Lexical Processing of Morphologically Complex Words

An information-theoretical perspective

ISBN: 978-90-76203-32-4
Cover design: Olga Goldovskaya
Printed and bound by Ponsen & Looijen bv, Wageningen

Lexical Processing of Morphologically Complex Words

An information-theoretical perspective

een wetenschappelijke proeve
op het gebied van de Letteren

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op donderdag 30 oktober 2008
om 15.30 uur precies

door

Victor Kuperman
geboren op 29 juli 1973 te Sint-Petersburg, Rusland

Promotor:	Prof. dr. R. Schreuder
Copromotores:	Prof. dr. R.H. Baayen (University of Alberta)
	Dr. M.T.C. Ernestus
Manuscriptcommissie:	Prof. dr. L.W.J. Boves Prof. dr. R. Kliegl (University of Potsdam) Prof. dr. H.J. Schriefers

Contents

Introduction

Chapter 1

One of the central questions of psycholinguistics is how people store words in their memory, and how they retrieve those words from memory when speaking, listening, reading or writing. This question is especially challenging for morphologically complex words, that is, words that are composed of two or more meaningful linguistic units, or morphemes (e.g., *un-reach-able*, *air-port*, or *talk-ed*). The main goal of this dissertation is to expand our knowledge about the role that morphological structure plays in speech production, as well as in reading. On the basis of experimental studies in languages with very different morphological systems, we aim to shed light on several areas of morphological inquiry that are either under-researched or controversial and to propose a new model of morphological processing. Specifically, we aim to address the following topics, which are hotly debated in current research on the role of morphology in lexical processing.

- Are (all) complex words stored in our memory along with their morphemes? The combinatorial dual-route model of Pinker (1999) claims that regular words (e.g., past-tense verb *talked*) do not necessarily have representations in our lexical memory (mental lexicon). Instead, the mental lexicon stores their morphemes, *talk* and *ed*. A deterministic rule is claimed to combine these morphemes into a complex word, while the meaning of that word is computed from the meanings of its morphemes. Irregular forms like *sang* cannot be rule-driven and hence are stored in memory. An alternative approach claims that our experience with using words and morphemes invariably leaves traces in memory, regardless of whether the words are regular or not. This approach suggests that morphological processing is best explained by probabilistic regularities between forms and meanings in language, rather than deterministic rules (cf., Hay & Baayen, 2005; Seidenberg & Gonnerman, 2000).

- In what order do morphemes and whole words become available in the production or comprehension of complex words? To illustrate this issue for visual word recognition, current psycholinguistic models differ widely in their predictions about what parts of morphological structure (full-forms, e.g., *dishwasher*, or morphological constituents, e.g., *dish* and *washer*) are activated during lexical processing of complex words, and in what order this activation proceeds. The class of connectionist models, such as the influential triangle model (Seidenberg & McClelland, 1989), make no explicit predictions as to the time-course of morphological effects. Sublexical and supralexical models advocate obligatory sequentiality in the activation of full-forms and morphological constituents. The former class of models posits that full-forms can only be accessed via morphological constituents (e.g., Pinker, 1999; Taft & Forster, 1975; Taft, 1979; Taft, 2004), and as a consequence, the properties of a word's morphological constituents are expected to influence processing times at an earlier stage than the properties of the word's full-form. The class of supralexical models makes the opposite claim that the activation of the full-form *precedes* the activation of constituents (e.g., Giraudo & Grainger, 2001). Yet another class of models, the dual route models, hypothesize that full-form based processing goes in parallel (simultaneously) with decomposition of complex words into their morphemes and the subsequent by-morpheme lexical access (cf., Allen & Badecker, 2002; Baayen & Schreuder, 1999; Baayen & Schreuder, 2000; Frauenfelder & Schreuder, 1992; Hay & Baayen, 2005; Laudanna & Burani, 1995; Schreuder & Baayen, 1995). In conclusion, the time-course of effects shown by morphemes and polymorphemic words on the lexical recognition of complex words is far from being a resolved issue. Yet it is important since such time-course may adjudicate between several competing models of morphological processing.

- Are words organized in the mental lexicon in hierarchical sets (i.e., paradigms) based on similarity of their morphemes in form or meaning, and if there is such a paradigmatic organization, how do those paradigms affect recognition or production of morphemes and complex words? Research of the last decade showed strong evidence that the size, the frequency-based characteristics and the internal structure of morphological families (i.e., paradigms of complex words sharing a morpheme, e.g., *dishcloth*, *dish soap*, *dish rack* or *talk*, *talked*, *talking*, or *happily*, *sadly*, *possibly*, etc.)

in codetermining the lexical processing of complex words (cf., De Jong, 2002; Krott, 2001; Moscoso del Prado Martín, 2003). Importantly, however, morphological families present only one of many possible organizations of complex words into paradigms, and the exploration of alternative hierarchical structures in the mental lexicon is still underway.

- What statistical, formal or semantic properties of morphemes and whole words can shift the balance between retrieving words from long-term memory as a whole versus online computing meanings of words from the constituent morphemes? The literature that assumes mental storage of both complex words and morphemes has proposed a wide range of factors that may bias lexical processing towards recognition of the whole word or towards parsing the complex word into its morphological constituents: These include orthographic properties of morphemes and whole words (e.g., length in characters, or transitional probabilities of n-grams, e.g., Andrews & Davis, 1999; Laudanna & Burani, 1995), their phonological and phonotactic properties (e.g., co-occurrence probabilities of n-phones and of other patterns across the morphemic boundary, e.g., Bertram, Pollatsek & Hyönä, 2004), their lexical properties (e.g., word formation type, homonymy or allomorphy, cf., Bertram, Schreuder & Baayen, 2000; Bertram, Laine, Karvinen, 1999; Järvikivi, Bertram & Niemi, 2006; Sereno & Jongman, 1997), and finally their distributional characteristics (e.g., Baayen, 1994; Hay, 2003). Importantly, many of these factors have been proposed on the basis of experiments that considered only a small number of predictors at a time (often experimenting on words differing in only one dimension). It is an open research question what the relative contributions of the proposed factors are to the processing costs of complex word recognition or production, when considered among many other factors.

- How does the lexical processing of morphologically complex words vary across languages, which show different morphological richness and different lexical-statistical patterns? For instance, the languages considered in the present dissertation – Dutch, English, Finnish and Serbian – differ in the size of their morphological paradigms, with Finnish being the morphologically richest language of the four (showing the largest derivational and compounding paradigms with up to 7000 members in compound families, and the richest inflectional system with 15 cases for nouns), and English being the language at the other extreme (with up to 200 family members in the compound families and only the remnants of an inflectional system for pronouns), and Serbian and Dutch being in between in terms of their lexical-statistical properties (cf., Moscoso del Prado Martín, 2003). While recent cross-linguistic studies begin to probe how language-specific morphological structure and its distributional characteristics modulate the role of lexical memory and that of morphological paradigms in recognition and production of complex words (cf., De Jong *et al.*, 2002; Lehtonen & Laine, 2003), this issue requires further investigation.

- Are morphemes also used in lexical processing when they are structurally embedded in larger morphological units (e.g., *wash-* and *-er* embedded in *washer*, which in turn is embedded in *dishwasher*)? Most experimental studies on morphological complexity considered words with only two morphemes (*sadly*). Yet words with three or more morphemes (e.g., *dishwasher*) are very common in the languages studied in this dissertation, accounting, for instance, for over 50% of the word types in Dutch and Finnish. Yet it is not self-evident that people are able to parse such multi-level morphological structures down to the very lower hierarchical levels. Hence, if observed, the effects of deeply embedded morphemes will expand our knowledge about how fine-grained morphological processing is.

The present dissertation addresses the outlined topics for speech production and visual comprehension of polymorphemic words in Dutch, English, Finnish and Serbian, by discussing earlier research, presenting novel experimental data and proposing a new probabilistic model of morphological processing. This model describes in a unified, principled way the insights of existing models and the complex patterns of morphological effects observed in our experimental studies. Our model uses the framework of information theory, which proved its worth for morphological processing in earlier studies by Kostić (1991; 1995) and Moscoso del Prado Martín, Kostić & Baayen (2004).

Outline of the Dissertation

Chapter 2 reports a study in speech production of Dutch compounds with linking elements, or interfixes. The focus of that chapter is on the probabilistic role of paradigms (morphological families) in codetermining the acoustic duration of the interfix, as realized in speech. Chapters 3-6 concentrate on visual recognition of morphologically complex words. Specifically, Chapter 3 combines eye-tracking and visual lexical decision as experimental techniques to investigate the time-course of morphological effects in the recognition of Dutch compounds. This chapter also offers specifications for a new model of morphological processing. The eye-tracking study in Chapter 4 has as its object Finnish trimorphemic compounds. Along with the time-course of morphological effects, it explores the role of derivational affixes as cues for the parsing of compounds into constituent morphemes, and formalizes our probabilistic model of morphological processing. Chapter 5 focuses on Dutch derived words and, using eye-tracking, examines further how the perceptual salience of affixes regulates the balance between storage and computation of complex words. The study in Chapter 6 considers the implications of our probabilistic model for the current body of research on the processing of inflectional paradigms in Serbian (e.g., Kostić, 1991), as well as for new experimental data on paradigmatic effects in the processing of English derived words. It specifically tackles the question of the paradigmatic organization of words in the mental lexicon. Chapter 7 summarizes the research presented in this dissertation and outlines topics for further investigation.

References

Allen, M. and Badecker, W. (2002). Inflectional regularity: Probing the nature of lexical representation in a cross-modal priming task. *Journal of Memory and Language*, 46:705–722.

Andrews, S., Miller, B., and Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, 16(1/2):285–311.

Baayen, R. H. (1994). Productivity in language production. *Language and Cognitive Processes*, 9:447–469.

Baayen, R. H. and Schreuder, R. (1999). War and peace: morphemes and full forms in a non-interactive activation parallel dual route model. *Brain and Language*, 68:27–32.

Baayen, R. H. and Schreuder, R. (2000). Towards a psycholinguistic computational model for morphological parsing. *Philosophical Transactions of the Royal Society (Series A: Mathematical, Physical and Engineering Sciences)*, 358:1–13.

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

De Jong, N. H. (2002). *Morphological Families in the Mental Lexicon*. MPI Series in Psycholinguistics. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and 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.

Frauenfelder, U. H. and Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In Booij, G. E. and Marle, J. v., editors, *Yearbook of Morphology 1991*, pages 165–183. Kluwer Academic Publishers, Dordrecht.

Giraudo, H. and Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review*, 8:127–131.

Hay, J. (2003). *Causes and Consequences of Word Structure*. Routledge, New York and London.

Hay, J. and Baayen, H. (2003). Phonotactics, parsing and productivity. *Italian Journal of Linguistics*, 1:99–130.

Hay, J. B. and Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, 9:342–348.

Järvikivi, J., Bertram, R., and Niemi, R. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and Cognitive Processes*, 21:394–431.

Kostić, A. (1991). Informational approach to processing inflected morphology: Standard data reconsidered. *Psychological Research*, 53(1):62–70.

Kostić, A. (1995). Informational load constraints on processing inflected morphology. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*. Lawrence Erlbaum Inc. Publishers, New Jersey.

Krott, A. (2001). *Analogy in morphology: The selection of linking elements in Dutch compounds*. MPI Series in Psycholinguistics. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Laudanna, A. and Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 345–364. Lawrence Erlbaum Associates, Hillsdale, N. J.

Lehtonen, M. and Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 6:213–225.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94:1–18.

Moscoso del Prado Martín, F. (2003). *Paradigmatic Structures in Morphological Processing: Computational and cross-linguistic experimental studies*. MPI Series in Psycholinguistics. Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Weidenfeld and Nicolson, London.

Schreuder, R. and Baayen, R. H. (1995). Modeling morphological processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 131–154. Lawrence Erlbaum, Hillsdale, New Jersey.

Seidenberg, M. S. and Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, 4(9):353–361.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96:523–568.

Sereno, J. and Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, 25:425–437.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition*, 7:263–272.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology*, 57A:745–765.

Taft, M. and Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14:638–647.

Morphological predictability and acoustic salience of interfixes in Dutch compounds

Chapter 2

This chapter is a slightly modified version of the paper published as Victor Kuperman, Mark Pluymaekers, Mirjam Ernestus, and R. Harald Baayen (2007). Morphological predictability and acoustic salience of interfixes in Dutch compounds. *Journal of the Acoustical Society of America* 121, 2261-2271.

Abstract

This chapter explored the effects of morphological predictability on the acoustic durations of interfixes in Dutch compounds. Two data sets were investigated: One for the interfix *-s-* (1155 tokens) and one for the interfix *-e(n)-* (742 tokens). Both datasets show that the more probable the interfix is given the compound and its constituents, the *longer* it is realized. These findings run counter to the predictions of information-theoretical approaches and can be resolved by the Paradigmatic Signal Enhancement Hypothesis. This hypothesis argues that whenever selection of an element from alternatives is probabilistic, the element's realization is predicted by the amount of paradigmatic support for the element: The most likely alternative in the paradigm of selection is realized with greater acoustic salience.

Introduction

One of the organizing principles of speech production is the trade-off between economy of articulatory effort and discriminability of the speech signal (Lindblom, 1990). Speech communication often takes place in noisy conditions. In order to ensure robust recognition of their acoustic output, speakers need to invest effort in articulation. Yet clear and careful articulation is costly and hence tends to be dispensed efficiently (cf., Aylett and Turk, 2004; Hunnicutt, 1985). As a consequence, elements with low information load (or high predictability) have shorter or otherwise less salient realizations than relatively more informative elements of an utterance.

The informational redundancy of speech elements is often operationalized in terms of the probability (relative frequency of occurrence) of a linguistic unit (e.g., phoneme, syllable, word, or phrase) in its context. High probability has been observed to correlate with acoustic reduction in a large variety of language domains: Syntactic, discourse-related, phonological and prosodic, and lexical (e.g., Aylett and Turk, 2004; Bard *et al.*. 2000; Fowler and Housum, 1987; Jurafsky *et al.*, 2001; Lieberman, 1963; McAllister *et al.*, 1994; Pluymaekers, Ernestus and Baayen, 2005a; Pluymaekers, Ernestus and Baayen, 2005b; Samuel and Troicki, 1998; Scarborough, 2004; Van Son and Pols, 2003; Van Son and Van Santen, 2005). The attested types of reduction include — apart from widely reported durational shortening of syllables and individual phonemes — deletion of phonemes and complete syllables (e.g., Ernestus, 2000; Johnson, 2004), decrease in spectral center of gravity (Van Son and Pols, 2003), decrease in mean amplitude (Shields and Balota, 1991), higher degree of centralization of vowels (Munson and Solomon, 2004), and lower degree of coarticulation (Scarborough, 2004). The informational redundancy associated with a particular unit is a juxtaposition of the unit's probabilities given all relevant contexts. For instance, a word can be predictable because it has a high frequency, but also because it is frequently used with the word that precedes it. Both factors diminish the word's informativeness and both are expected to correlate with durational shortening.

The information-theoretical framework developed by Shannon (1948) has been used to explain the association between acoustic salience and informational redundancy. The efficiency of information transmission is optimal if the information in the signal is distributed equally, or smoothly, per time unit (e.g., Aylett and Turk, 2004; Aylett and Turk, 2006). When an important element is transmitted for a longer time, the probability of losing this element to noise decreases and the probability of the element being recognized correctly increases. This theoretical paradigm views acoustic duration as a means of smoothing the amount of information in the signal over time.

The present paper shows how the information carried by morphological paradigmatic structure modulates acoustic duration. Previous research (cf., Hay, 2003; Losiewicz, 1992) reported morphological effects on the acoustic duration of affixes in complex words. A related line of research demonstrated the influence of lexical neighborhood density on durational characteristics and coarticulation in speech production (e.g., Munson and Solomon, 2004, Scarborough, 2004, Vitevitch, 2002). The morphological objects that are central in the present study are interfixes in Dutch noun-noun compounds. We will show that the acoustic duration of these interfixes creates an apparent paradox for the proposed information-theoretical principle of "less information, more reduction", which underlies the Smooth Signal Redundancy Hypothesis (Aylett and Turk, 2004), the Probabilistic Redundancy Hypothesis (Jurafsky *et al.*, 2001), and research on speech efficiency (e.g., Van Son and Pols, 2003). In our data, the more predictable the interfix is, the *longer* its articulation.

The distributional characteristics of the interfixes in Dutch compounds provide a clear-cut example of probabilistic, non-categorical morphological structure. Compounding is very productive in Dutch and is defined as the combination of two or more lexemes (or constituents) into a new lexeme (cf. Booij, 2002). In this paper we based our decisions of whether a given word is a compound and what its constituents are on the morphological parsing provided in the CELEX lexical database (Baayen, Piepenbrock and Gulikers, 1995). Compounds in Dutch can be realized with the interfix *-s-* (e.g., *oorlog-s-verklaring*, “announcement of war”), or with the interfix *-en-* (or its variant *-e-*) (e.g., *dier-en-arts* “veterinary”). Most compounds in Dutch, however, have no interfix (e.g., *oog-arts* “ophthalmologist”): For ease of exposition, we will henceforth refer to these latter words as compounds with the zero-interfix, or -∅-. In the frameworks that adopt deterministic rules, the distribution of interfixes in Dutch is enigmatic and inexplicable. Krott, Baayen and Schreuder (2001), however, have shown that the distribution of interfixes follows probabilistic principles defined over constituent families. The left (or right) constituent family of a compound is the set of all compounds which share the left (or right) constituent with this compound. For instance, the left constituent family of the compound *banknote* includes *bankbill*, *bankbook*, *bank-draft*, *bank-rate*, and *bankroll*. Krott, Baayen and Schreuder (2001), Krott *et al.* (2002) and Krott, Schreuder and Baayen (2002) show that the selection of the interfix is biased towards the interfix that is most commonly used with the given left constituent and, to a lesser extent, with the right constituent. Thus, besides having their own probability of occurrence, interfixes exhibit dependencies on larger morphological units both to the left and to the right. For this reason, interfixes serve as an appealing testing ground for studying the consequences of morphological predictability for acoustic realization.

The primary focus of the present study is the relationship between the predictability of the interfix given the morphological constituents of the compound, and its duration. We study the information-theoretical approach for two datasets with interfixed compounds and against the backdrop of multiple sources of redundancy, ranging from morphological to phonological and lexical information. Along the way, we replicate findings of laboratory studies of durational reduction for lively read-aloud speech.

Methodology

Materials

Acoustic materials were obtained from the Read Speech (or the “Library for the Blind”) component of the Spoken Dutch Corpus (Oostdijk, 2000). Within this corpus of approximately 800 hours of recorded speech, the Read Speech component comprises 100 hours of recordings of written texts read aloud by speakers of Northern Dutch from the Netherlands and Southern Dutch from the Flanders area of Belgium. In the preparation of the recordings, speakers were pre-screened for the quality of their voice and clarity of pronunciation, and texts were made available to the speakers beforehand for preparatory reading. We chose to concentrate on read speech primarily because of the low level of background noise of the recordings. Quality was essential, since Automatic Speech Recognition (henceforth, ASR) was used for obtaining the segmental durations (see below). It should be noted that since these texts of fiction were read for the collection of the Library for the Blind, the reading style was a lively, rather than monotonous recitation, especially in the dialogs, where readers often mimicked casual speech.

Two datasets of Dutch noun-noun compounds were compiled: One with tokens containing the interfix *-s-* and one with compounds containing the interfix *-e(n)-*. Tokens in which the interfix *-s-* was either preceded or followed by the phonemes

[s], [z] or [ʃ] were excluded from the dataset, since such an environment makes it difficult to reliably segment the interfix from its neighboring segments. The final dataset for the interfix *-s-* consisted of 1155 tokens. Similarly, tokens in which the second constituent begins with the segments [n] or [m] were taken out off the dataset of *-e(n)-* interfixes, resulting in a dataset of 742 tokens.

Measurements

Acoustic analysis of the selected tokens was performed using ASR technology. This was done for several reasons. First of all, the ASR technology allows to process a large volume of data in a relatively short time, which was important given the size of datasets used in this study. Moreover, it is possible to train an ASR device that bases its decisions purely on the characteristics of the acoustic signal, without reference to general linguistic knowledge. This is very difficult for human transcribers, who are bound to be influenced by expectations based on their knowledge of spelling, phonotactics, and so on (Cucchiarini, 1993). Second, ASR devices are perfectly consistent: Multiple analyses of the same acoustic signal always yield exactly the same result. Finally, the reliability of segmentations generated by an ASR system is equal to that of segmentations made by human transcribers (Vorstermans, Martens and Van Coile, 1996), provided that a phonemic transcription of the signal is available to the ASR algorithm.

