Interplay of word Formation Type, Affixal Homonymy, and Productivity in Lexical Processing: Evidence from a Morphologically Rich Language. *Journal of Psycholinguistic Research*, 28, 213–226.

Bertram, R., Pollatsek, A. & Hyönä, J. (2004) Morphological parsing and the use of segmentation cues in reading Finnish compounds. *Journal of Memory and Language*, 51, 325–345.

Bertram, R., Pollatsek, A. & Hyönä, J. (2007) Predictability Within a Word: Evidence from Finnish Compounds. Poster presented at the 13th Annual Conference on Architectures and Mechanisms for Language Processing. Turku, August 24-27, 2007.

Bertram, R., Schreuder, R. & Baayen, R.H. (2000b) 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.

Bick, E. (1996-2008) Interface for querying Korpus90 and Korpus2000. http://corp.hum.sdu.dk/cqp.html. Accessed September 22, 2008.

Blevins, J.P. (1999) Productivity and Exponence. *Behavioral and Brain Sciences*, 22, 1015–1016.

Bloomfield, L. (1933) *Language*. New York: Henry Holt.

Bölte, J. & Uhe, M. (2004) When is all understood and done? The psychological reality of the recognition point. *Brain and Language*, 88, 133–147.

Booij, G. (1993) Against split morphology. In G. Booij & J. van Marle, editors, *Yearbook of Morphology 1993*, pages 27–49. Dordrecht: Kluwer Academic Publishers.

Boudelaa, S. & Marslen-Wilson, W. (2004a) Abstract morphemes and lexical representation: the CV-Skeleton in Arabic. *Cognition*, 92, 271–303.

Boudelaa, S. & Marslen-Wilson, W.D. (2000) Non-concatenative morphemes in language processing: Evidence from Modern Standard Arabic. In A. Cutler, J.M. McQueen & R. Zondervaan, editors, *Proceedings of the Workshop on Spoken Word Access Processes*, pages 23–26. Nijmegen: Max-Planck-Institute for Psycholinguistics.

Boudelaa, S. & Marslen-Wilson, W.D. (2004b) Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language*, 90, 106–116.

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.

Burani, C., Dovetto, F.M., Thornton, A.M. & Laudanna, A. (1997) Accessing and naming suffixed pseudo-words. In G.E. Booij & J. van Marle, editors, *Yearbook of Morphology 1996*, pages 55–72. Dordrecht: Kluwer Academic Publishers.

Burani, C., Marcolini, S., De Luca, M. & Zoccolotti, P. (2008) Morpheme-based reading aloud: Evidence from dyslexic and skilled Italian readers. *Cognition*, 108, 243–262.

Burani, C. & Thornton, A.M. (2003) The interplay of root, suffix and whole-word frequency in processing derived words. In R.H. Baayen & R. Schreuder, editors, *Morphological structure in language processing*, pages 157–207. Berlin: Mouton de Gruyter.

Butterworth, B. (1983) Lexical representation. In B. Butterworth, editor, *Language Production (Vol. II): Development, writing and other language processes*, pages 257–294. London: Academic Press.

Bybee, J.L. (1995) Regular Morphology and the Lexicon. *Language and Cognitive Processes*, 10, 425–455.

Caramazza, A., Laudanna, A. & Romani, C. (1988) Lexical access and inflectional morphology. *Cognition*, 28, 297–332.

Carreiras, M. & Grainger, J. (2004) Sublexical representations and the "front end" of visual word recognition. *Language and Cognitive Processes*, 19, 321–331.

Chalker, S. & Weiner, E. (1994) *The Oxford Dictionary of English Grammar*. Oxford: Oxford University Press.

Chomsky, N. (1970) Remarks on nominalization. In R. Jacobs & P. Rosenbaum, editors, *Readings in transformational grammar*. Waltham, MA: Ginn & Co.

Chomsky, N. & Halle, M. (1968) *The sound pattern of English*. Cambridge, MA: MIT Press.

Chéreau, C., Gaskell, M.G. & Dumay, N. (2007) Reading spoken words: Orthographic effects in auditory priming. *Cognition*, 102, 341–360.

Clahsen, H. (1999) Lexical entries and the rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991–1013.

Cohen, J. (1983) The cost of dichotomization. *Applied Psychological Measurement*, 7, 249–254.

Colé, P., Beauvillain, C. & Segui, J. (1989) On the Representation and Processing of Prefixed and Suffixed Derived Words: A Differential Frequency Effect. *Journal of Memory and Language*, 28, 1–13.

Colé, P., Segui, J. & Taft, M. (1997) Words and Morphemes as Units for Lexical Access. *Journal of Memory and Language*, 37, 312–330.

Coltheart, M., Davelaar, E., Jonasson, J.T. & Besner, D. (1977) Access to the Internal Lexicon. In S. Dornic, editor, *Attention and Performance VI*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (2001) DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256.

Crawley, M.J. (2002) *Statistical computing: An introduction to data analysis using S-plus*. Chichester: Wiley.

Cutler, A. (1981) Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990. *Cognition*, 10, 65–70.

Cutler, A., Hawkins, J.A. & Gilligan, G. (1985) The suffixing preference: A processing explanation. *Linguistics*, 23, 723–758.

Daelemans, W., Zavrel, J., van der Sloot, K. & van den Bosch, A. (2003) TiMBL: Tilburg memory based leaner, version 5.0. Online at http://ilk.uvt.nl/software.html. Accessed September 10, 2008.

Davis, M.H., Marslen-Wilson, W.D. & Gaskell, M. G. (2002) Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218–244.

De Jong, N.H., Feldman, L., Schreuder, R., Pastizzo, M. & Baayen, R.H. (2002) The processing and representation of Dutch and English compounds: Peripheral morphological, and central orthographic effects. *Brain and Language*, 81, 555–567.

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 Saussure, F. (1916/1966) *Course in General Linguistics*. New York: McGraw-Hill.

de Vaan, L., Schreuder, R. & Baayen, R.H. (2007) Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon*, 2, 1–24.

Det Danske Sprog- og Litteraturselskab (1982-93) Korpus 90. Online. Available from: http://korpus.dsl.dk/e-resurser/k90_info.php?lang=dk. Accessed October 25, 2006.

Det Danske Sprog- og Litteraturselskab (1998-2002) Korpus 2000. Online. Available from: http://korpus.dsl.dk/e-resurser/k2000_info.php?lang=dk. Accessed October 25, 2006.

Det Danske Sprog- og Litteraturselskab (2008) Interface for querying Korpus90 and Korpus2000. http://korpus.dsl.dk/korpus2000/indgang.php?lang=dk. Accessed September 22, 2008.

Diderichsen, P. (1987) *Elementær Dansk Grammatik*. København: Gyldendal, third edition.

Diependaele, K., Sandra, D. & Grainger, J. (2005) Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes*, 20, 75–114.

Diependaele, K., Sandra, D. & Grainger, J. (2007) Masked priming with Dutch compounds: Is the whole greater than the sum of the parts? Paper presented at the 5th International Workshop on Morphological Processing, Marseilles, June 2007.

Drews, E. & Zwitserlood, P. (1995) Morphological and Orthographic Similarity in Visual Word Recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1098–1116.

Elbro, C. & Arnbak, E. (1996) The Role of Morpheme Recognition and Morphological Awareness in Dyslexia. *Annals of Dyslexia*, 46, 209–240.

Feldman, L.B. (1994) Beyond Orthography and Phonology: Differences between Inflections and Derivations. *Journal of Memory and Language*, 33, 442–470.

Feldman, L.B. (2000) Are Morphological Effects Distinguished From the Effects of Shared Meaning and Shared Form? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1431–1444.

Feldman, L.B., Basnight-Brown, D. & Kanai, Y. (2007) Effects of Morphological Family Size in the Forward Masked Lexical Decision Task. Poster presented at the 13th Annual Conference on Architectures and Mechanisms for Language Processing.

Feldman, L.B., Basnight-Brown, D. & Pastizzo, M.J. (2006) Semantic influences on morphological facilitation: Concreteness and family size. *The Mental Lexicon*, 1, 59–84.

Feldman, L.B. & Larabee, J. (2001) Morphological Facilitation Following Prefixed but Not Suffixed Primes: Lexical Architecture or Modality-Specific Processes. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 680–691.

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.

Feustel, T.C., Shiffrin, R.M. & Salasoo, A. (1983) Episodic and Lexical Contributions to the Repetition Effect in Word Identification. *Journal of Experimental Psychology: General*, 112, 309–346.

Fiorentino, R. & Poeppel, D. (2007) Compound words and structure in the lexicon. *Language and Cognitive Processes*, 22, 953–1000.

Fischer-Jørgensen, E. (1989) Phonetic analysis of the stød in standard Danish. *Phonetica*, 46, 1–59.

Ford, M.A., Marslen-Wilson, W.D. & Davis, M.H. (2003) Morphology and frequency: Contrasting methodologies. In R.H. Baayen & R. Schreuder, editors, *Morphological structure in language processing*, pages 89–124. Berlin: Mouton de Gruyter.

Forster, K. I. & Forster, J.C. (2006) DMDX version 3.1.4.5. On-line. Http://www.u.arizona.edu/ jforster/dmdx.htm.

Forster, K.I. & Azuma, T. (2000) Masked priming for prefixed words with bound stems: Does submit prime permit? *Language and Cognitive Processes*, 15, 539–561.