For the present analysis, we utilized a Hidden Markov Model (HMM) speech recognizer. This recognizer was trained using the software package HTK (Young *et al.*, 2002), comprises 37 phone models representing the 36 phonemes of Dutch and silence, and uses for each model 3-state HMMs with 32 gaussians per state (Kessens and Strik, 2004). The HTK recognizer operates in two modes: If it is provided with the transcription of the speech recording, it determines segmental temporal boundaries; if no such transcription is provided, it identifies both the phonemes and the positions of their temporal boundaries. The accuracy of segmentation is higher in the transcription-based mode. The sample rate of the HTK is 10ms. The reliability of the ASR's segmentation with predefined transcriptions was established in a test in which the positions of phoneme boundaries placed by the ASR were compared to the positions of the same boundaries placed by a trained phonetician. The materials used for this test consisted of 189 words spoken in isolation. Comparison between the ASR-generated and manual segmentations revealed that, after post-processing, 81% of the automatic boundaries were placed within 20 milliseconds of the corresponding hand-coded boundaries. This level of accuracy is in accordance with international standards (Vorstermans *et al.*, 1996), and we considered it sufficient for present purposes.

Acoustic analysis proceeded as follows. First, the speech signal corresponding to the target compound was manually excised from its utterance context and parameterized using Mel Frequency Cepstral Coefficients. The parameterized signal was then supplied to a Viterbi segmentation algorithm, along with a phonemic transcription of the word. This transcription was taken from the CELEX lexical database. However, for words with the interfix *-e(n)-*, a cursory inspection of sound files established that many instances of this interfix were not realized as [ə] (the canonical pronunciation in CELEX), but rather as [ən]. An inspection of the sound files from the dataset with the interfix *-s-* revealed cases where the interfix was realized as [s] instead of the CELEX transcription [z] due to the regressive voice assimilation. Therefore, two trained phoneticians independently transcribed the realization of interfixes in both datasets. Initially, they disagreed on 10% of tokens from the *en*-dataset and 13% of tokens from the *s*-dataset. In both cases, they subsequently carried out a joint examination of the problematic tokens and came up with consensus transcriptions. The resulting transcriptions were provided to the segmentation algorithm, which estimated the boundaries of the phonemes in the acoustic signal. In this way, we obtained information about the durations of all segments for all words.

The acoustic duration of the whole interfix (henceforth, *InterfixDuration*) was taken as the main dependent variable in this study.

Morphological Variables

As shown in Krott *et al.* (2001), the more frequent an interfix is for the left constituent family of a compound, the more biased speakers are to use this interfix in that compound. The measures for this morphologically based bias will be at the center of our interest. They are defined as the ratio of the number of compounds where the left constituent is followed by *-s-*, *-e(n)-*, or *-∅-* respectively, and the total number of compounds with the given left constituent (henceforth, the left family size). To give an example, the Dutch noun *kandidaat* “candidate” appears as the left constituent in one compound with the interfix *-s-*, *kandidaat-s-examen* “bachelor’s examination”, in one compound with the interfix *-en-*, *kandidat-en-lijst* “list of candidates”, and in one compound without an interfix *kandidaat-stelling* “nomination”. The type-based bias of this left constituent family towards the interfix *-s-* is $1/(1+2) = 0.33$. The bias of the interfix *-e(n)-* has the value of $1/(1+2) = 0.33$ as well, and so does the bias of the zero-interfix. The measures of bias are labeled *TypeSBias*, *TypeEnBias* and *TypeZeroBias*.

Alternative, token-based, estimates of the bias are defined in terms of the frequencies of occurrence, rather than the type count of the compounds. The performance of token-based measures is consistently worse in our models than that of the type-based ones. Therefore, the token-based measures are not reported here. Furthermore, we only consider left constituent families, since the effect of the right bias is reported as either weak or absent (Krott, Schreuder and Baayen, 2002; Krott *et al.*, 2004).

The predictivity of constituent families for the duration of the interfix may extend beyond the bias measures, which only estimate the ratio of variants in the constituent family, without taking the magnitude (size, frequency, or information load) of the constituent family into account. However, these magnitudes are expected to exhibit effects in our analysis, since they repeatedly emerged as significant predictors in both the comprehension and production of Dutch compounds (e.g., Bien, Levelt and Baayen, 2005; De Jong *et al.*, 2002; Krott *et al.*, 2004). To estimate the magnitude of constituent families, we incorporate in our study position-specific measures of entropy proposed by Moscoso del Prado Martín, Kostić and Baayen (2004). These measures employ the concept of Shannon's entropy (Shannon 1948), which estimates the average amount of information in a system on the basis of the probability distribution of the members of that system. The probability of each member (p_{sys}) is approximated as the frequency of that member divided by the sum of the frequencies of all members. The entropy of a system with *n* members is then the negative weighted sum of log-transformed (base 2) probabilities of individual members:

$$H = -\sum_{i=1}^{n} p_{sys} * \log_2 p_{sys}$$

Note that the entropy increases when the number of paradigm members is high (i.e. family size is large) and/or when the members are equiprobable.

Let us consider the positional entropy measure of the left constituent family of the Dutch noun *kandidaatstelling*. This family consists of three members: *kandidaatsexamen* has a lemma frequency of 22, *kandidaatstelling* has a lemma frequency of 15, and *kandidatenlijst* has a lemma frequency of 19 in the CELEX lexical database, which is based on a corpus of 42 million word forms. The cumulative frequency of this family is $22+15+19 = 56$, and the relative frequencies of these three family members are $22/56 = 0.39$ for *kandidaatsexamen*, $15/56 = 0.27$ for *kandidaatstelling* and $19/56 = 0.34$ for *kandidatenlijst*. The left positional entropy of this constituent family therefore equals $-(0.39 * \log_2 0.39 + 0.27 * \log_2 0.27 + 0.34 * \log_2 0.34) = 1.57$ bit.

We consider the positional entropy measures for both the left and the right constituent families, henceforth *LeftPositionalEntropy* and *RightPositionalEntropy* as potential predictors of the acoustic duration of the interfix. The informativeness of the right constituent family is meaningful as a measure of the cost of planning the right constituent: Planning upcoming elements with a low information load has been shown to predict reduction in the fine phonetic detail of the currently produced elements (Pluymaekers *et al.*, 2005a).

Other Variables

Since acoustic duration is known to depend on a wide range of factors, we used stepwise multiple regression to bring these factors under statistical control. Two sets of factors were considered: Lexical frequency-based probabilities, and phonetic, phonological and sociolinguistic variables.

Probabilistic factors

Phrasal level: A higher likelihood of a word given its neighboring words has been shown to correlate with vowel reduction, segmental deletion, and durational shortening (Bell *et al.*, 2003; Jurafsky *et al.*, 2001; Pluymaekers *et al.*, 2005a). To quantify this likelihood, for each compound token in our data we calculated its mutual information with the preceding and the following word (*BackMutualInfo, FwdMutualInfo*) by using the following equation (X and Y denote either the previous word and the compound, or they denote the compound and the following word; XY denotes the combination of the two words):

$$MI(X;Y) = -log \frac{\text{Frequency(XY)}}{\text{Frequency(X)} * \text{Frequency(Y)}}$$

The measures were computed on the basis of the Spoken Dutch Corpus, which contains 9 million word tokens. All frequency measures were (natural) log-transformed. Obviously, the values could not be computed for the instances where the target word was utterance-initial or utterance-final, respectively.

For those words for which mutual information with the preceding or the following word could be computed, we checked whether it was a significant predictor of the duration of the interfix over and beyond other factors. Neither *BackMutualInfo* nor *FwdMutualInfo* reached significance in our datasets. This result may originate in the properties of the datasets which comprise relatively low-frequency compounds. Obviously, these low-frequency compounds have even lower frequencies of cooccurrence with their neighboring words. For instance, for the *s*-dataset the average frequency of cooccurrence of the compounds with the preceding word is a mere 1.63 ($SD = 0.77$), and with the following word a mere 1.20 ($SD = 0.30$). Another explanation may be that effects of contextual predictability do not extend to phonemes in the middle of long compounds. They may only emerge for segments at word boundaries (e.g., Jurafsky *et al.*, 2001; Pluymaekers *et al.*, 2005a).

Word level: The lexical frequency of a word is known to codetermine articulation and comprehension (e.g., Jurafsky *et al.*, 2001; Pluymaekers *et al.*, 2005a; Scarborough, Cortese and Scarborough, 1977; Zipf, 1929). Moreover, previous research has shown that whole word frequency robustly affects production and comprehension of compounds even in the low-frequency range (cf. e.g., Bertram and Hyönä, 2003, Bien *et al.*, 2005). Therefore we include the natural log-transformed compound frequency (*WordFrequency*) as a control variable in the analyses. Together with the measure of the bias and the left positional entropy, this variable forms a cluster of predictors that capture different aspects of the same phenomenon. The measure of the bias estimates the *proportion* of the positional family of compounds that supports the interfix. The corresponding entropy estimates the number and average information load of the members in this family, i.e., it gauges the reliability of the knowledge base for the bias. Finally, a high compound frequency quantifies the evidence for the cooccurrence of the left and right constituents with the interfix. We expect these variables to behave similarly in predicting the durational characteristics of the interfix.

Segmental level: Another dimension of predictability for segmental duration is the amount of lexical information in the individual segment given the preceding fragment of the word (i.e., given the "word onset"). Following Van Son and Pols (2003), we define an information-theoretic measure that quantifies segmental lexical information (*TokenSegmentalInfo*):

$$I_L = -\log_2 \frac{\text{Frequency([word onset] + target segment)}}{\text{Frequency([word onset] + any segment)}}$$

Van Son and Pols (2003) interpret this measure as estimating the segment's incremental contribution to word recognition. The occurrence of a segment that is improbable given the preceding fragment of the word limits the cohort of matching words substantially and thus facilitates recognition. To give an example, the amount of lexical information of the segment [s] given the preceding English word fragment [kaʊ] is calculated as the negative log-transformed ratio of the cumulative frequency of words that begin with the string [kaʊs] (e.g., *cows, cowskin, cowslip, cowslips*) and the cumulative frequency of the words that begin with the string [kaʊ] plus any segment (e.g., *cows, cowpat, cowshed, cowskin, cowslip, cowslips, etc.*). In the present study, segmental lexical information measures are based on the frequencies of single words, such as made available in CELEX, and do not account for combinations of words, even if those may acoustically be valid matches for the phonetic string. For instance, the combination *cow stopped* is not included in the calculation of the lexical information for the segment [s] in the string [kaʊs].

A positive correlation of this token-based segmental lexical information and segmental duration was reported in Van Son and Pols (2003) for different classes of phonemes grouped by manner of articulation: For read speech, the r-values of correlations that reached significance ranged between 0.11 and 0.18 (55811 df). If segmental lexical information indeed modulates fine phonetic detail, it is a potential predictor of the duration of the interfix.

To this token-based measure of segmental lexical information (*TokenSegmentalInfo*), we add a type-based measure, *TypeSegmentalInfo*, which is based on the *number* of words matching the relevant strings, rather than their cumulated frequencies:

$$S_L = -\log_2 \frac{\text{Number([word onset] + target segment)}}{\text{Number([word onset] + any segment)}}$$

We validated both the token-based and the type-based measures of segmental lexical information against our own dataset to establish how the performance of the type-based estimate S_L compares with that of the token-based measure I_L. Our approach differs from that of Van Son and Pols (2003) in that it considers the divergence of phonemes from their mean durations, rather than the raw durations of these phonemes. Different phonemes, even those that share manner of articulation, intrinsically differ in their durations. Therefore, pooling the durations of large classes of phonemes introduces unnecessary noise in the correlation analyses. We gauged the divergence of each instantiation of every phoneme from the mean duration of this phoneme and tested whether this divergence can be explained by the amount of lexical information carried by the phoneme. Our survey is based on *all* segments in the *s*-dataset and in the compounds of the *en*-dataset in which the interfix is realized as [ə].

We collected the data on mean durations from the Read Text component of the IFA corpus, a hand-aligned phonemically segmented speech database of

Dutch (Van Son, Binnenpoorte, Van den Heuvel, Pols, 2001). We log-transformed the individual durations and computed the means and standard deviations of all tokens of each phoneme. Then, moving phoneme by phoneme through our compound dataset we calculated the z-score for each phoneme, that is, the difference between its actual log-transformed duration and its mean log duration, in units of standard deviation from the mean. The correlation between the observed durational difference and the corresponding amount of type-based segmental lexical information yields an *r*-value of 0.06 ($t(17694) = 7.41, p < 0.0001$). This order of magnitude is comparable with the results that Van Son and Pols (2003) obtained for the token-based measure of lexical information. The observed correlation is a rough estimate of the baseline effect that segmental lexical information may have on acoustic duration. The correlation is highly significant but the correlation coefficient is quite small. This is expected, given the multitude of phonetic, phonological, sociolinguistic and probabilistic factors that determine acoustic duration in speech production that are not taken into account here. As the type-based measure is predictive for durations of segments across the dataset, we decided to include it in our analyses of the interfix durations. Thus, we take as control variable the value of *TypeSegmentalInfo* for the (first) segment of the interfix.

Importantly, the durations show a weaker correlation with the token-based segmental lexical information, proposed by Van Son and Pols (2003) ($r = 0.03, t(17694) = 4.25, p < 0.0001$), than for its type-based counterpart ($r = 0.06$). This measure also performs worse in the models reported below. Since the token- and type-based measures are highly correlated, we incorporated only *TypeSegmentalInfo* in our analysis.

Phonetic, phonological and sociolinguistic variables

Speech rate is an obvious predictor of acoustic duration (e.g., Crystal and House, 1990; Fosler-Lussier and Morgan, 1999; Pluymaekers *et al.*, 2005a). Two different measures estimating speech rate were included as control variables. First, we defined an utterance-based rate of speech, *SpeechRate*, as the number of syllables in the utterance divided by the acoustic duration of the utterance. Utterance is defined here as the longest stretch of speech containing the compound and not containing an audible pause.

Second, we defined a more local speech rate for the interfix *-s-*. In the *s*-dataset, the interfix *-s-* always belongs to the coda of the preceding syllable. We measured the average segmental duration in the interfix-carrying syllable minus the *-s-* interfix, and considered it as an estimate of the local speed of articulation in the part of the syllable that precedes the interfix *-s-*, henceforth *SyllableSpeed*. The syllable from which the final segment [s] was subtracted is structurally complete, with an onset, a vowel and (in 83% of tokens) a coda of one or more consonants. Note that for words with the interfix *-e(n)-* this measure of local speech rate is not meaningful. It would subtract the complete rhyme of the relevant syllable, leaving only the onset, the duration of which is above all determined by the number and types of its consonants.

Nooteboom (1972) observed that segments are shorter the greater the number of syllables or segments in the word. We therefore considered the total number of segments in the word, *NumberSegments*, and the number of segments following the interfix, *AfterSegments*.

We also took into account the sex, age and language variety of the speaker (cf., Keune, Ernestus, Van Hout and Baayen, 2005). The binary variable *SpeakerLanguage* encodes the speaker's variant as Southern Dutch or Northern Dutch. If the information about age was missing, we filled in the average age of our speakers' population.

Prosody may affect the duration of segments as well. For instance, words at the beginning and the end of utterances show articulatory strengthening (e.g., Bell *et al.*, 2003; Cambier-Langeveld, 2000; Fougeron and Keating, 1997). To control for the word's position in the utterance, we coded each token with two binary variables *UtteranceInitial* and *UtteranceFinal*.

Furthermore, stressed syllables are pronounced longer than unstressed ones (e.g., Ladefoged, 1982). We coded each compound with the interfix *-s-* for whether its interfix-containing syllable carries a (primary or secondary) stress (the binary variable *Stressed*).

The interfix *-e(n)-* is never stressed. The common stress pattern for compounds with the interfix *-e(n)-* is for the primary stress to fall on the syllable immediately preceding the interfix-containing syllable, and the secondary stress on the syllable immediately following the interfix-containing syllable: The insertion of *-e(n)-* prevents a stress clash between the two constituents. The rhythmic structure of compounds has been proposed as a factor codetermining the selection of the interfix, in addition to lexical constituent families and several other factors (Neijt *et al.*, 2002). To test the acoustic consequences of the rhythmic pattern, we coded each compound in the *en*-dataset as to whether the interfix syllable intervenes between two immediately adjacent stressed syllables (the binary variable *Clash*).

Compounds with the interfix *-e(n)-* were coded for the presence or absence of [n] in the acoustic realization of the interfix (*NPresent*), as established by two phoneticians (see section Methodology). Similarly, compounds with the interfix *-s-* were coded for whether the interfix was realized as [z], variable *PhonemeZ*.

Finally, the immediate phonetic environment can make a segment more or less prone to reduction. Unstressed vowels in Dutch tend to lengthen before oral stops (cf., Waals, 1999). Therefore, each compound in the dataset with the *-e(n)-* interfix was coded for the manner of articulation of the following segment (binary variable *FollowedbyStop*).

Results

The interfix *-s-*

The dataset for the interfix *-s-* included 1155 tokens. The number of different word types was 680, and their token frequencies followed a Zipfian distribution ranging from 1 to 19. We fitted a stepwise multiple regression model with the acoustic duration of the interfix as the dependent variable. The values of this variable were (natural) log-transformed to remove skewness of the distribution. The resulting variable *InterfixDuration* has a mean of 4.37 of log units of duration ($SD = 0.35$). The log-transformation in this model and the models reported below was applied purely for statistical reasons, such as reducing the likelihood that the estimates of the coefficients are distorted by atypically influential outliers. The coefficients of the regression models that are presented here in log units of duration can easily be converted back into milliseconds by applying the exponential function e^F to the fitted values (F) of the model.

We identified 21 data points that fell outside the range of -2.5 to 2.5 units of SD of the residual error, or had Cook's distances exceeding 0.2. These outliers were removed from the dataset and the model was refitted. Below we only report variables that reached significance in the final model.

The strength of the bias for the *-s-* interfix, *TypeSBias*, emerged as a main effect with a positive slope: Surprisingly, the duration of *-s-* was longer for compounds with a greater bias for this interfix [$\hat{\beta} = 0.35, t(1125) = 5.20, p < 0.0001$], with a 33 ms lengthening of duration between the extreme values of the predictor. A positive correlation with duration was present for the predictor *RightPositionalEntropy* as well [$\hat{\beta} = 0.07, t(1125) = 4.10, p < 0.0001$], indicating that the duration of the interfix increases with the informational complexity of the right constituent (with a 47 ms difference between compounds with the maximum right positional entropy and those with the minimal one). These main effects were modulated by an interaction between *TypeSBias* and *RightPositionalEntropy* [$\hat{\beta} = -0.07, t(1125) = -3.67, p = 0.0003$]. Inspection of conditioning plots revealed that the influence of the bias measure was greater when the value of the right positional entropy was low. In addition, *WordFrequency* had an unexpected positive slope that just failed to reach significance: [$\hat{\beta} = 0.01, t(1125) = 1.95, p = 0.0510$]. We found no effect of the *LeftPositionalEntropy*.

Importantly, the lexical segmental information of the interfix was predictive in the expected direction: Segments conveying more information tended to be longer [*TypeSegmentalInfo*: $\hat{\beta} = 0.12, t(1125) = 3.86, p < 0.0001$], with a 75 ms lengthening of acoustic duration when comparing extreme values of *TypeSegmentalInfo*.

Among the phonological and phonetic variables, the measure of the speech rate also demonstrated the expected behavior. The greater the local speed of articulation, the shorter the realization of this interfix [*SyllableSpeed*: $\hat{\beta} = -0.51, t(1125) = -5.27, p < 0.0001$]. Whether the interfix-carrying syllable was stressed was a significant predictor as well, with stress predicting durational shortening of the interfix [*Stressed*: $\hat{\beta} = -0.09, t(1125) = -3.96, p < 0.0001$]. Finally, interfixes realized as [z] were shorter than those realized as [s], as expected given the findings by, for instance, Slis and Cohen (1969) [*PhonemeZ*: $\hat{\beta} = -0.16, t(1125) = -3.17, p = 0.0016$].

All significant predictors were tested for possible non-linearities; none reached significance. The bootstrap validated R^2 of the model was 0.104. The unique contribution of the morpholexical factors *TypeSBias*, *PositionalEntropyRight*, and *WordFrequency* to the explained variance over and above the other predictors was 2.0%, as indicated by the drop in R^2 when these variables were removed from the model.

Discussion

Three related morpholexical variables emerge as significant predictors of the duration of the interfix: *TypeSBias, RightPositionalEntropy* and (marginally) *WordFrequency*. The positive correlations of *TypeSBias* and *WordFrequency* with the duration of the interfix lead to the paradoxical conclusion that a greater likelihood for a linguistic unit may lead to a longer acoustic realization of that unit, contradicting the information-theoretical approach to the distribution of acoustic duration. We will address this issue in the General Discussion.

The interaction of the right positional entropy with the bias hints at planning processes at work. According to Pluymaekers *et al.* (2005b), the planning of upcoming linguistic elements may interfere with the planning and production of preceding elements. We interpret the right positional entropy measure as tapping into the costs of planning the right constituent. The observed interaction indicates that the bias allows greater durational lengthening of the interfix when planning the next constituent is easy.

In accordance with previous reports (e.g., Van Son and Pols, 2003), a high amount of lexical information carried by an individual segment (*TypeSegmentalInfo*) predicts the acoustic lengthening of this segment. In other words, segments with a larger contribution to the word's discriminability are produced with increased articulatory effort, and hence prolonged duration. This highlights the paradox with which we are confronted: Conventional measures, such as the segmental lexical information, behave as expected, while measures for the likelihood of the interfix exhibit exceptional behavior.