Fowler, C.A., Napps, S.E. & Feldman, L. (1985) Relations among regular and irregular morphologically related words in the lexicon as revealed in repetition priming. *Memory & Cognition*, 10, 680–698.

Frauenfelder, U.H. & Schreuder, R. (1992) Constraining psycholinguistic models of morphological processing and representation: the role of productivity. In G.E. Booij & J. van Marle, editors, *Yearbook of Morphology 1991*, pages 165–183. Dordrecht: Kluwer Academic Publishers.

Frost, R. (1998) Toward a Strong Phonological Theory of Visual Word Recognition: True Issues and False Trails. *Psychological Bulletin*, 123, 71–99.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M. & Marslen-Wilson, W. (2000) Morphological priming: Dissociation of phonological, semantic, and morphological factors. *Memory & Cognition*, 28, 1277–1288.

Frost, R., Forster, K.I. & Deutsch, A. (1997) What can we learn from the morphology of Hebrew? A masked priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 829–856.

Gagné, C.L. & Spalding, T.L. (2008) Constituent integration during the processing of compound words: Does it involve the use of relational structures? *Journal of Memory and Language*, in press.

Gaskell, M.G. & Marslen-Wilson, W.D. (1997) Integrating Form and Meaning: A Distributed Model of Speech Perception. *Language and Cognitive Processes*, 12, 613–656.

Giraudo, H. & Grainger, J. (2000) Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes*, 15, 421–444.

Giraudo, H. & Grainger, J. (2001) Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review*, 8, 127–131.

Giraudo, H. & Grainger, J. (2003a) On the role of derivational affixes in recognizing complex words. In R.H. Baayen & R. Schreuder, editors, *Morphological structure in language processing*, pages 209–232. Berlin: Mouton de Gruyter.

Giraudo, H. & Grainger, J. (2003b) A supralexical model for French derivational morphology. In E.M.H. Assink & D. Sandra, editors, *Reading Complex Words. Cross-Language Studies*, pages 139–156. New York: Kluwer Academic Publishers.

Goldinger, S.D. (1996a) Auditory Lexical Decision. *Language and Cognitive Processes*, 11, 559–567.

Goldinger, S.D. (1996b) Words and voices: Episodic traces in spoken word identification and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.

Gonnerman, L.M., Seidenberg, M.S. & Andersen, E.S. (2007) Graded Semantic and Phonological Similarity Effects in Priming: Evidence for a Distributed Connectionist Approach to Morphology. *Journal of Experimental Psychology: General*, 136, 323–345.

Grainger, J. (1990) Word Frequency and Neigborhood Frequency Effects in Lexical Decision and Naming. *Journal of Memory and Language*, 29, 228–244.

Grainger, J. & Segui, J. (1990) Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, 47, 191–198.

Grønnum, N. (2005) *Fonetik og fonologi almen og dansk.* København: Akademisk Forlag, third edition.

Grønnum, N. & Basbøll, H. (2001) Consonant Length, Stød and Morae in Standard Danish. *Phonetica*, 58, 230–253.

Grosjean, F. (1980) Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, 299–310.

Grosjean, F. (1996) Gating. *Language and Cognitive Processes*, 11, 597–604.

Hahn, U. & Nakisa, R.C. (2000) German Inflection: Single Route or Dual Route? *Cognitive Psychology*, 41, 313–360.

Halle, M. & Mohanan, K.P. (1985) Segmental Phonology in Modern English. *Linguistic Inquiry*, 16, 57–116.

Hansen, Aa. (1943) *Stødet i Dansk.* Det Kgl. Danske Videnskabernes Selskab, Historisk-Filologisk Meddelelser, Bind XXIX, Nr. 5. København: Munksgaard.

Hansen, Aa. (1967) *Moderne Dansk.* København: Grafisk Forlag.

Harley, T.A. (2008) *The Psychology of Language. From Data to Theory.* Hove & New York: Psychology Press, third edition.

Harrell, F. E. (2001) *Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York: Springer.

Hasher, L. & Zacks, R. T. (1984) Automatic Processing of Fundamental Information. The Case of Frequency of Occurrence. *American Psychologist*, 39, 1372–1388.

Haskell, T., MacDonald, M.C. & Seidenberg, M. (2003) Language learning and innateness: Some implications of compounds research. *Cognitive Psychology*, 47, 119–163.

Haspelmath, M. (2002) *Understanding Morphology.* London: Arnold.

Hay, J.B. (2001) Lexical frequency in morphology: is everything relative? *Linguistics*, 39, 1041–1070.

Hay, J.B. (2002) From speech perception to morphology: affix ordering revisited. *Language*, 78, 527–555.

Hay, J.B. & Baayen, R.H. (2005) Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, 9, 342–348.

Hay, J.B. & Plag, I. (2004) What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory*, 22, 565–596.