The effects of *TypeSegmentalInfo* and of *TypeSBias* may appear to contradict each other: For the same segment [s], the former variable predicts acoustic reduction, while the higher bias correlates with acoustic lengthening. Yet the two variables operate independently on different levels: The level of morphological word structure for the bias, and the segmental level for the lexical information. In the model, their (opposite) effects are simply additive.

The position of the compound in the utterance did not affect the durational characteristics of the interfix significantly, which is in line with observations by Cambier-Langeveld (2000). Cambier-Langeveld argues that final lengthening in Dutch only applies to the last syllable in the word or, if the vowel in this last syllable is [ə], to the penultimate syllable. Thus, the interfix lies beyond the scope of this effect. Similarly, the interfix emerges as outside the domain of influence of initial lengthening.

Segments are typically longer in a stressed syllable. This may have gone hand in hand with compensatory shortening of the duration of the following *-s-*. Compensatory reduction of the *-s-* in the coda of a stressed syllable may therefore provide an explanation for the observed effect of *Stressed*. Alternatively, acoustic reduction of the interfix may have arisen from the fact that stress on the syllable preceding the interfix *-s-* correlates with a higher local speech rate, which we calculated as the number of segments in the syllable (minus -s-) divided by the total duration of the syllable (minus -s-). This finding may appear counterintuitive, but it derives from the following observation. It is true that stressed syllables in our dataset have longer realizations than unstressed ones [two-tailed t-test: t(1097) = 30.0, p $<$ 0.0001], but more importantly, they consist of more segments [two-tailed t-test: t(1146) = 22, p $<$ 0.0001]. The net effect is the greater speech rate at stressed syllables. To test whether the latter finding is idiosyncratic to our dataset, we computed the number of segments for each syllable in Dutch monomorphemic words using CELEX phonological transcriptions. Again, we found that stressed syllables contained more segments than unstressed ones (2.76 vs. 2.17 segments per syllable, two-tailed t-test: t(192546) = 208.8, p $<$ 0.0001). This difference retained significance when the counts were corrected for ambisyllabicity. We conclude that a higher local speech rate may have contributed to the shortening of -s-interfixes that follow stressed syllables.

The interfix *-e(n)-*

The *en*-dataset contained 742 tokens of compounds. The number of different word types equalled 305, and the Zipfian distribution of tokens per type ranged from 1 to 74. We log-transformed the acoustic durations of the interfixes, which then had a mean of 4.065 log units of duration ($SD = 0.420$). We fitted a stepwise multiple regression model to these durations. This time, 19 data points fell outside the range of -2.5 to 2.5 units of SD of the residual error or had Cook's distances exceeding 0.2. These outliers were removed from the dataset, and the model was refitted. Only predictors that reached significance are reported.

The morpholexical predictors performed as follows: A higher bias for the interfix *-e(n)-*, *TypeEnBias*, correlated with longer interfixes: [$\hat{\beta} = 0.14, t(716) = 5.39, p < 0.0001$], with an 11 ms lengthening of acoustic duration for the most supported interfixes as compared to least supported ones . The positional entropy of the right constituent family also had a positive main effect [$\hat{\beta} = 0.08, t(716) = 4.56, p < 0.0001$], with a 42 ms difference between the extreme values of the positional entropy. The interaction of these two variables was not significant ($p > 0.4$). *LeftPositionalEntropy* and *WordFrequency* did not reach significance either ($p > 0.1$).

As in the model for the interfix *-s-*, a higher amount of lexical information, as attested by *TypeSegmentalInfo* for the first segment of the interfix, correlated with longer articulation [$\hat{\beta} = 0.07, t(716) = 3.09, p = 0.002$], with a 29 ms difference between the extreme values of *TypeSegmentalInfo*. This effect is again in line with predictions of the information-theoretical approach.

The interfixes of 226 tokens (29%) in the dataset were realized as [ən], while 561 tokens were pronounced with [ə]. As expected, the presence of [n] in the interfix implied a substantial increase in the total duration of the interfix. The factor *NPresent* was the most influential predictor [$\hat{\beta} = 0.71, t(716) = 37.80, p < 0.0001$], and its unique contribution to the explained variance of this duration was 55%.

Two phonetic factors contributed to the duration of the interfix. Unsurprisingly, the interfix was shorter when the utterance-based speech rate was higher [*SpeechRate*: $\hat{\beta} = -0.04, t(716) = -4.17, p < 0.0001$]. Factor *FollowedbyStop* also had an effect [$\hat{\beta} = 0.23, t(716) = 13.10, p < 0.0001$], which supports the observation by Waals (1999) that an unstressed vowel is pronounced longer before oral stops. It is noteworthy that Waals' observation, which was made under thoroughly controlled laboratory conditions, is replicated here in more natural read aloud speech.

All significant predictors in the model were checked for non-linearities, none of which reached significance. The bootstrap validated R^2 value for the model was 0.72. The unique contribution of the morphological predictors *TypeEnBias* and *RightPositionalEntropy* to the variance explained by the model was 2.3%, as indicated by the drop in R^2 after the removal of these variables from the model. This contribution is close to that provided by the morpholexical predictors in the *s*-dataset (2.0%).

Discussion

The analysis of the *en*-dataset replicates the unexpected direction of the influence of the morphologically-determined redundancy that we reported for the dataset with the interfix *-s-*: We found again that higher values for the bias estimates correlate with a longer duration of the interfix. We will return to this role of the bias in the General Discussion.

The positive simple main effect of the right positional entropy supports the hypothesis of continuous planning of articulation, according to which the planning complexity of upcoming elements may modulate acoustic characteristics of preceding elements.

Given the dominant contribution of the variable *NPresent* to the explained variance, we set out to establish what factors affected the selection of the variant [ən] versus [ə]. The interfix *-e(n)-* is spelled as either *-e-* or *-en-*, depending on orthographic rules. Compounds spelled just with *-e-* are unlikely to be pronounced with [ən]. The subset of compounds spelled with *-en-* contains 653 tokens. We fitted a logistic regression model that predicted the log odds of the selection of [ən] versus [ə] in this subset. The model uses the binomial link function and considers the presence of [n] in the realization of the interfix as a success, and its absence as a failure. The results demonstrate no effect of *TypeEnBias* on the selection of the phonetic variant ($p > 0.5$). Apparently the realization of an extra phoneme in the interfix is independent of the morphological likelihood of the interfix. The presence of [n] was more likely when *WordFrequency* was high [$\hat{\beta} = 0.63, p < 0.0001$], *RightPositionalEntropy* was high [$\hat{\beta} = 2.11, p < 0.0001$], the speaker's language was Southern Dutch [$\hat{\beta} = 1.37, p < 0.0001$], the number of segments after the interfix, *AfterSegments*, was high [$\hat{\beta} = 2.06, p < 0.0001$], and a stress clash was attenuated [$\hat{\beta} = 4.19, p < 0.001$]. The likelihood of [n] was lower when *LeftPositionalEntropy* was high [$\hat{\beta} = -0.60, p < 0.0001$].

In a second supplementary analysis, we investigated whether morpholexical factors are better predictors for acoustic duration if we consider the duration of [ə] as the dependent variable, rather than the duration of the whole interfix. In such a model, we expect the presence of [n] to exercise less influence and the morpholexical predictors to have greater explanatory value than in the model for the duration of the interfix as a whole. We fitted a stepwise multiple regression model to the data with the (natural) log-transformed acoustic duration of the phoneme [ə] in the interfix as the dependent variable. After removal of 25 outliers, the model was refitted against the remaining 717 datapoints.

In line with our expectations, we observe a decrease in the predictive power of *NPresent* to only 15% of the explained variance, while the share of morphological variables *TypeEnBias* and *RightPositionalEntropy*, which retain significance as predictors of acoustic lenghtening, increases to 4.3% of the explained variance. We conclude that morphological structure codetermines the acoustic characteristics of the interfix *-e(n)-* over and beyond major phonological and phonetic predictors[1].

[1]If a compound is spelled with *-e(n)-*, it can be realized as [ən] or [ə] in speech. We have shown that a higher word frequency favors the presence of [n] in the realization of the interfix. Might it be the case that the realization of the interfix as [ə] is longer in a compound that is more often realized with [ən]? To check this possibility, we computed the percentage of tokens realized as [ən] for each *-e(n)-*compound. This percentage was not a significant predictor of acoustic duration of [ə] (p $>$ 0.05). Thus we rule out an impact of the relative frequency of [ən]-realization (more probable in read speech) on [ə]-realization (more probable in spontaneous speech).

General Discussion

According to the information-theoretical approach to acoustic salience developed in the last decade, a higher likelihood of a linguistic unit is correlated with more acoustic reduction. The main finding of the present study is that the effect of morphologically-determined probability on the duration of interfixes in Dutch compounds runs counter to this prediction. This pattern of results is especially puzzling, since our data also provide evidence *in favor of* the information-theoretical approach in the form of an effect of segmental lexical information. Thus, we do find that a higher probability of a segment given the preceding word fragment leads to more acoustic reduction.

The speakers in the Spoken Dutch Corpus read the compounds and thus received unambiguous visual information about the correct interfix. It is therefore remarkable that we nevertheless observed effects of morpholexical factors on the planning and implementation of speech production. We note, however, that the bias of the interfix as determined by the left constituent family is known to predict the speed of reading comprehension of novel and existing compounds (Krott, Hagoort and Baayen, 2004). We therefore expect the acoustic consequences of the bias to have a larger scope when visual cues to the appropriate morphemes are absent, as in spontaneous speech genres.

What may be the solution for the problem that the present data appear to pose for the information-theoretical framework? One explanation might be that morphological information has a fundamentally different status from other types of linguistic information, and is typically associated with careful articulation. However, this line of reasoning is refuted by research on prefixes and suffixes in English (e.g., Hay, 2003) and Dutch (e.g., Pluymaekers *et al.*, 2005a, Pluymaekers *et al.*, 2005b).

Another solution might refer to the fact that interfixes are homophonous with plural markers in Dutch (cf., *boek-en* "books" and the compound *boek-en-kast* "bookshelf"). The frequency of the plural word forms might codetermine the duration of the interfix and be confounded with the bias. This explanation, however, can be discarded on the following grounds. First, there was no consistency in the correlation between the frequency of plural nouns and the bias of the interfix across datasets. For the *-s*-dataset the correlation was positive [$r = 0.12, t(1154) = 4.24, p < 0.0001$], while for the *-en*-dataset it was negative [$r = -0.28, t(740) = -8.15, p < 0.0001$]. Second, the frequency of the plural homophonous forms did not reach significance when included as a covariate in the regression models for both datasets. Finally, previous work on German compounds by Koester, Gunter, Wagner and Friederici (2004) has shown that plural suffixes and interfixes may not be perfectly homophonous in terms of systematic fine phonetic detail: Compound constituents followed by an interfix are shorter and have a higher pitch than their stand-alone plural counterparts.

The hypothesis that we would like to offer as a solution for the present paradox is that fine phonetic detail in speech is governed by two orthogonal dimensions, a syntagmatic dimension and a paradigmatic dimension. The information-theoretical approach that underlies the Smooth Signal Redundancy Hypothesis (Aylett and Turk, 2004) and the Probabilistic Reduction Hypothesis (Jurafsky *et al.*, 2001), as well as research on speech efficiency (Van Son and Pols, 2003; Van Son and Van Santen, 2005), views information from the syntagmatic perspective by considering the probability of a linguistic unit in its phonetic, lexical, or syntactic context. These syntagmatic relationships are inherently sequential and govern the temporal distribution of information in the speech stream. For instance, the extent to which a segment contributes to the identification of the word *given the preceding word fragment* (Van Son and Pols, 2003) is a syntagmatic measure that is positively correlated with duration: The greater the contribution of the segment, the longer its acoustic implementation.

The syntagmatic measures proceed upon the premise that there is no (probabilistic) variation in the elements forming the word or the syntactic clause to be realized by the speaker. When the speaker wants to express the concept *book*, there is no doubt that the element following [bʊ] is [k].

However, the identity of the elements is not always known with such certainty: The interfix in Dutch compounds is one such example. We label such elements "pockets of indeterminacy". Paradigmatic relations, here defined over constituent families, provide the probabilistic basis for resolving this indeterminacy. The bias measures quantify the extent of support provided by paradigmatics for the different interfixes available for selection: A greater support increases the likelihood of a given interfix. Our experimental results indicate that such a greater likelihood is paired with a longer acoustic realization. Moreover, we have shown that a higher frequency of a compound correlates with an increased chance of a more salient realization of the interfix *-e(n)-* as [ən], rather than [ə].

Whereas the syntagmatic dynamics of lexical disambiguation are intrinsically temporal, paradigmatic inference is a-temporal in nature. In the a-temporal domain of paradigmatic inference for positions of choice, a greater probability implies a broader empirical basis for selection of a given alternative, and comes with increased acoustic duration.

Importantly, paradigms as a source of support for alternatives for selection are not restricted to morphological structure: We consider paradigms in a general Saussurean sense, as sets of linguistic elements over which the operation of selection is defined (de Saussure, 1966).

The amount of evidence for the alternatives apparently determines the confidence with which an interfix is selected. That a lack of confidence may lead to a decrease in acoustic duration may be illustrated by an analogy: When producing case endings of German nouns, non-native speakers of German may hush up their realizations if they have doubts about the appropriate morpheme, but articulate the endings carefully and clearly if they are certain about which ending to choose. This example serves as an analogy only, and there is no implication that speakers make deliberate, conscious choices based on the morphological bias. The support measured as the bias is rather an estimate of the "naturalness" of the association between the available interfixes and the constituents of the compound.

Our hypothesis that paradigmatic inference for pockets of indeterminacy leads to longer (or otherwise more salient) realizations, henceforth the Paradigmatic Signal Enhancement Hypothesis, offers straightforward, testable predictions at various levels of linguistic structure. First consider the level of morphology. It is well known that English irregular verbs cluster into sets according to the kind of vocalic alternation that they exhibit in the past tense form (*keep/kept*, *run/ran*). The Paradigmatic Signal Enhancement Hypothesis predicts that a past-tense vowel — a pocket of indeterminacy — is realized with increased acoustic salience when the vocalic alternation is supported by a larger set of irregular verbs. Effects of paradigmatic gangs might even be found for the vowels of regular verbs (Albright and Hayes, 2003).

At the interface of morphology and phonology, we call attention to the phenomenon of final devoicing. In German and Dutch, a stem-final obstruent may alternate between voiced and voiceless, compare Dutch [hɔnt] *hond* ('dog') with [hɔndə] *honden* ('dogs'). Ernestus and Baayen (2003, 2004) have shown that this alternation, traditionally regarded as idiosyncratic, is affected by paradigmatic structures driven by the rhyme of the final syllable. In addition, they have shown that devoiced obstruents (e.g., the [t] of [hɔnt]) may carry residual traces of voicing, and that listeners are sensitive to these residual traces (Ernestus and Baayen, 2006). The Paradigmatic Signal Enhancement Hypothesis builds on these findings by predicting that greater paradigmatic support for voicing will correlate with enhanced acoustic salience of residual voicing in the devoiced obstruent.

Additional evidence for the Paradigmatic Signal Enhancement Hypothesis emerges from research on intrusive /r/ in New Zealand English (Hay and Maclagan, in press): The more likely speakers are to produce intrusive /r/ given a range of linguistic and sociolinguistic factors, the more salient its realization (as reflected in the degree of constriction).

Finally, the probabilistic dependencies between morphemes, such as exist between the interfix, the compound's left and right constituents, and the whole compound, challenge the fully decompositional theory of morphological encoding in speech production, developed by Levelt, Roelofs and Meyer (1999). According to this model, an abstract lemma representation provides access to a word's individual constituents. The planning for articulation of these individual constituents is fully encapsulated from all other morphemes and their paradigmatic relations. This model is challenged not only by the present findings, but also by those of Van Son and Pols (2003), Pluymaekers *et al.* (2005a), Pluymaekers *et al.* (2005b), Hay (2003), and Ernestus *et al.* (2006). What the present paper adds to this literature is the surprising observation that fine phonetic detail is not only determined by the properties of the word itself and its nearest phonological neighbors, but also by its morphological paradigmatic structure.

References

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

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47:31–56.

Aylett, M. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllabic nuclei. *Journal of the Acoustic Society of America*, 119:3048–3058.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Bard, E., Anderson, A., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language*, 42:1–22.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical society of America*, 113(2):1001–1024.

Bertram, R. and 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.

Bien, H., Levelt, W., and Baayen, R. (2005). Frequency effects in compound production. *PNAS*, 102:17876–17881.

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

Cambier-Langeveld, T. (2000). *Temporal marking of accents and boundaries*. LOT, Amsterdam.

Crystal, T. and House, A. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, 88:101–112.

Cucchiarini, C. (1993). *Phonetic transcription: A methodological and empirical study*. University of Nijmegen, Nijmegen.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and 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 Saussure, F. (1966). *Course in General Linguistics*. McGraw, New York.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch: A Corpus-Based Study of the Phonology-Phonetics Interface*. LOT, Utrecht.

Ernestus, M. and Baayen, R. (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In Goldstein, L., Whalen, D., and Best, C., editors, *Laboratory Phonology 8*, pages 27–49. Mouton de Gruyter, Berlin.

Ernestus, M. and Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, 79(1):5–38.

Ernestus, M. and Baayen, R. H. (2004). Analogical effects in regular past tense production in Dutch. *Linguistics*, 42:873–903.

Ernestus, M., Lahey, M., Verhees, F., and Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, 120:1040–1051.

Fosler-Lussier, E. and Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Communication*, 29:137–158.

Fougeron, C. and Keating, P. (1997). Articulatory strengthening at the edges of prosodic domains. *Journal of the Acoustical Society of America*, 101(6):3728–3740.

Fowler, C. and Housum, J. (1987). Talkers' Signalling of "New" and "Old" Words in Speech and Listeners' Perception and Use of the Distinction. *Journal of Memory and Language*, 26:489–504.

Hay, J. (2003). *Causes and Consequences of Word Structure*. Routledge, New York and London.

Hay, J. and Maclagan, M. (in press). Social and phonetic conditioners on the frequency and degree of "intrusive /r/" in New Zealand English. In Preston, D. and Niedzielski, N., editors, *Methods in Sociophonetics*.

Hunnicutt, S. (1985). Intelligibility versus redundancy - conditions of dependency. *Language and Speech*, 28:47–56.

Johnson, K. (2004). Massive reduction in conversational American English. In *Spontaneous speech: data and analysis. Proceedings of the 1st session of the 10th international symposium*, pages 29–54, Tokyo, Japan. The National International Institute for Japanese Language.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, *Frequency and the emergence of linguistic structure*, pages 229–254. Benjamins, Amsterdam.

Kessens, J. and Strik, H. (2001). Lower WERs do not guarantee better transcriptions. In *Proceedings of Eurospeech-2001*, pages 1721–1724, Aalborg, Denmark.

Keune, K., Ernestus, M., Van Hout, R., and Baayen, R. (2005). Social, geographical, and register variation in Dutch: From written 'mogelijk' to spoken 'mok'. *Corpus Linguistics and Linguistic Theory*, 1:183–223.

Koester, D., Gunter, T. C., Wagner, S., and Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: the auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience*, 16:1647–1668.

Krott, A., Baayen, R. H., and Schreuder, R. (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics*, 39(1):51–93.

Krott, A., Hagoort, P., and Baayen, R. H. (2004). Sublexical units and supralexical combinatorics in the processing of interfixed Dutch compounds. *Language and Cognitive Processes*, 19:453–471.

Krott, A., Krebbers, L., Schreuder, R., and Baayen, R. H. (2002a). Semantic influence on linkers in Dutch noun-noun compounds. *Folia Linguistica*, 36(1-2):7–22.

Krott, A., Schreuder, R., and Baayen, H. (2002b). Linking elements in Dutch noun-noun compounds: Constituent families as predictors for response latencies. *Brain and Language*, 81(1-3):708–722.

Ladefoged, P. (1982). *A course in phonetics*. Hartcourt, Brace, Jovanovich, New York, 2 edition.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22:1–38.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6:172–187.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W. and Marchal, A., editors, *Speech production and speech modeling*, pages 403–440. Kluwer, Dordrecht.

Losiewicz, B. (1992). *The effect of frequency on linguistic morphology*. PhD thesis, University of Texas.

McAllister, J., Potts, A., Mason, K., and Marchant, G. (1994). Word duration in monologue and dialog speech. *Language and Speech*, 37:393–405.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94:1–18.

Munson, B. and Solomon, N. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language and Hearing Research*, 47:1048–1058.

Neijt, A., Krebbers, R., and Fikkert, P. (2002). Rhythm and semantics in the selection of linking elements. In Broekhuis, H. and Fikkert, P., editors, *Linguistics in the Netherlands 2002*, pages 117–127. Benjamins, Amsterdam.

Nooteboom, S. G. (1972). *Production and perception of vowel duration: A study of the durational properties of vowels in Dutch*. University of Utrecht, Utrecht.

Oostdijk, N. (2000). The Spoken Dutch Corpus Project. *The ELRA Newsletter*, 5:4–8.

Pluymaekers, M., Ernestus, M., and Baayen, R. (2005a). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62:146–159.

Pluymaekers, M., Ernestus, M., and Baayen, R. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, 118(4):2561–2569.

Samuel, A. and Troicki, M. (1998). Articulation quality is inversely related to redundancy when children or adults have verbal control. *Journal of Memory and Language*, 39:175–194.

Scarborough, D. L., Cortese, C., and Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology:*

Human Perception and Performance, 3:1–17.

Scarborough, R. (2004). Degree of Coarticulation and Lexical Confusability. In *Proceedings of the 29th Meeting of the Berkeley Linguistics Society, February 14-17, 2003*.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Shields, L. and Balota, D. (1991). Repetition and associative context effects in speech production. *Language and Speech*, 34:47–55.

Slis, I. H. and Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction ii. *Language and Speech*, 12:80–102; 137–155.

Van Son, R., Binnenpoorte, D., van den Heuvel, H., and Pols, L. (2001). The IFA Corpus: a phonemically segmented Dutch Open Source speech database. In *Proceedings of Eurospeech-2001*, Aalborg, Denmark.

Van Son, R. and Pols, L. (2003). Information Structure and Efficiency in Speech Production. In *Proceedings of Eurospeech-2003*, Geneva, Switzerland.

Van Son, R. and Van Santen, J. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, 47:100–123.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28(4):735–747.

Vorstermans, A., Martens, J., and Van Coile, B. (1996). Automatic segmentation and labelling of multi-lingual speech data. *Speech Communication*, 19:271–293.

Waals, J. (1999). *An experimental view of the Dutch syllable*. Holland Academic Graphics, The Hague.

Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (2002). *The HTK Book 3.2*. Entropic, Cambridge.

Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, 15:1–95.

Reading Polymorphemic Dutch Compounds: Towards a Multiple Route Model of Lexical Processing

Chapter 3

This chapter is an article in press: Victor Kuperman, Robert Schreuder, Raymond Bertram, and R. Harald Baayen. Reading Polymorphemic Dutch Compounds: Towards a Multiple Route Model of Lexical Processing. *Journal of Experimental Psychology: Human Perception and Performance*

Abstract

This paper reports an eye-tracking experiment with 2500 polymorpemic Dutch compounds presented in isolation for visual lexical decision, while readers' eye-movements were registered. We found evidence that both full-forms of compounds (*dishwasher*) and their constituent morphemes (e.g., *dish*, *washer*, *er*) and morphological families of constituents (sets of compounds with a shared constituent) played a role in compound processing. We observed simultaneous effects of compound frequency, left constituent frequency and family size early (i.e., before the whole compound has been scanned), and also effects of right constituent frequency and family size that emerged after the compound frequency effect. The temporal order of these and other effects that we observed goes against assumptions of many models of lexical processing. We propose specifications for a new multi-route model of polymorphemic compound processing, which is based on time-locked, parallel and interactive use of all morphological cues, as soon as they become (even partly) available to the visual uptake system.

Introduction

Current models of morphological processing and representation in reading have explored a wide range of logically possible architectures. Sublexical models hold that complex words undergo obligatory parsing and that lexical access proceeds via their morphemes (cf., Taft, 1991; Taft & Forster, 1975, 1976). Supralexical models, by contrast, argue that morphemes are accessed only after the compound as a whole has been recognized (e.g., Giraudo & Grainger, 2001). Dual route models hypothesize that full-form based processing goes hand in hand with decompositional processing. The two access routes are usually assumed to be independent (Allen & Badecker, 2002; Baayen & Schreuder, 1999; Frauenfelder & Schreuder, 1992; Laudanna & Burani, 1995; Schreuder & Baayen, 1995), although an interactive dual route model has been proposed as well (Baayen & Schreuder, 2000). In connectionist models such as the triangle model (Seidenberg & McClelland, 1989), morphological effects are interpreted as arising due to the convergence of orthographic, phonological and semantic codes. What all these theories have in common is that they were developed to explain data obtained with chronometric measures for isolated reading of bimorphemic complex words. As a consequence, they tend to remain silent about the time-course of information uptake in the reading of complex words.

Establishing the temporal order of activation of full-forms (e.g., *dishwasher*) of complex words and of their morphological constituents (e.g., *dish* and *washer*) is critical for adjudicating between competing models of morphological processing. The present study addresses the time-course of morphological processing by considering the reading of long, polymorphemic Dutch compounds. Importantly, current models of morphological processing offer different predictions with regard to the visual recognition of such compounds. On supralexical models, one expects activation of the compound's full-form (diagnosed by the compound frequency effect) as the initial step of lexical access. After the full-form of the compound is activated, one expects to observe simultaneous activation of both the left and the right constituent (diagnosed by frequency-based properties of a constituent). On strict sublexical models, the predicted order of activation is as follows: first, the left constituent of a compound, second, its right constituent, and finally (either coinciding with activation of the right constituent, or following it) the full-form. The sublexical model of Taft and Forster (1976) argues that activation of the compound's left constituent is sufficient to trigger the retrieval of the compound's full-form. This model predicts sequential effects of the left constituent frequency and compound frequency, and no effects of the right constituent. On some dual-route models of parallel processing, one expects roughly simultaneous effects of compound frequency and left constituent frequency, since both routes are argued to be pursued simultaneously and independently (e.g., Baayen & Schreuder, 1999). Bertram and Hyönä (2003) have also proposed a dual-route architecture with a headstart for the decomposition route in case of long compounds, which predicts early effects pertaining to the compound's left constituent followed by the compound frequency effect.

Earlier eye-tracking studies not only confirmed the joint relevance of both constituents and full-form representations for reading posited by dual route models (Andrews, Miller & Rayner, 2004; Hyönä, Bertram & Pollatsek, 2004, Zwitserlood, 1994), they have also made more precise information about the time-course of morphological processing available. For instance, Hyönä *et al.* (2004) found that for long compounds there is early activation of the left constituent (*dish*) and later activation of the right constituent (*washer*). However, two important questions about the time-course of morphological processing are as yet unresolved. First, the temporal locus of compound frequency effects remains unclear. Several eye-tracking studies of compounds (cf., Andrews *et al.*, 2004; Bertram & Hyönä, 2003; Pollatsek *et al.*, 2000) have observed effects of compound frequency for the very first fixation, but these effects failed to reach significance. ERP studies of reading (Hauk & Pulvermüller, 2004; Penolazzi, Hauk & Pulvermüller, 2007; Sereno, Rayner & Posner, 1998) have repeatedly shown early effects of whole word frequency ($<$ 150-200 ms), but they focused on relatively short (4-6 characters) and morphologically simplex words. An early locus for the compound frequency effect in long compounds would challenge strict sublexical accounts of morphological processing, according to which whole word frequency effects would reflect post-access combinatorial processes instead of tapping into early visual information uptake.

Second, it is unclear whether the activation of the compound's full-form precedes, follows or coincides with the activation of the compound's constituents. The present evidence is controversial. For instance, Juhasz, Starr, Inhoff and Placke (2003) argued – on the basis of eye-tracking, lexical decision and naming experiments – that it is the compound's head, the last constituent to be read (e.g., *washer* in *dishwasher*), that plays the decisive role in the late stages of compound recognition, while the effects of the initial constituent emerge early and are weak (see, however, Juhasz, 2007). A possible reason for the dominance of the right constituent is its typical semantic convergence with the meaning of the whole compound (see also Duñabeitia, Perea & Carreiras, 2007). These results were argued to support models that argue for either co-activation of the right constituent and the full-form (Pollatsek, Hyönä & Bertram, 2000), or activation of the right constituent following activation of the full-form (Giraudo & Grainger, 2001). Their claim contrasts with chronometric studies by e.g., Taft and Forster (1976) who found evidence for the left constituent guiding lexical access to a compound's meaning. Taft and Forster (1976) saw these results as evidence that a compound's full-form gets activated *after* the left constituent of the compound receives activation.

The first aim of the present study is to address the temporal order of lexical access to the full-form and the morphological constituents of compounds. In other words, we explore how soon and in what order do the properties of the compound's full-form, and the properties of the compound's left and right constituents, emerge in the timeline of compound recognition. Second, we broaden the scope of constituent processing by probing whether morphological families of constituents (i.e., sets of compounds sharing a constituent, e.g., *ice pick*, *ice cube*, *ice box*) contribute to the speed of processing over and above properties of full-forms and those of constituents as isolated words. Lexical decision studies argued that the effects of constituent families are semantic in character, and hence emerge late, at the peripheral post-access stages of the complex word processing (e.g., De Jong, Schreuder & Baayen, 2000). In this study we tackle the temporal locus of the effects of constituent families using eye-tracking as a technique with a better temporal resolution than the one offered by lexical decision latencies. Third, we zoom in on the issue of independence of the full-form and decompositional processing routes claimed in some dual-route parallel processing models by considering the possibility that the effects elicited by the full-form properties might be modulated by constituent properties.

Instead of investigating bimorphemic compounds, we examined compounds with three to six morphemes. Type-wise, such polymorphemic compounds are more common in Dutch than the bimorphemic compounds that are traditionally studied in the experimental literature. For instance, perusal of CELEX (Baayen, Piepenbrock & Gulikers, 1995) shows that 54% of the nominal compounds has more than two morphemes. An additional dimension of morphological processing that we consider as the fourth goal of our study is the role of (free-standing and bound) morphemes deeply embedded in morphological structure (e.g., *wash-* and *-er* in *dishwasher*). Are morphemes at lower levels of morphological hierarchy recognized as independent units of meaning by the human lexical processor and used in compound identification, or are they invariably treated as parts of larger structural units (e.g., *washer*)? If, as we will argue, readers maximize their use of cues available for efficient compound identification, we may expect that the deeply embedded free and bound morphemes are used in the course of processing as well.

In what follows, we report a large regression experiment with 2500 target compounds that combined eye-tracking of isolated word reading with lexical decision as superimposed task to ensure sufficient depth of processing. We opted for this combination since it provides detailed insight into the time-course of morphological processing and it provides sufficient statistical power. In the General Discussion, we return in detail to the methodological consequences of our decision to make use of lexical decision rather than sentential reading. Here, we restrict ourselves to noting that a parallel study presenting Finnish compounds in sentential contexts reported in Chapter 4 of this dissertation (Kuperman, Bertram & Baayen, 2008) yielded a pattern of results that is highly consistent with the morphological effects reported below. Our present experiment provides evidence that current models of morphological processing are too restrictive in their architectures, and that a more flexible framework in which all opportunities for recognition are maximized (Libben, 2006) is called for.

Method

Participants

Nineteen students of the Radboud University of Nijmegen (12 females and 7 males) were paid 20 euro for participation in the study. All were native speakers of Dutch and reported normal or corrected-to-normal vision and right-handedness.

Apparatus

Eye movements were monitored by the head-mounted video-based EYELINK II eye-tracking device produced by SR Research (Mississauga, Canada). The average gaze position error of EYELINK II is $<0.5^o$, while its resolution is 0.01^o. Recording of the eye movements was performed on the left eye only and in the pupil-only mode. The sampling rate of recording used in this study was 250 Hz. The 17-inch computer monitor used for the display of the stimuli had a 60 Hz refresh rate.

Stimuli

In total, 2500 lexical items (1250 existing words and 1250 nonce compounds) were included as stimuli. A list of existing polymorphemic Dutch compounds (triconstituent compounds, or biconstituent compounds with at least one and at most four derivational affixes) was selected from the CELEX lexical database (Baayen, Piepenbrock & Gulikers, 1995), for instance, *werk+gev-er*, "work-giver", i.e., "employer". Additionally, a list of multiply complex nonce compounds was created by blending existing words into novel combinations (i.e., combinations that are not registered in the CELEX database), for instance, *alarmijsbaan*, composed of *alarm* "alarm" and the compound word *ijsbaan* "skating ring". At the level of immediate constituents, the resulting targets and fillers represented a mixture of noun-noun, adjective-noun and verb-noun compounds.

The average number of morphemes per stimulus was 3.2 (SD = 0.4). The maximum length of a stimulus was set at 12 characters. The resulting range of 8-12 characters (mean length = 11.62, SD = 0.74) allowed for a tight experimental control of word length, and kept collinearity of such measures as word length and frequency, and left constituent length and frequency within reasonable bounds. Stimuli were displayed one at a time in a fixed-width font Courier New size 12. With a viewing distance of about 80 cm, one character space subtended approximately 0.36^o of visual angle.

Procedure

Participants were instructed to read words at their own pace. They were also informed that nonce compounds were built of existing Dutch words and were asked to evaluate the *whole* stimulus as an existing word or a non-word by pressing the right button ("Yes" response) or the left button ("No" response) of a dual button box. Prior to the presentation of the stimuli, the eye-tracker was calibrated using a nine-point grid that extended over the entire computer screen. Prior to each stimulus, a fixation point was presented in the central position of the screen for 500 ms. After each third stimulus a drift correction was performed using the screen-central fixation point as a mark. After 500 ms or after the calibration was corrected, a stimulus was displayed in black lower-case characters on a white background. When one of the dual box buttons was pressed, the stimulus was removed from the screen and a fixation point appeared. If no response was registered after 5000 ms, a stimulus was removed from the screen and the next trial was initiated. Participants' responses and response times were recorded along with their eye movements.

Stimuli were displayed centralized vertically, and slightly off-center horizontally such that the space between the fourth and the fifth characters of a stimulus was always at the center of the screen where the fixation point was shown. This position is closest to the preferred viewing position (the most frequent position where the eyes initially land) reported in eye movements studies for Finnish, English and French words with the lengths that we used, mostly 12 characters, (e.g., Bertram & Hyönä, 2003; McDonald & Shillcock, 2004; Vergilino-Perez, Collins & Doré-Mazars, 2004).

The presentation order of stimuli was randomized. Stimuli were presented in two separate sessions each consisting of three blocks. The order of presentation of the blocks and the order of the words within each block were the same for each participant (see Appendix 2 for the discussion of randomization procedures). For each participant, sessions were run on two different dates, while blocks within one session were separated by a five to ten minute break. After each break the eye-tracker was calibrated again. A single session lasted 70 minutes at most, and the total time of the experiment lasted a maximum of 130 minutes.

Dependent variables

For the analysis of the lexical decision data, we considered as dependent variables the (natural) log-transformed response times (*RT*), as well as the accuracy of responses (*Correct*).

In the eye-tracking data analysis, we selected as early measures of lexical processing the first fixation duration, *FirstDur*, and the subgaze duration on the compound's left constituent, *SubgazeLeft* (the summed duration of all fixations on the left constituent before exiting it). As measures that tap into later stages of compound recognition, we considered subgaze for the right immediate constituent, *SubgazeRight* (the summed duration of all fixations on the right constituent before exiting it). Gaze duration, *GazeDur*, served as the global measure of processing difficulty. In this study, gaze duration was defined as the summed duration of all fixations on the target word that were completed before one of two events took place: Either the reader fixated away from the word, or the lexical decision was made[1]. All durational measures were natural log-transformed to reduce

[1]Note that *SubgazeLeft* and *SubgazeRight* are not strictly additive in the measure of gaze duration. In the situation where fixation 1 is on the left constituent, fixation 2 on the right one and fixation 3 on the left one, *SubgazeLeft* is equal to the duration of fixation 1, and *SubgazeRight* to the duration of fixation 2. The measure of gaze duration, however, would be equal to the sum of 1, 2 and 3, and could show an effect that differs in size from the sum of effects found for both subgazes. Also we fitted the statistical models to the subgaze measures with the non-zero duration. There are words, however, in which all fixations fall on one constituent, and there is no subgaze duration

the influence of atypical outliers. We considered several other eye-movement measures as well: These included single, second and third fixation durations; initial fixation position; the amplitude of the first within-word saccade; the probability of a given fixation being the last one on the word; the probability of a given fixation being to the left of the previous fixation; and the total number of fixations on a word. The data patterns for these measures were in line with the ones we reported, but did not offer substantial additional insight into our research questions.

Predictors

Morphological variables. The measures of morphological characteristics of stimuli included: whole word (compound) frequency, *WordFreq*; the word frequency of the left constituent as an isolated word, *LeftFreq*; and the word frequency for the right constituent as an isolated word, *RightFreq*. All these frequencies were lemma frequencies, i.e., summed frequencies of a compound word and of its inflectional variants (e.g., sum of frequencies of the singular form *newspaper*, the plural form *newspapers* and the singular and plural genitive forms *newspaper's* and *newspapers'*).

All frequency-based measures in this study, including the ones reported in the remainder of this section, were obtained from CELEX (counts based on a corpus of 42 million word forms) and log-transformed to reduce the influence of outliers.

We also considered measures of morphological connectivity for the constituents of our compounds. We refer to the set of compounds that share the left (right) constituent with the target as the left (right) morphological family of that constituent (e.g., the left constituent family of *ice cream* includes *ice pick*, *ice cube* and *ice box*). Words that appear as constituents in many compounds (i.e., have large morphological families) or in frequent compounds (i.e., have high family frequency) have been repeatedly shown across languages to elicit shorter lexical decision latencies, whether presented visually or auditorily (cf., e.g., De Jong, Schreuder & Baayen, 2000; De Jong, Feldman, Schreuder *et al.*, 2002; Dijkstra, Moscoso del Prado Martín, Schulpen *et al.*, 2005; Moscoso del Prado Martín, Bertram, Häikiö *et al.*, 2004). Left constituent family size is also known to modulate gaze duration in interaction with semantic opacity of Finnish compounds, cf., Pollatsek and Hyönä (2005)[2].

for the other constituent. In such cases there is only one subgaze component contributing to the composite measure of gaze duration.

[2]For both the left and the right constituents, the alternative measure of family frequency (the summed token frequency of the members in the morphological family) consistently elicited weaker effects than family size of the respective constituents in all statistical models, in contrast to findings

Morphological family size for the left constituents in our compounds strongly correlated with the frequencies of these left constituents as isolated words. We orthogonalized these collinear measures by fitting a regression model where left constituent family size was predicted by left constituent frequency. We then considered the residuals of this model, *ResidLeftFamilySize*, as our new left family size measure. It was highly correlated with the original measure ($r = 0.95, p < 0.0001$), but the effects of constituent frequency were now partialled out. Using the same procedure for the right constituent family size and frequency we obtained *ResidRightFamilySize*, which again closely approximated right constituent family size ($r = 0.93, p < 0.0001$), and was orthogonal to *RightFreq*. We decorrelated family size and frequency for analytical clarity, in order to be better able to assess the independent contributions of predictors (beta coefficients) to the model.

The presence of each subconstituent morpheme and its position in the morphological structure were coded by the multi-level factor *Affix* with the following levels: "Initial" (for compounds with prefixed left constituents), "Medial" (for compounds with a suffixed left constituent, an interfix, a prefixed right constituent, or with any combination of these affixes), "Final" (for compounds with suffixed right constituents), "Multiple" (for compounds with multiple affixes[3]) and "Tri" (for 'pure triconstituent' compounds with three word stems and no affixes; for the sake of analytical clarity, we excluded from our analyses 112 compounds with three word stems and further affixes). The resulting counts of stimuli representing each type of morphological complexity are summarized in Table 3.1.

We also considered affix productivity, *AffixProd* (the type count of derived words in which the affix occurs). The total number of morphemes in the compounds was included as an index of the compound's morphological *Complexity*.

---

of De Jong *et al.* (2002) for Dutch compounds. The difference in effect sizes was revealed in smaller regression (beta) coefficients for family frequencies, when constituent family frequencies and family sizes were included, separately, as predictors in our statistical models. For instance, in the model for gaze duration, the regression coefficient was -0.026 for left constituent family frequency and -0.036 for left constituent family size. As the distinction between family size and family frequency effects is not crucial for our research questions, we do not discuss this measure further. We rather note that the entropy measure proposed by Moscoso del Prado Martín *et al.* (2004) may be a possible resolution for the relative impacts of the family-based alternatives.

[3]We classified compounds with more than one affix at the immediate constituent boundary, such as *rov-er-s-hol*, "robbers' den", as Medial rather than as Multiple. In other words, the category Medial comprises compounds with at least one medial affix, while the category Multiple comprises compounds with affixes at more than one position in the compound. We opted not to differentiate between compounds with different numbers of medial affixes, since the effects of these affixes considered separately were very similar across our analyses.

Table 3.1: Counts of compounds partitioned by type of morphological complexity.

	Type of Complexity	Number of stimuli
1	Triconstituent	580
2	Initial	158
3	Medial	541
4	Multiple	407
5	Final	702

Other variables. We also considered word length (*WordLength*) (in the range of 8-12 characters), as well as left constituent length (*LeftLength*). The longitudinal effect of the experimental task on the participants' behavior (e.g., fatigue or habituation as the participant works through the experiment) was estimated by means of the position of the stimulus in the experimental list, *TrialNum*. We also took into account the influence that carried over from trial $N-1$ to trial N (see Baayen, Davidson & Bates, 2008; De Vaan *et al.*, 2007) by considering the log-transformed response time from the trials immediately preceding the current one (*RT1*). Other control predictors that reached significance in codetermining either the lexical decision latencies or reading times as revealed in eye-movements are presented in Appendix 1.

Table 3.3 in Appendix 1 lists the distributions of the continuous variables used in this study, including their ranges, and mean and median values.

Statistical Considerations

In this study we made use of mixed-effects multiple regression models with random intercepts for *Subject* and *Word* (and occasionally by-subject random slopes and contrasts for item-bound predictors), and the predictors introduced above as fixed effect factors and covariates (cf., Baayen, 2008; Bates & Sarkar, 2005; Pinheiro & Bates, 2000).