Hjorth, E., Kristensen, K., Lorentzen, H. & Trap-Jensen, L., editors (2003-2005) *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab.

Hockett, C.F. (1954) Two models of grammatical description. *Word*, 10, 210–234.

Hyönä, J. & Olson, R. (1995) Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1430–1440.

Inhoff, A.W., Briihl, D. & Schwartz, J. (1996) Compound word effects differ in reading, on-line naming, and delayed naming tasks. *Memory and Cognition*, 24, 466–476.

Jackendoff, R. (2002) *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jaeger, J.J., Lockwood, A.H., Kemmerer, D.L., Van Valin, R.D., Murphy, B.W. & Khalak, H.G. (1996) A Positron Emission Tomography study of regular and irregular verb morphology in English. *Language*, 72, 451–497.

Jarema, G. & Libben, G. (2007) Introduction: Matters of Definition and Core Perspectives. In G. Jarema & G. Libben, editors, *The Mental Lexicon: Core Perspectives*, pages 1–6. Oxford: Elsevier.

Jarvad, P. (1999) *Nye Ord. Ordbog over nye ord i dansk 1955-1998*. København: Gyldendal.

Ji, H., Gagné, C.L. & Kemps, R.J.J.K. (2006) Compounds are processed more quickly than monomorphemic words. Paper presented at the 5th International Conference on the Mental Lexicon, Montreal. October 2006.

Joanisse, M.F. & Seidenberg, M.S. (1999) Impairments in verb morphology after brain injury: a connectionist model. *Proceedings of the National Academy of Science*, 96, 7592–7597.

Josefsson, G. (2005) How could Merge be free and word formation restricted: The case of compounding in Romance and Germanic. *Working Papers in Scandinavian Syntax*, 75, 55–96.

Järvikivi, J., Bertram, R. & Niemi, J. (2006) Affixal Salience and the Processing of Derivational Morphology: The Role of Suffix Allomorphy. *Language and Cognitive Processes*, 21, 394–431.

Kazanina, N., Dukova-Zheleva, G., Geber, D., Kharlamov, V. & Tonciulescu, K. (2008) Decomposition into multiple morphemes during lexical access: A masked priming study of Russian nouns. *Language and Cognitive Processes*, 23, 800–823.

Kempley, S.T. & Morton, J. (1982) The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, 73, 441–454.

Kemps, R., Wurm, L., Ernestus, M., Schreuder, R. & Baayen, R.H. (2005a) Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes*, 20, 43–73.

Kemps, R.J.J.K., Ernestus, M., Schreuder, R. & Baayen, R.H. (2005b) Prosodic cues for morphological complexity: the case of Dutch plural nouns. *Memory & Cognition*, 33, 430–446.

Keuleers, E. (2008) *Memory-based learning of inflectional morphology.* Ph.D. thesis, University of Antwerp.

Kimura, Doreen (1999) *Sex and Cognition.* Cambridge, MA: The MIT Press.

Kiparsky, P. (1982) From cyclic phonology to lexical phonology. In N. van der Hulst & N. Smith, editors, *The structure of phonological representations (Part I).* Dordrecht: Foris.

Klepousniotou, E. & Baum, S.R. (2007) Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, 20, 1–24.

Kostic, A., Markovic, T. & Baucal, A. (2003) Inflectional morphology and word meaning: Orthogonal or co-implicative cognitive domains? In R.H. Baayen & R. Schreuder, editors, *Morphological structure in language processing*, pages 1–43. Berlin: Mouton de Gruyter.

Krott, A., Baayen, R.H. & Schreuder, R. (2001) Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics*, 39, 51–93.

Kuperman, V. (2008) *Information-theoretical approaches to morphology.* Ph.D. thesis, Radboud University, Nijmegen.

Kuperman, V., Bertram, R. & Baayen, R.H. (2008a) Morphological Dynamics in Compound Processing. *Language and Cognitive Processing*, in press.

Kuperman, V., Schreuder, R., Bertram, R. & Baayen, R.H. (2008b) Reading of multimorphemic Dutch compounds: towards a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, accepted for publication.

Laine, M., Vainio, S. & Hyönä, J. (1999) Lexical Access Routes to Nouns in a Morphologically Rich Language. *Journal of Memory and Language*, 40, 109–135.

Landauer, T.K., Foltz, P. W. & Laham, D. (1998) Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259–284.

Laudanna, A. & Burani, C. (1995) Distributional Properties of Derivational Affixes: Implications for Processing. In L.B. Feldman, editor, *Morphological Aspects of Language Processing*, pages 131–154. Hillsdale, NJ: Lawrence Erlbaum Associates.

Laudanna, A., Burani, C. & Cermele, A. (1994) Prefixes as processing units. *Language and Cognitive Processes*, 9, 295–316.