Unless noted otherwise, only those fixed effects are presented below that reached significance at the 5%-level in a backwards stepwise model selection procedure. All random effects included in our models significantly improved the explanatory value of those models, as indicated by significantly higher values of the maximum likelihood estimate of the model with a given random effect as compared to the model without that random effect (all ps < 0.0001 using likelihood ratio tests), for detailed treatment of random effects in mixed-effects models see Pinheiro and

Bates (2000). Below we report which predictors required random slopes in addition to the random intercepts for *Subject* and *Word*, see Table 3.9 in Appendix 1.

All models were fitted and atypical outliers were identified, i.e., points that fell outside the range of -2.5 to to 2.5 units of SD of the residual error. Such outliers were removed from the respective datasets (and were not used in the composite eye-movement measures) and the models were refitted in order to avoid distortion of the model estimates due to atypical extreme observations. Below we report statistics of those refitted models.

Due to the large number of models fitted in this study, we only report in Appendix 1 the full specifications of the model for lexical decision latencies for existing words, and of the four models for the eye movements measures (first fixation duration, subgazes for the left and the right constituent, and gaze duration).

Results and Discussion

Lexical Decision

The initial lexical decision data pool consisted of 2500 words x 19 participants = 47500 trials. From this dataset we excluded one word that was misspelled, as well as the trials in which the (log) RT value fell beyond 3 units of standard deviation from the mean. Since no participant exceeded the threshold of a 30% error rate in either nonce compounds or the existing words, none were excluded. The resulting dataset consisted of 47206 trials, of which 41245 were correct replies. The error rate reached 23% for existing words and 3% for nonce compounds. Thus, in the lexical decision task participants exhibited a clear bias towards "no"-responses, which does not come as a surprise given that many of the existing compounds are fairly low-frequency words and also semantically opaque words, the meaning of which is conceptually difficult to construct from the individual constituents, just as is the case with many nonce compounds. For correct replies, the average lexical decision latency was 763 ms (SD = 246) for existing words and 801 ms (SD = 261) for nonce compounds.

Below we only discuss the analysis of the lexical decision latencies for the 18217 trials with existing compounds that were correctly identified in the lexical decision task.

Morphological Variables. Column *RT* in Table 3.2 summarizes the effects of compound frequency and frequency-based measures of a compound's constituents on the lexical decision latencies (see Table 3.4 in Appendix 1 for the full specification of the model). The column provides effect sizes for morphological predictors (see Appendix for the explanation as to how these were computed) and p-values for main effects, as well as indicates interactions between predictors of interest. For clarity of exposition, we leave out from the table the effects of morphemes deeply embedded in the compound structure: These are discussed separately.

Both compound frequency (*WordFreq*) and morpheme-based frequencies (*LeftFreq*, *RightFreq*), and morphological connectivity measures (*ResidLeftFamilySize*, *ResidRightFamilySize*) entered into negative correlations with the RTs, i.e., higher frequencies or larger families facilitated compound processing.

Of these predictors, compound frequency showed the greatest effect (-96 ms). These facilitatory morphological effects are in accord with previous reports of visual lexical decision experiments with Dutch and English compounds (cf., e.g., Andrews, 1986; De Jong *et al.*, 2000; De Jong *et al.*, 2002; Juhasz *et al.*, 2003).

Interestingly, compound frequency interacted with left constituent frequency in such a way that the effect of compound frequency was strongest in compounds with the low-frequency left constituents and was weaker in compounds where left constituents were relatively frequent, see Figure 3.1.

Suppose, following Libben (2006), that both compound frequency and left constituent frequency are among the morphological cues that the lexical processor may use to facilitate recognition of the compound. Then the observed interaction is the evidence that the magnitude of one such cue (e.g., left constituent frequency) appears to modulate the extent to which the other cue (e.g., compound frequency) contributes to the identification of the complex word.

We also observed an interaction between right constituent frequency and left constituent family size, see Figure 3.2. The effect of right constituent frequency was strongest in compounds with large left constituent families (i.e., with a large number of possible morphemic continuations for the left constituent, e.g., *shoelace*, *shoe cream*, *shoe shop*), and decreased with decreasing morphological family size.

Apparently, ease of access to the lexical representation of the right constituent (diagnosed by its frequency) speeds up compound recognition more when there is more uncertainty about which candidate to choose from a larger number of possible right constituents. In case the competition in the family is relatively weak, due to a low number of choices, the right constituent may be relatively easy to predict and

Table 3.2: Summary of morphological effects on durational measures

Predictor	RT	FirstDur	SubgazeLeft	SubgazeRight	GazeDur
LeftFreq	-32 ms (<0.001)	-35 ms (<0.001)	-48 ms (<0.001)	ns	-34 ms (<0.001)
- interaction with	*WordFreq* (0.006), Fig. 3.1	*WordFreq* (0.006)	*WordFreq* (0.001)		*WordFreq* (0.01)
ResidLeftFamSize	-43 ms (<0.001)	-47 ms (<0.001)	-83 ms (<0.001)	-24 ms (0.028)	-61 ms (<0.001)
- interaction with	*RightFreq* (0.004), Fig. 3.2	*WordLength* (0.008)			*RightFreq* (0.008)
RightFreq	-37 ms (0.002)	ns	ns	-27 ms (0.001)	-27 ms (<0.001)
- interaction with	*ResidLeftFamSize* (0.004), Fig. 3.2				*ResidLeftFamSize* (0.008)
ResidRightFamSize	-40 ms (0.001)	ns	ns	-54 ms (0.001)	ns
WordFreq	-96 ms (<0.001)	-24 ms (<0.001)	-42 ms (<0.001)	-47 ms (<0.001)	-73 ms (<0.001)
- interaction with	*LeftFreq* (0.006), Fig. 3.1	*LeftFreq* (0.01)	*LeftFreq* (0.001)		*LeftFreq* (0.006)

Numbers in columns 2-6 show sizes of statistically significant main effects. In the case where the main effect is qualified by the interaction (e.g., *WordFreq* by *LeftFreq*, we report the numerical estimate of the main effect size (*WordFreq*) for the median value of the interacting term (*LeftFreq*). The column *RT* provides references to figures illustrating the interactions, which are qualitatively similar for both lexical decision latencies and reading times. Numbers in brackets provide p-values for the main effects and interactions. "ns" stands for non-significant.

Figure 3.1: Interaction of compound frequency by left constituent frequency for lexical decision latencies. The lines plot the effect of compound frequency for the quantiles of left constituent frequency (quantile values provided at the right margin). Compound frequency comes with the strongest negative effect at the 1st quantile (solid line), the effect gradually levels off at the 2nd quantile (dashed line), the 3d quantile (dotted line) and the 4th quantile (dotdash line), and is weakest for the compounds with highest-frequency left constituents, the 5th quantile (longdash line).

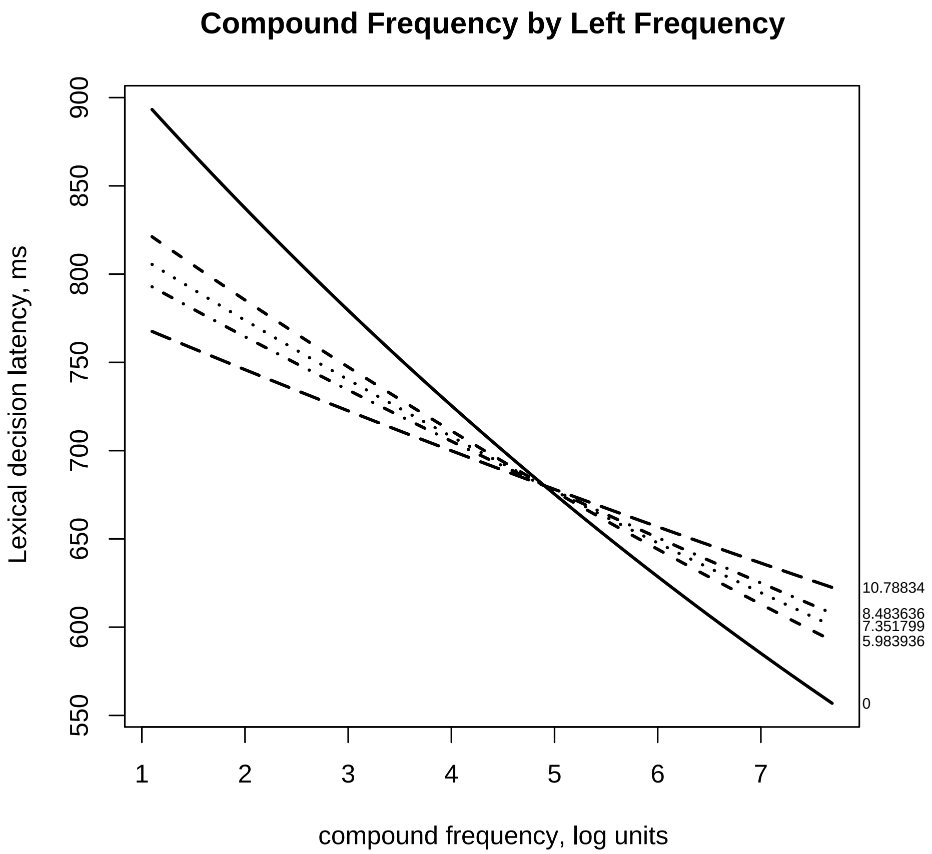
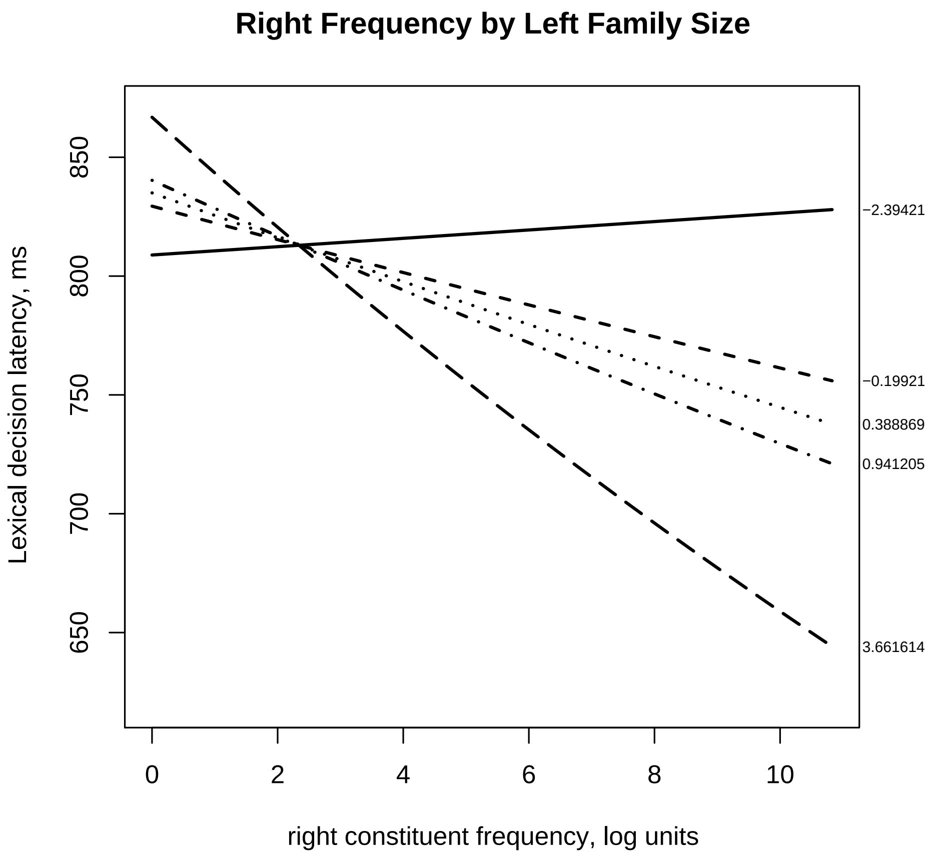


Figure 3.2: Interaction of right constituent frequency by (residualized) left constituent family size for lexical decision latencies. The lines plot the effect of right constituent frequency for the quantiles of left constituent family size (quantile values provided at the right margin). Right constituent frequency has no substantial effect for smallest left constituent families, represented as the 1st quantile (solid line). The effect gradually increases at the 2nd quantile (dashed line), the 3d quantile (dotted line) and the 4th quantile (dotdash line), and it is strongest for compounds with the largest left constituent families, the 5th quantile (longdash line).



additional morphological information in the form of right constituent frequency is not as useful for the lexical processor. Again, we find that the magnitude of one cue for compound recognition affects the utility and magnitude of other such cues.

The effects of lower-level, subconstituent, morphemes revealed that compounds with two stems (of which at least one was a derivation) were processed significantly faster than triconstituent compounds (by about 20 ms, averaged across levels of *Affix*). Moreover, stimuli that comprised more morphemes, as measured by *Complexity* elicited longer latencies (effect size = 86 ms), as expected.

Other Control Variables. We observed habituation of participants to the task: The further they were into the experiment (as estimated by the trial position in the experimental list), the faster their lexical decisions were (effect size = -34 ms).

Longer RTs to the immediately preceding trial (*RT1*) went hand in hand with longer lexical decision latency at the current trial (effect size = 223 ms). These findings make a clear case that both the longitudinal effects of the experimental task and those related to immediately preceding trials contribute substantially to modulating lexical decision latencies.

Eye movements

We considered only the first-pass reading (i.e., the sequence of fixations made before the fixation is made outside of the word boundaries) and only those fixations that were completed before a response button was pressed. Trials with blinks and misreadings (i.e., trials for which no fixations were recorded by the eye-tracking device, due to the machine error) were removed, as well as the trials with lexical decision latencies exceeding 3 units of SD from the mean. The resulting dataset comprised 85908 fixations. We also removed from the dataset of fixations and from composite eye-movement measures those fixations that exceeded 2.5 units of SD from the mean log-transformed duration, whereas the mean duration and the standard deviation were calculated separately for each participant. In this way we avoided penalizing very slow or very fast readers. In total, 2227 (2.6%) outliers were removed, and the resulting range of fixation durations was 49 ms to 1197 ms. Subsequently, fixations that bordered microsaccades (fixations falling within same character) were removed (122 x 2 = 244 fixations, 0.1%). The resulting pool of data points consisted of 83437 valid fixations.

Eighteen percent of the stimuli required a single fixation for reading, 36% required exactly two fixations, 26% required exactly three fixations, and it took four or more fixations to read the remaining 20% of the stimuli. The average number of fixations on a stimulus was 2.6 (SD = 1.2). Regressive fixations (within-word fixations located to the left of the previous fixation) constituted 12.6% of our data pool. The average fixation duration was 262 ms (SD = 117), and the average gaze duration was 620 ms (SD = 382). Eighty-one percent of initial fixations was located either on the fourth or the fifth character of the presented stimulus, which is the area where we intented those fixations to be[4]. Seventy-seven percent of initial fixations were located on the left constituent. Since we had compounds with 2-4 character-long left constituents, a relatively large proportion of initial fixations was located at the right constituent (23%). Seventy-eight percent of second progressive fixations landed on the right constituent.

We further report our findings for the trials with existing compounds and only those that elicited correct responses. Our findings are based on four statistical models: for first fixation duration (14232 data points), for subgaze duration on the left constituent (11684 data points), for subgaze duration on the right constituent (8495 data points), and for gaze duration (14616 data points).

Morphological effects: Compound and immediate constituents. Columns 3 to 6 in Table 3.2 are a summary of the effects that morphological structure elicits in eye-movements across four statistical models (see full specifications for the models in Tables 3.5-3.8 in Appendix 1). Considered jointly, the results of the statistical models in Table 3.2 outline the temporal flow of compound recognition. First, we found evidence that both immediate constituents and the whole compound affect lexical processing of compound words (cf., e.g., Andrews *et al.*, 2004; Bertram & Hyönä, 2003; Hyönä, Bertram & Pollatsek, 2004). In fact, every single morphological predictor that we considered (compound frequency, constituent frequencies and family sizes, as well as properties of deeply embedded morphemes discussed below) had a role to play in the time-course of visual compound recognition. This hints at the possibility that morphological structure offers more cues for the task of compound identification than previously thought.

Second, properties of the left constituents of compounds showed earlier effects

[4]It should be noted that the positions of almost 90% of initial fixations were within the measurement error ($<0.5^o$ of the visual angle) of EYELINK II, that is no more than 1.4 character (14 pixels) away from the displayed fixation point. The shape of the distribution of initial fixation positions was close to normal with the mean of 40.7 pixels (that is, between the 4th and 5th letter) and standard deviation of 8.4 pixels. The initial fixations at the tails of the distribution (in the beginning or the end of the word) may be explained by the somewhat long presentation of the fixation point (500 ms), which may have caused people to occasionally saccade away from that fixation point prior to word presentation.

than the respective properties of the right constituents: the latter were only present in the late measures, *SubgazeRight* and *GazeDur*. Moreover, the impact of the right constituent on compound recognition was considerably weaker than that of the left constituent: The effects of the right constituent were smaller in size and often qualified by interactions with other predictors. These findings may reflect that fact that the left constituent is available earlier to the lexical processor than the right constituent. The typical sequence of fixations in our dataset supported this claim: Initial fixations tended to be located at the left constituent (77% of first fixations), while subsequent fixations mostly landed on the right constituent (78% of progressive second fixations)[5]. We note that the size of the left constituent family codetermined the speed of identification of a compound's right constituent. Apparently, the relative ease of processing of the left constituent spills over to the processing of the right constituent, which is consistent with the spillover effect of word N on word N+1 observed in sentential reading (e.g., Rayner & Duffy, 1986; Reichle, Rayner & Pollatsek, 2004).

Third, the compound frequency effect emerged as early as the first fixation and lingered on throughout the entire time-course of compound processing. That the strong and statistically significant effect of compound frequency shows so early resolves the question raised by Bertram and Hyönä (2003: 627) of whether compound frequency might affect the early stages of visual processing in long compounds. The answer is that it does for 8-12 character-long words[6]. The likelihood that our stimuli, which are mostly 12 character long, are appreciated in one fixation is quite low, in fact, only 18% of our stimuli elicited a single fixation. We conclude that we found evidence that full-form access (diagnosed by the compound frequency effect) is initiated before all characters of the compound have been foveally inspected (for the discussion of the early locus of word frequency effect

[5]Given the lengths of our compounds and the initial fixation positions, it is likely that some characters from the right constituent are identified during an initial fixation on the left constituent. However, the absence of early effects associated with the compound's right constituent implies that the available orthographic information on the right constituent is apparently not sufficient for early activation of that morpheme (cf., Hyönä *et al.*, 2004).

[6]The effect of compound frequency was still significant in the statistical model for the first fixation duration from which single-fixation cases were excluded (model not shown, $p < 0.0001$). We did not observe an interaction of word length by compound frequency, but as the range of word lengths in our study is small, with most words having a length of 12 characters, our data do not shed light on the visual acuity hypothesis of Bertram and Hyönä (2003), according to which compound frequency effects would be more prominent for shorter words with less than 9 characters (Bertram & Hyönä, 2003; cf., also Pollatsek *et al.*, 2000; Niswander-Klement & Pollatsek, 2006).

see also Cleland, Gaskell, Quinlan & Tamminen, 2006).

Fourth, the fact that the effect of compound frequency was simultaneous with the left constituent frequency and family size effect and preceded the right constituent frequency and family effect, poses a problem for strictly sequential sublexical models of morphological processing. In such models, one would expect full-form activation to occur in time after activation of the left and the right constituent. In the Taft and Forster (1976) variant of this model, properties of the right constituent should never exert any influence on compound word identification.activation of the right constituent.

Our set of findings is also problematic for supralexical models, as those models argue for initial activation of the full-form and subsequent spreading activation of constituent morphemes. On this view, the properties of the left and the right constituents are expected to receive activation from the full-form and left and right constituent frequency effects should therefore kick in later than the full-form frequency effect. In fact, however, our data show that at least right constituent effects only emerge in later or global processing measures, i.e., subgaze duration for the right constituent and gaze duration.

Fifth, we observed two surprising effects of constituent morphological paradigms. Left constituent family size effect showed up at the first fixation, which is unexpectedly early given the traditional interpretation of family size effects as a post-access semantic effect reflecting activation spreading through morphological paradigms (cf., e.g., Bertram, Schreuder & Baayen, 2000; De Jong *et al.*, 2000; De Jong, Feldman, Schreuder *et al.*). To explain the finding one has to assume that either the family size effect is formal rather than semantic in nature, or that semantic effects can emerge earlier than usually claimed. As we outline in the General Discussion, we believe that both the formal and the semantic components contribute to the family size effect. On the other hand, we found a late effect of *ResidRightFamSize* on subgaze duration for the right constituent. Recall that the right constituent family is a set of compounds (e.g., *vanilla cream, ice cream, shoe cream*, etc.) beginning in morphemes that can combine with the given right constituent (*cream*). The effect is surprising since by the time when the right constituent is scanned, it is quite plausible that the one left constituent that actually occurs in the compound (e.g., *vanilla*) has already been (partly) identified and then activation of a paradigm of possible left constituents (e.g., *vanilla, ice, shoe*, etc.) appears unwarranted. It is likely that the effect of the right constituent family may be driven by cases in which lexical processing of the left constituent is not complete at the first fixation (for instance, due to difficult lexical processing of the left constituent or suboptimal visual uptake of word-initial information) and continues as a spillover effect even as the eyes move to the right constituent. We return to the role of morphological families in the General Discussion.