Lüdeling, A. & De Jong, N. H. (2002) German particle verbs and word-formation. In N. Dehé, R. Jackendoff, A. McIntyre & S. Urban, editors, *Verb-particle explorations*, pages 315–333. Berlin: Mouton de Gruyter.

Lehtonen, M., Cunillera, T., Rodríguez-Fornells, A., Hultén, A., Tuomainen, J. & Laine, M. (2007) Recognition of morphologically complex words in Finnish: Evidence from event-related potentials. *Brain Research*, 1148, 123–137.

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.

Levelt, W.J.M. (1989) *Speaking. From Intention to Articulation.* Cambridge, MA: MIT Press.

Libben, G. (2006) Why Study Compound Processing? An overview of the issues. In G. Libben & G. Jarema, editors, *The Representation and Processing of Compound Words*, pages 1–22. Oxford, UK: Oxford University Press.

Libben, G., Gibson, M., Yoon, Y.-B. & Sandra, D. (2003) Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, 84, 50–64.

Lieber, R. (1992) *Deconstructing morphology: word formation in syntactic theory.* Chicago: University of Chicago Press.

Lieberman, E., Michel, J.B., Jackson, J., Tang, T. & Nowak, M.A. (2007) Quantifying the evolutionary dynamics of language. *Nature*, 449, 713–716.

Longtin, C.-M. & Meunier, F. (2005) Morphological decomposition in early visual word processing. *Journal of Memory and Language*, 53, 26–41.

Longtin, C.-M., Segui, J. & Hallé, P.A. (2003) Morphological priming without morphological relationship. *Language and Cognitive Processes*, 18, 313–334.

Luce, P.A. (1986) A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, 39, 155–158.

Lukatela, G., Eaton, T., Lee, C.H., Carello, C. & Turvey, M.T. (2002) Equal Homophonic Priming With Words and Pseudohomophones. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 3–21.

Lukatela, G. & Turvey, M.T. (1994a) Visual Lexical Access Is Initially Phonological: 1. Evidence From Associative Priming by Words, Homophones, and Pseudohomophones. *Journal of Experimental Psychology: General*, 2, 107–128.

Lukatela, G. & Turvey, M.T. (1994b) Visual Lexical Access Is Initially Phonological: 2. Evidence From Phonological Priming by Homophones and Pseudohomophones. *Journal of Experimental Psychology: General*, 123, 331–353.

MacCallum, R.C., Zhang, S., Preacher, K.J. & Rucker, D.D. (2002) On the Practice of Dichotomization of Quantitative Variables. *Psychological Methods*, 7, 19–40.

Magnuson, J.S., Dixon, J.A., Tanenhaus, M.K. & Aslin, R.N. (2007) The Dynamics of Lexical Competition During Spoken Word Recognition. *Cognitive Science*, 31, 1–24.

Marshall, C.R. & van der Lely, H.K.J. (2006) A challenge to current models of past tense inflection: The impact of phonotactics. *Cognition*, 100, 302–320.

Marslen-Wilson, W.D. (1984) Function and process in spoken word recognition. In H. Bouma & D.G. Bouwhuis, editors, *Attention and Performance X: Control of language processes*, pages 125–150. Hillsdale, NJ: Lawrence Erlbaum Associates.

Marslen-Wilson, W.D. (1987) Functional parallelism in spoken word-recognition. *Cognition*, 25, 71–102.

Marslen-Wilson, W.D. (1990) Activation, competition and frequency in lexical access. In G. Altmann, editor, *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, pages 148–172. Cambridge, MA: MIT Press.

Marslen-Wilson, W.D., Bozic, M. & Randall, B. (2008) Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes*, 23, 394–421.

Marslen-Wilson, W.D. & Tyler, L.K. (2003) Capturing underlying differentiation in the human language system. *Trends in Cognitive Science*, 7, 62–63.

Marslen-Wilson, W.D., Tyler, L.K., Waksler, R. & Older, L. (1994) Morphology and Meaning in the English Mental Lexicon. *Psychological Review*, 191, 3–33.

Marslen-Wilson, W.D. & Warren, P. (1994) Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101, 653–675.

Marslen-Wilson, W.D. & Welsh, A. (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29–63.

Marslen-Wilson, W.D. & Zhou, X. (1999) Abstractness, Allomorphy, and Lexical Architecture. *Language and Cognitive Processes*, 14, 321–352.

Matthews, P. (1991) *Morphology. 2nd edition.* Cambridge, UK: Cambridge University Press, second edition.

McClelland, J.L. (1999) Cognitive Modelling, Connectionist. In R.A. Wilson & F.C. Keil, editors, *The MIT Encyclopedia of the Cognitive Sciences*, pages 137–141. Cambridge, Mass: Bradford Books/MIT Press.

McClelland, J.L. & Elman, J.L. (1986) The TRACE Model of Speech Perception. *Cognitive Psychology*, 16, 1–86.

McClelland, J.L. & Patterson, K. (2002) Rules or connections in past-tense inflections: what does the evidence rule out? *Trends in Cognitive Sciences*, 6, 465–472.

McQueen, J. & Cutler, A. (2001) Spoken word access procedures: An introduction. *Language and Cognitive Processes*, 16, 469–490.

Meunier, F. & Longtin, C.-M. (2007) Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, 56, 457–471.

Meunier, F. & Marslen-Wilson, W.D. (2004) Regularity and irregularity in French verbal inflection. *Language and Cognitive Processes*, 19, 561–580.

Meunier, F. & Segui, J. (1999a) Frequency Effects in Auditory Word Recognition: The Case of Suffixed Words. *Journal of Memory and Language*, 41, 327–344.

Meunier, F. & Segui, J. (1999b) Morphological Priming Effect: The Role of Surface Frequency. *Brain and Language*, 68, 54–60.

Meunier, F. & Segui, J. (2002) Cross-Modal Morphological Priming in French. *Brain and Language*, 81, 89–102.

Meunier, F., Seigneuric, A. & Spinelli, E. (2008) The morpheme gender effect. *Journal of Memory and Language*, 58, 88–99.

Milin, P., Kuperman, V., Kostic, A. & Baayen, R.H. (2008) Words and paradigms bit by bit: An information-theoretic approach to the processing of inflection and derivation. In J.P. Blevins & J. Blevins, editors, *Analogy in Grammar: Form and Acquisition.* Oxford: Oxford University Press. In press.

Morrison, C.M. & Ellis, A.W. (1995) Roles of Word Frequency and Age of Acquisition in Word Naming and Lexical Decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 116–133.

Moscoso del Prado Martín, F. (2007) How can we measure morphemic complexity? Paper presented at the 5th International Workshop on Morphological Processing, Marseilles, June 2007.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R. & Baayen, R.H. (2004a) Morphological Family Size in a Morphologically Rich Language: The Case of Finnish Compared With Dutch and Hebrew.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 1271–1278.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N.H. & Baayen, R.H. (2005) Changing places: A cross-language perspective on frequency and family counts in Dutch and Hebrew. *Journal of Memory and Language*, 53, 496–512.

Moscoso del Prado Martín, F., Kostic, A. & Baayen, R.H. (2004b) Putting the bits together: an information theoretical perspective on morphological processing. *Cognition*, 94, 1–18.

New, B., Brysbaert, M., Segui, J., Ferrand, L. & Rastle, K. (2004) The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, 51, 568–585.

Norris, D. (1994) Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52, 189–234.

Orsolini, M. & Marslen-Wilson, W. (1997) Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes*, 12, 1–47.

Pastizzo, M.J. & Feldman, L.B. (2002) Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 244–249.

Pastizzo, M.J. & Feldman, L.B. (2004) Morphological processing: A comparison between free and bound stem facilitation. *Brain and Language*, 90, 31–39.

pgAdmin Development Team (2005) *pgAdmin III PostgreSQL tools, version 1.2.1.* Available from www.pgadmin.org.

Pinker, S. (1991) Rules of Language. *Science*, 253, 530–535.

Pinker, S. (1997a) *How the mind works.* Harmondsworth: Penguin.

Pinker, S. (1997b) Words and rules in the human brain. *Nature*, 387, 547–548.

Pinker, S. (1999) *Words and Rules: The Ingredients of Language*. London: Phoenix.

Pinker, S. & Jackendoff, R. (2005) The faculty of language: what's special about it? *Cognition*, 95, 201–236.

Pinker, S. & Ullman, M.T. (2002) The past and future of the past tense. *Trends in Cognitive Sciences*, 6, 456–463.

Plag, I. & Baayen, R.H. (2008) Suffix ordering and morphological processing. *Language*, accepted for publication.

Plaut, D.C. & Gonnerman, L.M. (2000) Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, 15, 445–485.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S. & Patterson, K. (1996) Understanding Normal and Impaired Word Reading in Quasi-Regular Domains. *Psychological Review*, 103, 56–115.

Pollatsek, A., Reichle, E. & Rayner, K. (2003) Modeling eye movements in reading: Extensions of the E-Z Reader model. In Y. Hyönä, R. Radach & H. Deubel, editors, *The mind's eye: Cognitive and applied aspects of eye movement research*, pages 361–390. Amsterdam: Elsevier.

Pulvermüller, F. (1999) Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253–279.

R Development Core Team (2006) *R: A Language and Environment for Statistical Computing, version 2.4.1.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raab, D.H. (1962) Statistical Facilitation of Simple Reaction Times. *Transactions of the New York Academy of Sciences*, 24, 574–590.