Sixth, the interactions between morphological predictors that we saw in lexical decision latencies were replicated in eye-movement measures. As early as the first fixation, left constituent frequency modulated the compound frequency effect, such that compound frequency contributed most to recognition of those compounds in which left constituent frequency was lower, and the compound frequency effect diminished as the left constituent frequency increased (see Figure 3.1). Importantly, compound frequency still has a large role to play even when the left constituent frequency is high and the traditional decompositional route is supposed to be the preferred route of compound processing. This interaction indicates that activation of compounds' full-forms and of morphemes is not independent as claimed in several dual-route models of morphological processing, and that the lexical processor is not identifying compounds by strictly selecting between decomposition or full-form processing. Instead, the processing appears to be flexible and co-operative, taking advantage of both (or more, see below) routes, even when it is prompted to rely more upon one of the routes. Thus, identification of the compound through its full-form is optimal when the other route is less beneficial for identification purposes, and vice versa morphological decomposition preferentially takes place when full-form access is less favorable for compound recognition. Moreover, balanced utilization of the two routes is in place from the earliest stages of complex word recognition.

Also, in subgaze duration for the right constituent we observed the interaction of *ResidLeftFamSize* by *RightFreq*, which showed the strongest effect of right constituent frequency in compounds with large left constituent families, and thus with many potential right constituents that might follow the left constituent (see Figure 3.2). As we argued above, we take this interaction as evidence that (morphological or other) properties of morphemes and complex words serve as cues to recognition of morphologically complex structures and that some cues modulate the presence and magnitude of the effect of other cues.

Morphological effects: Deeply embedded morphemes. Thus far we have considered morphological structure at the level of the whole compound and its immediate constituents. We now consider the effects of the internal structure of these immediate constituents.

Similarly to the lexical decision latencies, triconstituent compounds (i.e., those combining three lexemes) consistently elicited longer reading times in the eye-movement record than compounds with two lexemes (one of which additionally included derivational morphemes). The divergence in the processing of the two compound types did not emerge immediately, at the first fixation, rather it presented itself in subgaze and gaze durations. As effects related to meaning are assumed to occur late, we conclude that the divergence reflects a relative difficulty of semantic integration of three, rather than two, free-standing lexemes (on the temporal order of morphological and semantic effects in compounds, see e.g., Cunnings & Clahsen, 2007).

The role of affix position in a complex word varied in accordance with the temporal order of the visual uptake. Obviously, compound-final affixes are viewed with more acuity when the compound's right constituent, rather than the left one, is under foveal inspection. Indeed, compound-final affixes elicited shorter subgaze durations and gaze durations, but their effect was five times stronger in the model for *SubgazeRight* ($\hat{\beta} = -0.10, p = 0.0001$) than it was in the model for *SubgazeLeft* ($\hat{\beta} = -0.02, p = 0.0001$). Furthermore, multiple affixes appeared to facilitate processing even more than other types of affixation, as revealed in subgaze duration for the left constituent (see Table 3.6). This finding is consistent with the hypothesis that affixes function as segmentation cues in locating the boundaries of morphological constituents (Chapter 4). The observed advantage of compounds with multiple affixes may indicate the relative ease of identifying a higher-level morphological hierarchy in complex words with multiple segmentation cues.

An analysis of the subset of words with exactly one affix (9790 fixations) showed that more productive affixes (i.e., affixes that occur in more word types) came with shorter gaze durations ($\hat{\beta} = -0.009, t(9790) = -6.403, p < 0.001$; effect size = -15 ms, model not shown). This result converges with lexical decision studies in Finnish (cf., Bertram, Laine & Karvinen, 1999) reporting shorter RTs for derived words with more productive affixes than for words with unproductive affixes.

Orthographic and Visuo-Motor Variables. Compound length (*WordLength*) went hand in hand with shorter first fixations (-37 ms) and with longer gaze durations (26 ms). This trade-off between the number and duration of fixations in correlation with word length is well-attested in the eye-movement literature (cf., Vergilino-Perez *et al.*, 2004 and references therein). Compounds with longer left constituents (*LeftLength*) elicited longer first fixations and subgaze durations for left constituents, which is as expected. In subgaze durations for the right constituents and gaze durations, the effect of left constituent length appeared to be reverse: *LeftLength* correlated negatively with durations. However, since we set the maximum for compound length, longer left constituents implied shorter right constituents. So the longer the compound's left constituent, the shorter its right constituent, and the faster it takes to complete the visual uptake of the right constituent (hence shorter subgaze duration for the right constituent), which is in line with the direction of the corresponding effect for the left constituent length.

At first fixation, the nonlinear effect of fixation position on fixation duration showed the inverse-U shape (see the linear term *FixPos* and the quadratic term *FixPos2* in Table 3.5). The fixations between the 4th and the 5th character (i.e., the position of the displayed fixation point in our experiment) had a longer duration (on average by about 70 ms) than did fixations at the word's extremes, the first and the twelfth character of the stimuli. This Inverted-Optimal Viewing Position effect is well attested in the literature on eye-movements for single word recognition and sentential reading (for an overview of available theoretical accounts see Vitu, Lancelin & d'Unienville, 2007). Initial fixation position did not interact with any predictors of our interest.

Other Control Variables. We observed longitudinal effects of the course of the experiment on participants' performance. The more the participants progressed into the experiment (as measured by the position of trial in the experimental list), the shorter their first fixations were (effect size = -9 ms), and their gaze durations were also shorter (effect size = -8 ms). In other words, the eye-movement record, just as the lexical decision latencies, shows that participants become familiarized with the task as the experiment proceeds, in line with e.g., Meeuwissen, Roelofs and Levelt (2003) and De Vaan *et al.* (2007).

The longer the lexical decision latency to the immediately preceding trial was (*RT1*), the longer the first fixations were (effect size = 51 ms). Longer *RT1* also came with a substantial lengthening of gaze duration (effect size = 282 ms). The "spillover" effect on the current trial of the processing difficulty of the preceding trial is noticeable not only in the visual lexical decision latencies, but apparently co-determines the entire time-course of morphological processing starting from the first fixation onwards. There may be two components to the effect of the RT on the preceding trial. First, this effect may reflect the spillover of the lexical processing load, which is clearly increased in the cases with longer *RT1*. In other words, word N-1 may still be processed even when the lexical decision has been made and word N has been presented. Second, and perhaps more likely, the dynamics of going through the experiment may be such that the local processing speed at word N adapts to the speed developed at previous trials (in our case, the immediately preceding trial). Being fast in a recent decision-making and motoric action of the lexical decision may influence the availability of resources and expected speed of processing for the current trial (regardless of the actual lexical characteristics of the currently presented word). We leave disentangling these possibilities to further research. Yet we note that neglecting this predictor in the statistical analysis may have profound consequences. For instance, when *RT1* was removed from the statistical model for gaze durations, the amount of variance explained by the fixed effects dropped by 1.3% percent. From a methodological perspective, bringing longitudinal and local effects in the course of the experiment may be crucial for coming to a proper understanding of the data (cf., De Vaan *et al.*, 2007; Kinoshita & Mozer, 2006; Taylor & Lupker, 2006).

General Discussion

This study primarily addressed the role of morphological structure in compound recognition. This section begins with a summary of findings, then we elaborate on the methodology of this study, and finally, we formulate requirements for a model of compound processing which would account for the present set of results.

To explore computation for multiply complex words, we considered a range of diagnostic measures traditionally interpreted as indicating decompositional processing. In our data, we observed facilitatory effects of the left and right constituent lemma frequencies, as well as the facilitatory effects of the left and right constituent family sizes. In addition, we found facilitatory effects of the compound lemma frequency, the traditional hallmark for non-decompositional processing.

The time-course of all these effects was tied to the time-course and direction of reading. Properties associated with the left constituent played a role in the early measures of eye movements, while the role of the right constituent emerged relatively late (cf., Hyönä *et al.*, 2004). Moreover, the effect sizes observed for the right constituent were considerably smaller.

The constituent frequency and family size effects may have arisen at the level of form processing, at the level of semantic processing, or possibly at both levels. At the level of form, the effect of a constituent's frequency may reflect the reader's experience with identifying that constituent's string of characters. The effect of morphological family may tap into a reader's more specific experience with parsing out and recognizing the constituent as part of a larger word. At the level of word meaning, a constituent's frequency may gauge the ease of access to its meaning. A constituent's family size would then estimate the resonance that activation of a constituent morpheme gives rise to in its morphological family.

The effect of compound frequency emerged already at the first fixation duration, a point in time when most compounds have not yet been fully scanned. There are several ways in which this surprising effect can be interpreted. This full-form frequency effect may result from unstructured form processing in which the available visual input at the first fixation (the initial characters, the previewed characters in the middle of the word, as well as the word's length, cf., Pollatsek & Rayner, 1982; Rayner, Well, Pollatsek & Bertera, 1982) is matched against stored form representations. The more entrenched this full-form representation is, the earlier the benefits of its availability emerge in the eye-movement record. Importantly, this interpretation presupposes that full-form representations do not require full visual inspection of the input and may be accessed on the basis of partially matching information (cf., de Almeida & Libben, 2002). The fact that the effect of compound frequency is also visible in later measures implies that the full-form representation of a compound is actively involved in the process of compound recognition even when other sources of lexical information become available, possibly for checking the new input for consistency with the already activated full-form and/or deactivating other competitors in the morphological family.

It is unlikely, however, that unstructured form processing would fully account for the compound frequency effect and especially for its presence in the late eye movement measures. The compound frequency effect survives inclusion in the statistical model of the frequency of the initial quadrogram summed over words that match the target compound in length (model not shown). This indicates that it is unlikely that the compound frequency effect can be reduced specifically to the earliest available visual information. Following Wurm, Aycock and Baayen (2008), it is conceivable that full-form frequency effects reflect, at least in part, memory traces of constituent morphemes having been combined together into one lexical unit. The higher the frequency of a complex word in language, the stronger the association between that word and its morphemes, and the more experience the reader has with integrating a given morpheme into that embedding word. If so, a high-frequency compound may benefit more from identification of one of its constituents than a low-frequency compound.

At the present stage of our knowledge, we cannot exclude that the compound frequency effect is also indicative of facilitation from semantic processing, given that semantic effects have been observed for very short initial time spans (cf., e.g., Diependaele, Grainger & Sandra, 2005; Hauk & Pulvermüller, 2004; Penolazzi *et al.*, 2007; cf., also Baayen, Feldman & Schreuder, 2006, for evidence concerning a strong semantic component to the word frequency effect).

In addition to constituent frequency and family size effects, and in addition to the compound frequency effect, we obtained ample evidence for a role of morphemes that are embedded inside the immediate constituents of compounds. Thus, embedded affixes that are more productive elicited shorter gaze durations, as expected given previous studies of bimorphemic derivations (cf., e.g., Bertram *et al.*, 1999). We also observed that compounds embedded in compounds require more reading time than derivations embedded in compounds. We have two possible explanations for that. First, compounds with three free-standing lexemes are more difficult to integrate semantically than those with two such lexemes. For instance, readers need to determine whether a compound with three lexemes is left-branching (i.e., the first two constituents modify the third, as in *voet-bal+bond* "football association") or right-branching (i.e., the first constituent is a modifier of the two latter constituents, as in *zaal+voet-bal* "indoor football"). Second, the derivational morpheme may have served as a parsing cue to identification of immediate constituents, and using such cues allows faster access to morphological constituents and faster semantic wrap-up of the complex word (see Chapter 4 for a more detailed discussion on this issue).

Methodological considerations. A comparison of the results obtained with the visual lexical decision task and those obtained with the cumulative eye-movement measures (subgaze and gaze durations) show remarkable convergence. In the RTs, just like in eye movements, we observe facilitatory effects of constituent frequencies and family sizes, and also those of compound frequencies. We also find qualitatively similar interactions between morphological predictors (*WordFreq* by *LeftFreq*, and *ResidLeftFamSize* by *RightFreq*) in lexical decision latencies and eye-movement durational measures. Furthermore, embedded morphemes and experimental control variables give rise to very similar patterns of results in the two datasets, lexical decision latencies and eye-movements. What the analysis of the eye-movements adds is detailed information about the time-course of morphological processing, including the early and lingering compound frequency effect, the early left constituent family size effect, and the temporal sequence of the effects pertaining to the compounds' left and right constituents.

Our choice of investigating the processing of isolated existing and nonce compounds in visual lexical decision has offered us both advantages and disadvantages. The main advantage of using isolated words is the ability to collect large numbers of data points from the same subject relatively quickly. As a result, our statistical analyses enjoy the benefit of enhanced power. In addition, combining lexical decision, the task that has been used most intensively to study morphological processing, with eye-tracking allows us to evaluate to what extent the two paradigms converge (cf., Grainger's (2003) program of investigating functional overlap between tasks). As noted above, there is indeed remarkable convergence in our data.

Our choice for using isolated words in lexical decision also comes with several disadvantages, most of which concern the issue of the ecological validity of our results. In single word reading, there is no parafoveal preview from the preceding word, and there is no natural spillover effect from the target word to the next word to be investigated. More importantly, lexical decision may induce rather different kinds of processing strategies than those used for the natural integration of word meaning into the sentence and discourse.

Another methodological decision that we had to make is whether to include a look away point on the screen, that is, whether to instruct participants to complete their lexical decision task by fixating either the word "Yes" or the word "No", which would be displayed in two different areas on the screen equally distant from the area where the target word was displayed (for the full description of this technique, see Hyönä, Laine & Niemi, 1995). For compatibility with the existing body of literature, we stayed as close to the conventional lexical decision paradigm as possible, and did not make use of such a look away point. Instead, we considered in our analyses only those fixations that were completed before the button press registering a lexical decision. The price we pay is the possibility of some more noise in the eye-movements measures, especially in the gaze durations. Yet in our data, gaze durations and RTs are not that highly correlated: $R^2 = 0.46$ only. Thus, both gaze durations and RTs serve as dependent variables in their own right.

We also note that the presence of nonce compounds and of many low-frequency existing compounds in our experiment may have enhanced decompositional processing and inhibited full-form processing. In the light of this possibility, it is all the more surprising that an effect of compound frequency is observed at the very first fixation.

Whatever the disadvantages of our methodology may be, the pattern of results that we have obtained and reported either in the body of the paper or in Appendix 1, dovetails perfectly with many of the results obtained in the literature for sentential reading, such as visuo-oculomotor effects (cf., e.g., O'Regan *et al.*, 1994; Rayner, 1998; Vitu, McConkie, Kerr & O'Regan, 2001), effects of compound length and frequency, as well as of constituent frequencies (cf., e.g., Andrews *et al.*, 2004; Duñabeitia, Perea & Carreiras, 2007; Hyönä & Pollatsek, 1998; Hyönä *et al.*, 2004; Juhasz *et al.*, 2003; Taft & Forster, 1976), and effects of orthographic n-grams (reported in Appendix 1, cf., e.g., Lima & Inhoff, 1985). Furthermore, in a recent sentential reading study (Chapter 4), in which Finnish compounds were embedded in context, a highly similar pattern of results was observed, including early effects of compound frequency, left constituent frequency and family size, later and weaker effects of right constituent frequency and family size, interactions between morphological predictors, as well as longitudinal experimental effects.

Towards a theory of compound processing

According to Libben (2006), readers and listeners maximize their opportunities for comprehension by the simultaneous use of all processing cues available to them, and all processing mechanisms that they have at their disposal, including retrieval from memory and compositional computation. The present study provides support for Libben's hypothesis of maximization of opportunity. All constituent morphemes, the whole compound itself and morphological families that share one of the compound's constituents play a noticeable role in lexical processing of compounds. This indicates that there are multiple routes at work in compound processing, and readers use these routes interactively, at different times and to a different extent, to efficiently and accurately recognize compounds. The early compound frequency effect shows that readers do not wait for all the characters of the word to be seen before making inferences about the word's identity. The early compound frequency effect also shows that readers do not gain access to compound representations only after having accessed its constituents. The interactions of morphological predictors (compound frequency by left constituent frequency and left constituent family size by right constituent frequency) show that the cues modulate each other, and that decompositional processes and full-form driven processes are not independent. Using one kind of morphological information for compound identification as if other sources of information do not exist amounts to missing out on the cumulative use of informations and on concomitant facilitation of performance.

In what follows, we take as the point of departure the basic assumption of parallel dual route models, given the evidence in our data for both processing routes. As the detailing of a full-fledged model of morphological processing is beyond the scope of this study, we restrict ourselves to listing a number of requirements that are not satisfied by the current parallel dual route models proposed in the literature (e.g., Schreuder & Baayen, 1995). While our results were obtained in the visual domain, we believe that the requirements outlined below would equally hold for the models of the auditory processing of compounds.

First, current models of morphological processing almost always discuss complex words as if they are read with only one fixation. An example of a model that addresses the temporal dynamics of reading complex words is the one proposed by Pollatsek, Reichle and Rayner (2003), and they conclude that a parallel dual route architecture is unable to approximate the empirical data, unless the two routes of lexical processing are allowed to interact. It is clear, also from the present data, that the details of the time-course of information entering the system needs to be explicitly included in models of morphological processing in reading.

In the typical left-to-right reading of long compounds the very first opportunities for comprehension of the compound present themselves already during parafoveal preview, when information about the initial characters and word length becomes available (Rayner *et al.*, 1982). In single-word reading, this information is also available very early, during the low-level attentional scan of the word that occurs in the beginning of fixation, cf., Reichle, Rayner and Pollatsek, 2003. Following Clark and O'Regan (1999) and O'Regan (1979), word length may play a disambiguating role in word recognition (for the opposing view, see Inhoff & Eiter, 2003). For words embedded in the sentential context, additional information may come from contextual predictability (e.g., Ehrlich & Rayner, 1981), collocational strength (e.g., McDonald & Shillcock, 2001) and constructional cues (e.g., Frazier *et al.*, 2006).

The next opportunities for restricting the range of possible interpretations for the visual input arise at the first fixation, where a range of properties of the first constituent come into play: not only the frequency of the left constituent, its length, and its morphological family, but also the combinatorial likelihood of morphemes within the whole compound, in conjunction with information about the compound's length. Later opportunities (at second and subsequent fixations) include properties of the right constituent. New information obtained at this stage is processed against the backdrop of the information already extracted about the word.

Second, models of morphological processing in reading need to allow for a simultaneous processing of information at different levels without requiring strict sequentiality of processing stages, as witnessed, for instance, by the simultaneous early effects in our data of compound frequency, left constituent frequency and family size, and orthographic n-gram effects[7]. Our results challenge sublexical models, which allow full-form access only after morphological constituents have been recognized (cf., Pinker, 1999; Taft, 2004; Taft & Forster, 1976; Taft, 1991). Our results also challenge supralexical models, which only allow constituents to come into play after the compound as a whole has been recognized (Giraudo & Grainger, 2001).

Third, models of compound processing should allow for the modulation of the weight of one opportunity by the presence and strength of other opportunities, as witnessed by the interaction of compound frequency and left constituent frequency (for related discussion of cue trade-offs in speech processing see e.g., Mattys, White & Melhorn, 2005; McClelland & Elman, 1986). Current parallel dual route models tend to simplify morphological processing to activation of autonomous lexical representations that are blind to each other's activation (cf., Laudanna & Burani, 1985; Frauenfelder & Schreuder, 1991, and Schreuder & Baayen, 1995; see however Baayen & Schreuder, 2000). In general, the fact that we find, also in the parallel study, early constituent frequency effects and whole-word frequency effects at the same time, tells us that one cue or route is not cancelling out the other completely, a prediction that would directly derive from a strict dual route model. Depending on the strength of the available cues, the fine-tuning of this kind of co-operative system depends on the specific properties of the complex word.

Fourth, models of morphological processing should come to grips with fast activation of morphological paradigms (families) associated with a compound's constituents. One important constraint on morphological models is our finding that left constituent families are activated immediately upon access to those constituents, and not after full-form access.

Effectively, a model that meets these requirements is no longer a dual route model, but rather a multiple route model that, in morphological terms, allows access to full-forms, immediate constituents, embedded morphemes and

[7] A modeling framework that may prove to be useful here is the hierarchical temporal memory framework proposed by Hawkins and George (2006), see also Hawkins and Blakeslee (2004). In the hierarchical temporal memory framework, the simultaneous processing would be accomplished by generation skip, i.e., lower-level detectors in the hierarchy propagating information about the input to higher levels, skipping intermediate levels.

morphological families. More generally, such a model will have as its basic principle maximization of all opportunities, both morphological, orthographic, phonological, and contextual, for comprehension of the visual input. We believe that probabilistic and information-theoretical approaches to lexical processing developed recently in morphological and syntactic research (cf. e.g., Moscoso del Prado Martín *et al.*, 2004; Levy, 2008) hold promise for formalization of those opportunities and for computational implementation of the multiple-route model of compound recognition.