Radeau, M., Morais, J., Mousty, P. & Bertelson, P. (2000) The Effect of Speaking Rate on the Role of the Uniqueness Point in Spoken Word Recognition. *Journal of Memory and Language*, 42, 406–422.

Radeau, M. & van Berkum, J.J.A. (1996) Gender Decision. *Language and Cognitive Processes*, 11, 605–610.

Ramscar, M. (2002) The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, 45, 45–94.

Rastle, K., Davis, M.H., Marslen-Wilson, W.D. & Tyler, L.K. (2000) Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, 15, 507–537.

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.

Raveh, M. & Rueckl, J.G. (2000) Equivalent Effects of Inflected and Derived Primes: Long-term Morphological Priming in Fragment Completion and Lexical Decision. *Journal of Memory and Language*, 42, 103–119.

Reid, A. & Marslen-Wilson, W.D. (2003) Organising Principles in Lexical Representation: Evidence from Polish. In R.H. Baayen & R. Schreuder, editors, *Morphological structure in language processing*, pages 287–336. Berlin: Mouton de Gruyter.

Reimer, J.F., Lorsbach, T.C. & Bleakney, D.M. (2008) Automatic semantic feedback during visual word recognition. *Memory & Cognition*, 36, 641–658.

Rodd, J., Gaskell, G. & Marslen-Wilson, W. (2002) Making Sense of Semantic Ambiguity: Semantic Competition in Lexical Access. *Journal of Memory and Language*, 46, 245–266.

Rubenstein, H., Lewis, S.S. & Rubenstein, M.A. (1971) Evidence for Phonemic Recoding in Visual Word Recognition. *Journal of Verbal Learning and Verbal Behavior*, 10, 645–657.

Rueckl, J.G. & Aicher, K.A. (2008) Are CORNER and BROTHER morphologically complex? Not in the long term. *Language and Cognitive Processes*, in press.

Rueckl, J.G., Mikolinski, M., Raveh, M., Miner, C.S. & Mars, F. (1997) Morphological Priming, Fragment Completion, and Connectionist Networks. *Journal of Memory and Language*, 36, 382–405.

Salverda, A., Dahan, D. & McQueen, J. (2003) The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51–89.

Salverda, A., Dahan, D., Tanenhaus, M.K., Crosswhite, K., Masharov, M. & McDonough, J. (2007) Effects of prosodically modulated sub-phonetic variation on lexical competition. *Cognition*, 105, 466–476.

Schreuder, R. & Baayen, R.H. (1995) Modeling Morphological Processing. In L.B. Feldman, editor, *Morphological Aspects of Language Processing*, pages 131–154. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schreuder, R. & Baayen, R.H. (1997) How Complex Simplex Words Can Be. *Journal of Memory and Language*, 37, 118–139.

Schreuder, R., Grendel, M., Poulisse, N., Roelofs, A. & van der Voort, M. (1990) Lexical processing, morphological complexity and reading. In D.A. Balota & G.B. Flores d'Arcais, editors, *Comprehension processes in reading*, pages 125–141. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schriefers, H., Zwitserlood, P. & Roelofs, A. (1991) The Identification of Morphologically Complex Spoken Words: Continuous Processing or Decomposition. *Journal of Memory and Language*, 30, 26–47.

Seidenberg, M. (1987) Sublexical structures in visual word recognition: Access units or orthographic redundancy? In M. Coltheart, editor, *Attention and performance XII: The psychology of reading.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Seidenberg, M.S. & Gonnerman, L.M. (2000) Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, 4, 353–361.

Seidenberg, M.S. & McClelland, J.L. (1989) A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.

Seidenberg, M.S., Waters, G.S., Sanders, M. & Langer, P. (1984) Pre- and post-lexical loci of contextual effects on word recognition. *Memory & Cognition*, 12, 315–328.

Sereno, J. & Jongman, A. (1997) Processing of English inflectional morphology. *Memory & Cognition*, 25, 425–437.

Skousen, R., Lonsdale, D. & Parkinson, D.B., editors (2002) *Analogical modeling : an exemplar-based approach to language.* Amsterdam: John Benjamins.

Soveri, A., Lehtonen, M. & Laine, M. (2007) Word frequency and morphological processing in Finnish revisited. *The Mental Lexicon*, 2, 359–385.

Spencer, A. (2001) The paradigm-based model of morphosyntax. *Transactions of the Philological Society*, 99, 279–313.

Stadthagen-Gonzalez, H., Bowers, J.S. & Damian, M.F. (2004) Age-of-acquisition effects in visual word recognition: evidence from expert vocabularies. *Cognition*, 93, B11–B26.

Stanners, R.F., Neisser, J.J., Hernon, W.P. & Hall, R. (1979) Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, 18, 399–412.

Stone, G.O., Vanhoy, M. & Van Orden, G.C. (1997) Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language*, 36, 337–359.

Stump, G. (1998) Inflection. In A. Spencer & A.M. Zwicky, editors, *The Handbook of Morphology*, pages 13–43. Oxford: Blackwell.

Stump, G. (2001) *Inflectional Morphology. A Theory of Paradigm Structure.* Cambridge, UK: Cambridge University Press.

Tabak, W., Schreuder, R. & Baayen, R.H. (2005) Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch. In S. Kepser & M. Reis, editors, *Linguistic evidence - Empirical, Theoretical, and Computational Perspectives*, pages 529–555. Berlin: Mouton de Gruyter.

Taft, M. (1979) Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7, 263–272.

Taft, M. (1991) *Reading and the mental lexicon.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Taft, M. (1994) Interactive-activation as a Framework for Understanding Morphological Processing. *Language and Cognitive Processes*, 9, 271–294.

Taft, M. (2004) Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, 57, 745–765.

Taft, M., Castles, A., Davis, C., Lazendic, G. & Nguyen-Hoan, M. (2008) Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language*, 58, 366–379.

Taft, M. & Forster, K. (1975) Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14, 630–647.

Taft, M., Hambly, G. & Kinoshita, S. (1986) Visual and Auditory Recognition of Prefixed Words. *The Quarterly Journal of Experimental Psychology*, 38, 351–366.

Toman, J. (1998) Word Syntax. In A. Spencer & A.M. Zwicky, editors, *The Handbook of Morphology*, pages 306–321. Oxford: Blackwell.

Tyler, L.K., Marslen-Wilson, W.D., Rentoul, J. & Hanney, P. (1988) Continuous and discontinuous access in spoken word-recognition: The role of derivational prefixes. *Journal of Memory and Language*, 27, 368–381.

Ullman, M.T. (2001) The Declarative/Procedural Model of Lexicon and Grammar. *Journal of Psycholinguistic Research*, 30, 37–66.

Ullman, M.T. (2004) Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231–270.

Ullman, M.T., Estabrooke, I.V., Steinhauer, K., Brovetto, C., Pancheva, R., Ozawa, K., Mordecai, K. & Maki, P. (2002) Sex differences in the neurocognition of language. *Brain and Language*, 83, 141–143.

van Casteren, M. (2006) *Mix.* Available from: http://www.mrc-cbu.cam.ac.uk/~maarten/Mix.htm. Accessed December 4, 2006.

Vannest, J., Bertram, R., Järvikivi, J. & Niemi, J. (2002) Counterintuitive Cross-Linguistic Differences: More Morphological Computation in English than in Finnish. *Journal of Psycholinguistic Research*, 31, 83–106.

Venables, W.N. & Ripley, B. (2002) *Modern Applied Statistics with S.* New York: Springer, 4th edition.

Vitevich, M.S. & Luce, P.A. (1998) When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*, 9, 325–329.

Vitevich, M.S. & Luce, P.A. (1999) Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374–408.

Wehberg, S., Vach, W., Bleses, D., Thomsen, P., Madsen, T.O. & Basbøll, H. (2007) Danish children's first words: Analysing longitudinal data based on monthly CDI parental reports. *First Language*, 27, 361–383.

Westbury, C. & Buchanan, L. (2002) The Probability of the Least Likely Non-Length-Controlled Bigram Affects Lexical Decision Reaction Times. *Brain and Language*, 81, 66–78.

Wurm, L.H. (1997) Auditory Processing of Prefixed English Words Is Both Continuous and Decompositional. *Journal of Memory and Language*, 37, 438–461.

Wurm, L.H. (2000) Auditory Processing of Polymorphemic Pseudowords. *Journal of Memory and Language*, 42, 255–271.

Wurm, L.H., Ernestus, M., Schreuder, R. & Baayen, R.H. (2006) Dynamics of the auditory comprehension of prefixed words. *The Mental Lexicon*, 1, 125–146.

Wurm, L.H. & Ross, S.E. (2001) Conditional Root Uniqueness Points: Psychological Validity and Perceptual Consequences. *Journal of Memory and Language*, 45, 39–57.

Yates, M., Friend, J. & Ploetz, D.M. (2008) The effect of phonological neighborhood density on eye movements during reading. *Cognition*, 107, 685–692.

Zevin, J. D. & Seidenberg, M. S. (2002) Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, 47, 1–29.

Ziegler, J., Petrova, A. & Ferrand, L. (2008) Feedback Consistency Effects in Visual and Auditory Word Recognition: Where Do We Stand After More Than a Decade? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 643–661.

Ziegler, J.C., Muneaux, M. & Grainger, J. (2003) Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, 48, 779–793.