References

Allen, M. and Badecker, W. (2002). Inflectional regularity: Probing the nature of lexical representation in a cross-modal priming task. *Journal of Memory and Language*, 46:705–722.

Andrews, S. (1986). Morphological influences on lexical access: lexical or non-lexical effects? *Journal of Memory and Language*, 25:726–740.

Andrews, S., Miller, B., and Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, 16(1/2):285–311.

Baayen, R. H. (2007). Storage and computation in the mental lexicon. In Jarema, G. and Libben, G., editors, *The Mental Lexicon: Core Perspectives*, pages 81–104. Elsevier.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge University Press, Cambridge.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *In press*.

Baayen, R. H., Feldman, L. B., and Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55:290–313.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Baayen, R. H. and Schreuder, R. (1999). War and peace: morphemes and full forms in a non-interactive activation parallel dual route model. *Brain and Language*, 68:27–32.

Baayen, R. H. and Schreuder, R. (2000). Towards a psycholinguistic computational model for morphological parsing. *Philosophical Transactions of the Royal Society (Series A: Mathematical, Physical and Engineering Sciences)*, 358:1–13.

Baayen, R. H., Wurm, L. H., and Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *Mental Lexicon*, 2:419–463.

Bates, D. M. and Sarkar, D. (2005). The lme4 library. *[On-line], Available: http://lib.stat.cmu.edu/R/CRAN/*.

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

Clark, J. and O'Regan, J. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research*, 39:842–857.

Cleland, A., Gaskell, M., Quinlan, P., and Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology:Human Perception and Performance*, 32:104–119.

Cunnings, I. and Clahsen, H. (2007). The time-course of morphological constraints: Evidence from eye-movements during reading. *Cognition*, 104:476–494.

de Almeida, R. and Libben, G. (2002). Compound pre-access decomposition: Effects of constituent disruption. *Folia Linguistica*, 36:97–115.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and 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., and Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, 15:329–365.

De Vaan, L., Schreuder, R., and Baayen, R. H. (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon*, 2(1):1–24.

Diependaele, K., Sandra, D., and Grainger, J. (2005). Masked cross-modal morphological priming: unravelling morpho-orthographic and morpho-semantic inluences in early world recognition. *Language and Cognitive Processes*, 20:75–114.

Dijkstra, T., Moscoso del Prado Martín, F., Schulpen, B., Schreuder, R., and Baayen, R. (2005). A roommate in cream: Morphological family size effects on interlingual homograph recognition. *Language and Cognitive Processes*, 20:7–41.

Duñabeitia, J. A., Perea, M., and Carreiras, M. (2007). The role of the frequency of constituents in compound words: Evidence from Basque and Spanish. *Psychonomic Bulletin & Review*, 14:1171–1176.

Ehrlich, S. F. and Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20:641–655.

Frauenfelder, U. H. and Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In Booij, G. E. and Marle, J. v., editors, *Yearbook of Morphology 1991*, pages 165–183. Kluwer Academic Publishers, Dordrecht.

Frazier, L., Carminati, M., Cook, A., Majewski, H., and Rayner, K. (2006). Semantic evaluation of syntactic structure: Evidence from eye movements. *Cognition*, 99:B53–B62.

Giraudo, H. and Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review*, 8:127–131.

Grainger, J. (2003). Moving eyes and reading words: How can a computational model combine the two. In Hyönä, Y., Radach, R., and Deubel, H., editors, *The mind's eye: Cognitive and applied aspects of eye movement research*. Elsevier, Amsterdam.

Hauk, O. and Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115:1090–1103.

Hawkins, J. and Blakeslee, S. (2004). *On Intelligence*. Henry Holt and Co, New York.

Hawkins, J. and George, D. (2006). Hierarchical temporal memory. concepts, theory and terminology. *[On-line], Available:*

http://www.numenta.comtechnology.php.

Hyönä, J., Bertram, R., and Pollatsek, A. (2004). Are long compound words identified serially via their constituents? Evidence from an eye-movement-contingent display change study. *Memory and Cognition*, 32:523–532.

Hyönä, J., Laine, M., and Niemi, J. (1995). Effects of a word's morphological complexity on readers' eye fixation pattern. In Findlay, J. M., Kentridge, R., and Walker, R., editors, *Eye movement research: mechanisms, processes and applications*, pages 207–231. Elsevier, Amsterdam.

Hyönä, J. and Pollatsek, A. (1998). Reading finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance*, 24:1612–1627.

Inhoff, A. and Eiter, B. (2003). Knowledge of word length does not constrain word identification. *Psychological Research*, 67:1–9.

Juhasz, B. (2007). The influence of semantic transparency on eye movements during english compound word recognition. In Van Gompel, R., Fischer, M., Murray, W., and Hill, R., editors, *Eye Movements: A Window on Mind and Brain*, pages 373–389. Elsevier, Amsterdam.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., and Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, 94:223–244.

Kinoshita, S. and Mozer, M. (2006). How lexical decision is affected by recent experience: Symmetric versus asymmetric frequency-blocking effects. *Memory and Cognition*, 34:726–742.

Krott, A., Baayen, R. H., and Schreuder, R. (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics*, 39(1):51–93.

Kuperman, V., Bertram, R., and Baayen, R. H. (2008). Morphological dynamics in compound processing. *In press, Language and Cognitive Processes*.

Laudanna, A. and Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 345–364. Lawrence Erlbaum Associates, Hillsdale, N. J.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177.

Libben, G. (2006). Why study compound processing? In Libben, G. and Jarema, G., editors, *The representation and processing of compound words*, pages 1–23. Oxford University Press, Oxford.

Lima, S. D. and Inhoff, A. (1985). Lexical access during eye fixations in reading: Effects of word-initial letter sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 11:272–285.

Mattys, S., White, L., and Melhorn, J. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134:477–500.

McClelland, J. L. and Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18:1–86.

McDonald, S. and Shillcock, R. (2001). Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing. *Language and Speech*, 44(3):295–323.

McDonald, S. and Shillcock, R. (2004). The potential contribution of preplanned refixations to the preferred viewing location. *Perception and Psychophysics*, 66(6):1033–1044.

Meeuwissen, M., Roelofs, A., and & Levelt, W. J. M. (2003). Planning levels in naming and reading complex numerals. *Memory & Cognition*, 31:1238–1248.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., and Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30:1271–1278.

Niswander-Klement, E. and Pollatsek, A. (2006). The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Memory and Cognition*, 34:685–702.

O'Regan, J. (1979). Saccade size control in reading: Evidence for the linguistic control hypothesis. *Perception and Psychophysics*, 25:501–509.

Penolazzi, B., Hauk, O., and Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 74:374–388.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Statistics and Computing. Springer, New York.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Weidenfeld and Nicolson, London.

Pollatsek, A. and Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, 20:261–290.

Pollatsek, A., Hyönä, J., and Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26:820–833.

Pollatsek, A. and Rayner, K. (1982). Eye-movement control in reading - the role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 8:817–833.

Pollatsek, A., Reichle, E., and Rayner, K. (2003). Modeling eye movements in reading: Extensions of the E-Z Reader model. In Hyönä, Y., Radach, R., and Deubel, H., editors, *The mind's eye: Cognitive and applied aspects of eye movement research*, pages 361–390. Elsevier, Amsterdam.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422.

Rayner, K., Well, A., Pollatsek, A., and Bertera, J. (1982). The availability of useful information to the right of fixation in reading. *Perception and Psychophysics*, 31:537–550.

Reichle, E., Reyner, K., and Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26:445–526.

Schreuder, R. and Baayen, R. H. (1995). Modeling morphological processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 131–154. Lawrence Erlbaum, Hillsdale, New Jersey.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96:523–568.

Sereno, S. C., Rayner, K., and Posner, M. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *Neuroreport*, 9:2195–2200.

Taft, M. (1991). *Reading and the mental lexicon*. Lawrence Erlbaum, Hove, U.K.

Taft, M. and Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14:638–647.

Taft, M. and Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 15:607–620.

Taylor, T. E. and Lupker, S. J. (2006). Time perception and word recognition: An elaboration of the time-criterion account. *Perception and Psychophysics*, 68:933–945.

Vergilino-Perez, D., Collins, T., and Doré-Mazars, K. (2004). Decision and metrics of refixations in reading isolated words. *Vision Research*, 44:2009–2017.

Vitu, F., Lancelin, D., and Marrier d'Unienville, V. (2007). A perceptual-economy account for the inverted-optimal viewing position effect. *Joural of Experimental Psychology: Human Perception and Performance*, 33:1220–1249.

Vitu, F., McConkie, G., Kerr, P., and O'Regan, J. (2001). Fixation location effects on fixation durations during reading: An inverted optimal viewing position effect. *Vision Research*, 41:3513–3533.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, 9:341–368.

Appendix 1

Table 3.3: Summary of Continuous Variables

Variable	Range (Adjusted Range)	Mean(SD)	Median
RT	270:2208 ms (5.6:7.7 log units)	6.7(0.3)	6.6
InitPos	0.1:11.9 characters (1:119 pixels)	40.7(8.4)	40
FirstDur	50:1200 ms (3.9:7.1 log units)	5.6(0.4)	5.6
SubgazeLeft	60:1808 ms (4.1:7.5 log units)	5.8(0.5)	5.7
SubgazeRight	82:1097 ms (4.4:7.0 log units)	5.6(0.4)	5.5
GazeDuration	50:2208 ms (3.9:8.2 log units)	6.5(0.5)	6.5
TrialNum	1:2500	12.0(7.2)	12
RT1	148:4023 ms (5.0:8.3 log units)	6.73(0.3)	6.7
WordLength	8:12 characters	11.6(0.7)	12
LeftLength	2:10 characters	5.4(1.6)	5
FinTrigram	1:984609 (0:13.8 log units)	9.6((2.6)	9.9
WordFreq	3:2207 (1.1:7.7 log units)	2.2(1.1)	1.9
LeftFreq	1:24343 (0.0:10.1 log units)	5.0(2.9)	5.4
RightFreq	1:49020 (0:10.8 log units)	4.5(3.0)	4.2
ResidLeftFamilySize	3:298 (-2.3:3.7)	0.0(1.0)	0.0
ResidRightFamilySize	3:270 (-3.5:7.4)	0.0(1.1)	-0.1
AffixProd	3:6002 (0.7:8.7 log units)	6.8(1.3)	6.99
Complexity	3:6 morphemes	3.2(0.4)	3

Numbers in the second column show original value ranges for predictors. If any transformations have been made with the original values for statistical reasons (i.e., natural log transformation, decorrelation with other predictors or scaling), the numbers in the brackets show the ranges actually used in statistical models. Means, standard deviations and median values refer to the predictor values used in the models. Values for frequency and family size measures are based on the corpus with 42 million word-forms.

Key to Table 3.3: Predictors of primary interest for this study are presented in the main body of paper. Additional control variables that show significant effects in our statistical models are as follows: *Correct1*, the binary indicator of whether the previous trial was a correct lexical decision; *FixPos* and *FixPos2*, first fixation position and its squared value; *FinTrigram*, frequency of the word-final trigram; and *Nomore*, indicator of whether the fixation is word-final. In addition to these, we have considered a large number of control variables that were not significant predictors of reading times, fixation probabilities or lexical decision latencies. These included variables listed in the subsection Dependent variables as well as initial trigram frequency, mean bigram frequency of the word, position of the minimal bigram, affix

Table 3.4: Fixed Effects of the Model for Lexical Decision RT for Existing Compounds

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.9740	5.9771	5.8176	6.1336	0.001	0.0000
WordLength	0.0148	0.0149	0.0083	0.0226	0.002	0.0000
LeftFreq	-0.0181	-0.0183	-0.0250	-0.0115	0.001	0.0000
RightFreq	-0.0095	-0.0096	-0.0129	-0.0060	0.001	0.0000
Complexity	0.0639	0.0634	0.0302	0.0953	0.001	0.0002
Trial	-0.0041	-0.0042	-0.0050	-0.0032	0.001	0.0000
RT1	0.1288	0.1286	0.1144	0.1413	0.001	0.0000
Correct1Y	-0.0160	-0.0159	-0.0285	-0.0031	0.012	0.0146
ResidLeftFamSize	0.0114	0.0111	-0.0106	0.0353	0.354	0.3557
ResidRightFamSize	-0.0122	-0.0121	-0.0194	-0.0049	0.001	0.0010
AffixFinal	-0.0527	-0.0526	-0.0796	-0.0295	0.001	0.0001
AffixInitial	-0.0178	-0.0169	-0.0613	0.0339	0.500	0.4801
AffixMedial	-0.0382	-0.0378	-0.0653	-0.0116	0.006	0.0062
AffixMultAffix	-0.0897	-0.0887	-0.1371	-0.0394	0.001	0.0004
WordFreq	-0.0717	-0.0722	-0.0904	-0.0533	0.001	0.0000
LeftFreq:WordFreq	0.0037	0.0037	0.0012	0.0062	0.002	0.0047
RightFreq:ResidLeftFamSize	-0.0049	-0.0048	-0.0079	-0.0017	0.001	0.0040

length, branching of triconstituents, and frequencies of deeply embedded stems in triconstituents.

Specifications of statistical models, Tables 3.4-3.9

Specifications include estimates of the regression coefficients; 95% highest posterior density intervals (HPDs), which are a Bayesian estimate of the most likely values of a parameter, roughly comparable to traditional 95% confidence intervals; p-values estimated by the Monte Carlo Markov chain (MCMC) method; and p-values obtained with the t-test for fixed effects using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom (see Pinheiro & Bates, 2000 for discussion of the method). We also report the estimated standard deviations for each random intercept (e.g., *Subject* or *Word*) and each random slope (e.g., *Subject* by *WordLength*), together with the estimates based on the MCMC samples and HPD intervals, such as the MCMC mean and 95% HPDs (Table 3.9 for all models), see Pinheiro and Bates (2000) for detailed treatment of random effects in mixed-effects models.

Computation of effect sizes

Effect sizes were estimated as follows. For factors, for which we used contrast coding, effect size was calculated as the difference between (i) the sum of the intercept and the contrast coefficient, $\hat{\beta}$, and (ii) the intercept. For log-transformed dependent variables (fixation duration, gaze duration, RT), effect sizes were calculated for back-transformed values, so that effect sizes are reported in ms. Effect sizes for simple main effects of a covariate were calculated as the difference

Table 3.5: Model for First Fixation Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.8489	5.8532	5.5993	6.1337	0.001	0.0000
NomoreTRUE	0.2345	0.2356	0.1773	0.3067	0.001	0.0000
WordLength	-0.0394	-0.0390	-0.0575	-0.0202	0.001	0.0000
LeftLength	-0.0261	-0.0260	-0.0304	-0.0209	0.001	0.0000
FixPos	0.0088	0.0088	0.0065	0.0113	0.001	0.0000
FixPos2	-0.0001	-0.0001	-0.0001	0.0000	0.001	0.0000
WordFreq	-0.0347	-0.0346	-0.0490	-0.0171	0.001	0.0001
LeftFreq	-0.0172	-0.0172	-0.0231	-0.0115	0.001	0.0000
ResidLeftFamSize	0.0728	0.0733	0.0001	0.1559	0.058	0.0690
Trial	0.0000	0.0000	0.0000	0.0000	0.001	0.0000
RT1	0.0352	0.0348	0.0168	0.0515	0.001	0.0001
WordFreq:LeftFreq	0.0031	0.0031	0.0010	0.0052	0.010	0.0058
WordLength:ResidLeftFamSize	-0.0085	-0.0086	-0.0156	-0.0019	0.018	0.0171

Table 3.6: Model for Subgaze Duration for the Left Constituent

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.9312	5.9380	5.7110	6.1295	0.001	0.0000
WordLength	-0.0628	-0.0627	-0.0729	-0.0529	0.001	0.0000
LeftLength	0.0777	0.0774	0.0687	0.0856	0.001	0.0000
WordFreq	-0.0591	-0.0598	-0.0846	-0.0341	0.001	0.0000
LeftFreq	-0.0272	-0.0275	-0.0361	-0.0175	0.001	0.0000
RightFreq	-0.0028	-0.0029	-0.0070	0.0015	0.184	0.2056
ResidLeftFamSize	-0.0378	-0.0382	-0.0472	-0.0280	0.001	0.0000
ResidRightFamSize	-0.0023	-0.0024	-0.0116	0.0062	0.624	0.6280
Trial	0.0000	0.0000	0.0000	0.0000	0.004	0.0026
RT1	0.0748	0.0747	0.0529	0.0988	0.001	0.0000
AffixMedial	-0.0472	-0.0468	-0.0831	-0.0151	0.008	0.0084
AffixFinal	-0.0216	-0.0218	-0.0588	0.0138	0.266	0.2556
AffixMultAffix	-0.0805	-0.0808	-0.1153	-0.0472	0.001	0.0000
WordFreq:LeftFreq	0.0052	0.0053	0.0017	0.0085	0.001	0.0019

Table 3.7: Model for Subgaze Duration for the Right Constituent

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.6285	5.6278	5.3264	5.9335	0.001	0.0000
WordLength	0.0289	0.0289	0.0139	0.0408	0.001	0.0000
LeftLength	-0.1105	-0.1105	-0.1211	-0.0994	0.001	0.0000
WordFreq	-0.0340	-0.0339	-0.0444	-0.0244	0.001	0.0000
LeftFreq	-0.0016	-0.0016	-0.0068	0.0037	0.550	0.5612
RightFreq	-0.0103	-0.0103	-0.0167	-0.0043	0.001	0.0009
ResidLeftFamSize	-0.0154	-0.0153	-0.0296	-0.0027	0.028	0.0292
ResidRightFamSize	-0.0188	-0.0188	-0.0317	-0.0074	0.001	0.0052
Trial	0.0000	0.0000	0.0000	0.0000	0.040	0.0528
RT1	0.1029	0.1032	0.0676	0.1364	0.001	0.0000
AffixMedial	-0.0598	-0.0594	-0.1149	-0.0120	0.020	0.0210
AffixFinal	-0.1022	-0.1016	-0.1513	-0.0529	0.001	0.0000
AffixMultAffix	-0.0005	-0.0010	-0.0520	0.0486	0.966	0.9852

Table 3.8: Model for Gaze Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.6415	5.6385	5.4218	5.8684	0.001	0.0000
WordLength	0.0173	0.0173	0.0032	0.0319	0.012	0.0174
LeftLength	-0.0173	-0.0173	-0.0291	-0.0061	0.002	0.0029
WordFreq	-0.0912	-0.0908	-0.1194	-0.0621	0.001	0.0000
LeftFreq	-0.0253	-0.0253	-0.0354	-0.0142	0.001	0.0000
RightFreq	-0.0080	-0.0081	-0.0132	-0.0026	0.008	0.0054
ResidLeftFamSize	0.0073	0.0075	-0.0291	0.0446	0.672	0.6975
ResidRightFamSize	-0.0070	-0.0072	-0.0177	0.0025	0.164	0.2102
Trial	0.0000	0.0000	-0.0001	0.0000	0.001	0.0000
RT1	0.1506	0.1509	0.1273	0.1752	0.001	0.0000
AffixMedial	-0.0812	-0.0798	-0.1201	-0.0400	0.001	0.0001
AffixFinal	-0.1001	-0.0985	-0.1413	-0.0585	0.001	0.0000
AffixMultAffix	-0.0834	-0.0826	-0.1250	-0.0470	0.001	0.0001
FinTrigram	-0.0070	-0.0071	-0.0124	-0.0015	0.012	0.0185
RightFreq:ResidLeftFamSize	-0.0068	-0.0068	-0.0119	-0.0018	0.008	0.0087
WordFreq:LeftFreq	0.0053	0.0053	0.0016	0.0091	0.008	0.0055

Table 3.9: Random effects for *RT*, *FirstDur*, *SubgazeLeft*, *SubgazeRight* and *GazeDur*

A. Lexical decision latency				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.095	0.095	0.090	0.101
Subject	0.151	0.155	0.110	0.215
Residual	0.241			
B. First fixation duration				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.049	0.050	0.042	0.057
Subject	0.42	0.415	0.286	0.608
Subject by Nomore	0.099	0.098	0.070	0.148
Subject by WordLength	0.035	0.035	0.024	0.051
Residual	0.289			
C. Subgaze duration for the left constituent				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.088	0.087	0.078	0.096
Subject	0.114	0.12	0.087	0.167
Residual	0.335			
D. Subgaze duration for the right constituent				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.010	0.097	0.075	0.116
Subject	0.107	0.110	0.079	0.158
Residual	0.456			
E. Gaze duration				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.014	0.122	0.113	0.132
Subject by LeftLength	0.015	0.014	0.008	0.023
Subject by WordLength	0.017	0.018	0.012	0.025
Subject	0.082	0.022	0.001	0.172
Residual	0.386			

between the model's predictions for the minimum and maximum values of that covariate.

Comparison between effect sizes of numeric variables obtained in our study and in previous studies that set these variables to a discrete number of levels for factorial designs is not straightforward. Our estimates are defined over the entire range of values of the variable, while a factorial contrast is defined as a difference between group means, where groups are formed (in the simplest case) by dichotomization of a given predictor. The best approximation to factorial estimates is one half of our effect sizes, which is equivalent (for linear effects) to the factorial contrast where the variable of interest is dichotomized and where the group means are positioned at the first and the third quartiles. Obviously, factors do not pose such a problem and are directly comparable across reports.

Appendix 2

In the present study, we chose to present readers with a fixed order of items in the block and the fixed order of blocks, so that each reader saw the words in the same order (even though that one order was set randomly). We hypothesized that by using one list order we would have tighter experimental control, especially as we have the position of an item in the experimental list as a covariate in the model, so that longitudinal effects of practice or fatigue are modeled explicitly. By using the fixed list order we also attempted to avoid the increase in between-participant variance, which derives from the random ordering of items across participants. By that, we aimed at gaining increased statistical power. Using the fixed list order, however, goes against the common practice of counterbalancing (or otherwise randomizing) the presentation order of items across participants. The problem that is usually claimed to follow from using the fixed item order is that the variance that one attributes to a predictor of interest might in fact be due to the influence of the item order. In other words, the item order is a potential confound for estimates of other effects.

It turns out that the item order is no a priori reason for worry. Linear mixed effects models are very well able to disentangle the various sources of variance for a design such as we used. In a simulation study, we considered a repeated measures design with 20 participants, 1000 items, list position (the rank or trial number in the list) as a predictor, and five predictors specifying properties linked to the items (standing for word length, word frequency, left constituent family size, left constituent and right constituent frequencies). In other words, the simulated data have the same design as our experimental data, albeit with fewer predictors and fewer items. The question of interest is whether the mixed-effects modeling algorithm can adequately separate the different sources of variance under two conditions, one in which each participant is exposed to the items in the same order (as in our manuscript) and one in which each participant is exposed to the items in a different random order. This simulation does not aim to assess the significance of our predictors or to validate our statistical models. Rather we simulate two types of experimental designs (with a fixed number of items, participants and predictors) to see whether, under these conditions, predictions of statistical models would differ across designs.

More formally, let i index participants, j index items, and k index trial number. Furthermore, let X_1 denote *Trial*, X_2 to X_6 item-bound properties, and β_{0-6} the intercept and regression coefficients, respectively. We further denote participant random effect as b_{S_i} (normally distributed with standard deviation σ_S) and item random effect as b_{W_j} (normally distributed with standard deviation σ_W), and we denote the error term as ε_{ijk} (normally distributed with standard deviation σ). Our simulated data have the general form

$$y_{ijk} = \beta_0 + \beta_1 X_{1_k} + \beta_2 X_{2_j} + \beta_3 X_{3_j} + \beta_4 X_{4_j} + \beta_5 X_{5_j} + \beta_6 X_{6_j} + b_{S_i} + b_{W_j} + \varepsilon_{ijk}, \qquad (3.1)$$
$$b_{S_i} \sim \mathcal{N}(0, \sigma_S), b_{W_i} \sim \mathcal{N}(0, \sigma_W), \varepsilon_{ijk} \sim \mathcal{N}(0, \sigma), b_{S_i} \perp b_{W_j} \perp \varepsilon_{ijk}.$$

In building a simulation there are many choices to be made. In this simulation we make a simplifying assumption that item-bound predictors X_2 to X_6 are uncorrelated, while the predictors in the empirical data show mild collinearity. Also, it is not necessarily the case that there is a unique value for a predictor for each trial. Say, if we take word length to range from 4 to 12 characters, there will not be 1000 different values of word length for 1000 trials, rather integer values for 4 to 12 will be repeated multiple times, just like in the original data.

We distinguish between a model with fixed order for all participants, so $k = j$, henceforth $M_{k=j}$, and a model in which each participant has a different order, so $k \neq j$, henceforth $M_{k\neq j}$. We studied the behavior of both models for 20 participants and 1000 items, across 1000 simulation runs. Columns 1-3 in Table 3.10 specify which fixed and random predictors were used in the simulation, what their coding is in (3.1), and what the values were that we set for those predictors. We based the ranges of values for item-bound predictors on the actual ranges in the experimental data. Values of X_2 to X_6 varied randomly (uniformly) in the corresponding value ranges. For all predictors, only integer values were considered. Our estimates for regression coefficients and the intercept closely follow the output of the statistical model for lexical decision latencies (Table 3.4), with a few exceptions. We increased variance in data by setting higher values for random errors, and we diminished the influence of the strongest lexical predictors, word length and word frequency, by dividing their regression coefficients by 10. We increased noise and weakened some effects, because the simulation ran on the original data showed the almost perfect accuracy in estimating the coefficients, and it reported significance of predictors correctly in almost 100% of cases. Results of the simulation are summarized in Table 3.10.

Columns 4 and 6 in Table 3.10 show the means of the estimates of the coefficients for the fixed effects and for the standard deviations of the random effects, obtained with the model with fixed order of items and the model with the randomized order of items for each participant, correspondingly. Columns 5 and 7 show proportions of correctly reported significance across simulation runs for

Table 3.10: Parameters, estimates of the parameters, and power (for $\alpha = 0.05$) for the models without ($M_{k=j}$) and with ($M_{k \neq j}$) random orders of items for each participant. Averages over 1000 simulation runs.

predictor	parameter	value	$M_{k=j}$		$M_{k \neq j}$	
			estimate	power	estimate	power
Intercept	β_0	5.82	5.8074	1.00	5.8084	1.000
Trial	β_1	-0.0048	-0.0010	1.00	-0.0010	1.00
WordFreq	β_2	-0.0043	-0.0045	0.35	-0.0045	0.35
WordLength	β_3	0.0013	0.0010	0.05	0.0010	0.05
LeftFamSize	β_4	-0.0084	-0.0082	0.27	-0.0082	0.27
LeftFreq	β_5	-0.0026	-0.0027	0.21	-0.0027	0.22
RightFreq	β_6	-0.0072	-0.0072	0.88	-0.0072	0.88
Subj	σ_S	0.32	0.3100		0.3100	
Item	σ_W	0.20	0.1999		0.1999	
Resid	σ	0.60	0.5996		0.5996	

both types of models. It is evident that for large data samples, such as we used in this study, there is no appreciable difference across presentation orders in the performance of statistical models, neither in the accuracy of estimates for model coefficients or for standard deviations of random effects, nor in the power to detect the effect of the item-bound predictors. For smaller samples, we have seen cases where a single experimental list (fixed order) comes with slightly reduced power than the list with the random presentation of items.

We have carried out more simulations with different values for the fixed and random effect parameters, time and again the pattern is like the one summarized in Table 3.10. Importantly, these simulations show no a priori reason to believe that sources of variance are confounded: at least, not with the number of items and the number of uncorrelated predictors that we used here. It is important to realize that the strength of linear mixed-effects models lies precisely in their ability to 'unconfound' different sources of variance.

We have double-checked whether there were interactions of this longitudinal effect with item-bound predictors, including lexical, distributional or orthographic characteristics of compounds as whole words or their morphemes, but there were none. This gives additional assurance that the morphological effects of our primary interest are not modulated by the longitudinal effects of the experimental list. We have also investigated whether other longitudinal effects might be present (ranging from priming effects due to constituents that appeared earlier in the list to effects of sharing onset or rhyme). None turned out to be significant. In other words, the morphological and orthographic effects that we report are not artifacts nor confounds of experimental control variables, as can be demonstrated both in a simulation and in an experiment.

Morphological Dynamics in Compound Processing

Chapter 4

This chapter is an article in press: Victor Kuperman, Raymond Bertram and R. Harald Baayen. Morphological Dynamics in Compound Processing. *Language and Cognitive Processes.*

Abstract

This chapter explores the time-course of morphological processing of trimorphemic Finnish compounds. We find evidence for the parallel access to full-forms and morphological constituents diagnosed by the early effects of compound frequency, as well as early effects of left constituent frequency and family size. We also observe an interaction between compound frequency and both the left and the right constituent family sizes. Furthermore, our data show that suffixes embedded in the derived left constituent of a compound are efficiently used for establishing the boundary between compounds' constituents. The success of segmentation of a compound is demonstrably modulated by the affixal salience of the embedded suffixes. We discuss implications of these findings for current models of morphological processing and propose a new model that views morphemes, combinations of morphemes and morphological paradigms as probabilistic sources of information that are interactively used in recognition of complex words.

Introduction

Current models of morphological processing vary widely in their assumptions about what morphological information is used, and in what order, to identify and interpret complex words, for instance *dish+wash-er* or *happi-ness*. For instance, sublexical and supralexical models advocate obligatory sequentiality: The former class of models posits that full-forms can only be accessed via morphological constituents (e.g., Taft & Forster, 1975; Taft, 1979; Taft, 1991), while the latter class claims that the activation of the full-form precedes the activation of constituents (e.g., Giraudo & Grainger, 2001). Some parallel dual-route models allow for simultaneous activation of both the full-forms of complex words and their morphological constituents, but assume that the two routes proceed independently of each other (e.g., Schreuder & Baayen, 1995; Baayen & Schreuder, 1999). The computational model MATCHEK (Baayen & Schreuder, 2000) implements the interaction between the two processing routes, but is silent about the time-course of visual information uptake, and assumes that all words are read with a single fixation. The present eye-tracking study adresses the temporal unfolding of visual recognition of trimorphemic Finnish compounds, in order to establish whether the requirements posed by current models (e.g., obligatory sequentiality or independence of processing stages) hold for reading of long words. We present evidence that more sources of morphological information are at work and interacting with each other in compound processing than previously reported.

The central research issue that this chapter addresses is the hotly debated topic of the time-course of morphological effects in recognition of long compounds. It is a robust finding that full-form representations of compounds are involved in compound processing, as indicated by the effect of compound frequency (e.g., De Jong, Feldman, Schreuder, Pastizzo & Baayen, 2002; Hyönä & Olson, 1995; Van Jaarsveld & Rattink, 1988). The question that remains open, however, is how early this involvement shows up. Several studies of English and Finnish compounds found a weak non-significant effect of compound frequency as early as the first fixation on the compound (cf., Andrews, Miller & Rayner, 2004; Bertram & Hyönä, 2003; Pollatsek, Hyönä & Bertram, 2000). The presence or absence of compound frequency effects at the earliest stages of word identification may inform us about the order of activation of the full-forms of compounds and their morphological constituents. Specifically, an early effect of compound frequency may be problematic for obligatory decompositional models.

The role of constituents in compound processing is also controversial. Taft and

Forster (1976) claimed that the left constituent of a compound serves as the point of access to the meaning of the compound, while Juhasz, Starr, Inhoff and Placke (2003) argued for the primacy of the right constituent (see also Duñabeitia, Perea & Carreiras, 2007). Several studies of Finnish compounds established the involvement of both the left and the right constituent in reading of compounds (cf., e.g., Hyönä & Pollatsek, 1998; Pollatsek *et al.*, 2000). Moreover, Bertram and Hyönä (2003) argued on the grounds of visual acuity that the longer the compound, the more prominent the role of its morphological structure becomes.

An eye-tracking visual lexical decision study of 8-12 character-long isolated Dutch compounds by Kuperman, Schreuder, Bertram and Baayen (Chapter 3 of this dissertation) (with as nonce words non-existing compounds composed of existing nouns) established a significant effect of compound frequency emerging as early as the first fixation. Given the length of target words and constraints of visual acuity, the compound frequency effect at the first fixation is likely to precede the identification of all characters of the compound. This is supported by the fact that most compounds in their study elicited more than one fixation. The authors suggest that readers aim at identifying the compound on the basis of partial information obtained during the first fixation (e.g., initial characters, compound length and possibly an identified left constituent, see also the General Discussion). They also observed an interaction between compound frequency and left constituent frequency, which is not predicted by models that posit obligatory sequentiality in activation of the full-form and the constituent morphemes. Furthermore, they reported effects of frequency and family size for both the left and the right constituents of the compound[1].

Chapter 3 of this dissertation explained its findings within the conceptual framework of maximization of opportunity (Libben, 2006). This framework argues that readers simultaneously use, as opportunities for compound recognition, multiple sources of information (as soon as those are available to them), and multiple processing mechanisms that they have at their disposal, including full-form retrieval from the mental storage and on-line computation. In Chapter 3 we propose that an adequate model of compound processing needs to meet at

[1]The left (right) morphological family of a compound is the set of compounds that share the left (right) constituent with that compound (e.g., the left constituent family of *bankroll* includes *bankbill*, *bank holiday*, *bank draft*, etc.). The size of such family is the number of its members, while the family frequency is the cumulative frequency of family members. We included in the left (and right) families all complex words that began (or ended) with the given constituent, including, for instance, triconstituent compounds and derivations that embedded compounds

least the following four requirements: (i) explicit consideration of the temporal order of information uptake, (ii) absence of strict sequentiality in the processing of information, i.e., simultaneous processing of information at different levels in representational hierarchies; (iii) the possibility for one processing cue to modulate the presence and strength of other cues; and (iv) fast activation of constituent families, along with activation of constituents and full-forms.

The present study explores the role of morphological structure in compound processing in a way that differs from the experiment with Dutch compounds reported in Chapter 3 in several crucial respects. We use a different experimental technique (reading of compounds in sentential contexts, no lexical decisions on compounds presented in isolation), a different language (Finnish) and a different range of word lengths (10-18 characters, mean 15). We specifically address the following questions. Does the pattern of results obtained with the visual lexical decision paradigm generalize to a more natural task of sentential reading with words in normal context? Will compound frequency have an early effect in longer words, where more characters fall outside of the foveal area with high visual acuity? Will morphological families show the same facilitation in reading as they show in lexical decision? The effect of constituent family size may differ across tasks, since a more "word-like" target with a large family may facilitate a positive lexical decision. In normal reading, however, the members of the family might function as competitors and hamper the integration of the word in the sentence, which would show as inhibition in the eye movement record (for similar dualilty in the effect of orthographic neighborhood size, see Pollatsek, Perea & Binder, 1999). Finally, is there evidence in the eye movement record that different routes of lexical processing interact, when compounds are placed in sentential contexts? Another task that we set for ourselves is to formalize the specifications for a model of morphological processing outlined in Chapter 3. We propose such a model in the General Discussion.

Additionally, we consider the processing of compounds with more than two morphemes. Current research on visual processing of morphologically complex words is largely constrained to bimorphemic words (for exceptions see e.g., De Almeida & Libben, 2005; Inhoff, Radach & Heller, 2000; Krott, Baayen & Schreuder, 2001; Krott, Libben, Jarema *et al.*, 2004; Chapter 3 of this dissertation). At the same time, such complexity is anything but rare in many languages: In German, Dutch and Finnish words with three or more morphemes account for over 50% of word types. Similarly, words in the length range of 10-18 characters that we use in this study account for over 60% of word types and over 20% word tokens in Finnish. In the present experiment, we zoomed in on one type of morphological structure, where the left constituent is a derived word with a suffix and the right constituent is a simplex noun (e.g., *kirja-sto/kortti* "library card", where *kirja* is "book", *kirjasto* is "library" and *kortti* is "card").

We took into consideration two suffixes: the suffix *-stO*[2], which attaches to nouns forming collective nouns (e.g., *kirja*, "book", and *kirjasto*, "library"), and the suffix *-Us*, which attaches to verbs and forms nouns with the meaning of the act or the result of the verb (analogous to the English *-ing*, e.g., *aloittaa* "to begin" and *aloitus* "beginning"), cf., Järvikivi, Bertram and Niemi (2006). Bertram, Laine and Karvinen (1999) and Järvikivi *et al.* (2006) argue that these two suffixes differ in their affixal salience, defined as the likelihood of serving as a processing unit in identification of the embedding complex form (cf., Laudanna & Burani, 1995). The suffix *-stO* is arguably more salient and less ambiguous than the suffix *-Us*. Järvikivi *et al.* (2006) attribute this difference in salience to the fact that the suffix *-stO* has no allomorphs (i.e., is structurally invariant across inflectional paradigms), nor homonyms. Conversely, the suffix *-Us* has a very rich allomorphic paradigm (cf., several inflectional variants of *räjähdys* "explosion": *-ysken, -yksien, -ysten, -ystä, -yksiä, -yksenä*, Table 2 in Järvikivi *et al.*, 2006) and is homonymous with the deadjectival suffix *-(U)Us*.

The difference in affixal salience has demonstrable consequences for the processing of derived words. In particular, Järvikivi *et al.* (2006) showed in a series of lexical decision experiments that Finnish derived words ending in relatively salient affixes, like *-stO*, show facilitatory effects of both the surface frequency of the derived form (e.g., *kirjasto*) and the base frequency of its stem (e.g., *kirja*). At the same time, complex words that carry less salient affixes, like *-Us*, show facilitation only for surface frequency. In other words, salient affixes tend to shift the balance towards decomposition of complex words into morphemes and towards subsequent computation of a word's meaning from these constituent morphemes (e.g., Baayen, 1994; Bertram, Schreuder & Baayen, 2000; Järvikivi *et al.*, 2006; Laudanna & Burani, 1995; Sereno & Jongman, 1997).

Crucially, in bimorphemic derivations, one of the affix boundaries is explicitly marked by a space, which makes the task of parsing morphemes out of the embedding word easier. Our goal was to determine the role of affixal salience

[2]The capital characters in suffixes refer to the archiphoneme of the vowel that has back and front allophones. Realization of Finnish suffixes alternates due to the vowel harmony with the vowels in the stem, e.g., *-stO* may be realized either as /sto/ or /stœ/, and *-Us* either as /us/ or /ys/.

for suffixes orthographically and morphologically embedded in larger words. We envisioned several possible states of affairs. First, the suffix may, depending on its salience, facilitate activation of the base of the derived left constituent of the compound (i.e., *kirja* "book" in *kirjastokortti* "library card"), as shown for bimorphemic derivations by Järvikivi *et al.* (2006). On this account, one expects an interaction of base frequency by suffix type. Specifically, compounds with a relatively salient suffix *-stO* would show effects of both the base and the surface frequency of the left immediate constituent, while for the less salient suffix *-Us*, we expect to only witness the effects of left constituent surface frequency, in line with findings by Järvikivi *et al.* (2006). Second, the suffix demarcates the boundary between the two immediate constituents of the compound (i.e., *kirjasto* "library" and *kortti* "card" in *kirjastokortti*). If so, it is plausible that a more salient affix serves as a better segmentation cue and facilitates decomposition of a compound into its major constituents (for the discussion of segmentation cues in compound processing, see e.g., Bertram, Pollatsek & Hyönä, 2004). The finding expected on this account is the interaction between characteristics of the compound's constituents and the suffix type. For instance, we would expect the effects of left constituent frequency or family size to interact with the salience of our suffixes. Third, suffixes might pave the way for both parsings (*kirja* in *kirjastokortti* and *kirjasto* in *kirjastokortti*), as they may demarcate both the boundary of the base in the derived left constituent and the boundary between the compound's major constituents. If this is the case, we would expect the frequencies (or other morphological characteristics) of both the base and the full-form of the left constituent to interact with the suffix type.

As the time-course of morphological effects is essential for this study, we opted for using the eye-tracking experimental paradigm, which allows for a good temporal resolution of cognitive processes as reflected in eye movements. Furthermore, multiple regression mixed-effects modeling with participants and items as crossed random effects satisfied our need to explore simultaneously many predictors, both factors and covariates, while accounting for between-participants and between-items variance (cf., Baayen, Davidson & Bates, 2007; Bates & Sarkar, 2005; Pinheiro & Bates, 2000).

Method

Participants

Twenty-seven students of the University of Turku (18 females and 9 males) participated in this experiment for partial course credit. All were native speakers of Finnish and had normal or corrected-to-normal vision.

Apparatus

Eye movements were recorded with an EyeLink II eye-tracker manufactured by SR Research Ltd. (Canada). The eyetracker is an infrared video-based tracking system combined with hyperacuity image processing. The eye movement cameras are mounted on a headband (one camera for each eye), but the recording was monocular (right eye) and in the pupil-only mode. There are also two infrared LEDs for illuminating the eye. The headband weighs 450 g in total. The cameras sample pupil location and pupil size at the rate of 250 Hz. Recording is performed by placing the camera and the two infrared light sources 4-6 cm away from the eye. Head position with respect to the computer screen is tracked with the help of a head-tracking camera mounted on the center of the headband at the level of the forehead. Four LEDs are attached to the corners of the computer screen, which are viewed by the head-tracking camera, once the participant sits directly facing the screen. Possible head motion is detected as movements of the four LEDs and is compensated for on-line from the eye position records. The average gaze position error of EYELINK II is $<0.5^o$, while its resolution is 0.01^o. The stimuli were presented on a 21 inch ViewSonic computer screen, which had a refresh rate of 150 Hz.

Stimuli

The set of target words included 50 noun-noun compounds with the derivational first constituent ending in the suffix *-stO* (e.g., *tykistötuli* "cannon fire"), 50 noun-noun compounds with the derivational first constituent ending in the suffix *-Us* (e.g., *hitsaustyö* "a piece of welding"), and 50 bimorphemic compounds with two noun stems (e.g., *palkkasotilas* "a soldier of fortune"). Average values for frequency and length measures for the three types of compounds are summarized in Table 4.3 in the Appendix. All target words were selected from an unpublished Finnish newspaper corpus of 22.7 million word forms with the help of the WordMill database program (Laine & Virtanen, 1999). Each target word in the nominative case was embedded in a separate sentence, and it never occupied the sentence-initial or sentence-final position. All critical sentences had semantically neutral initial parts up to the target word. In a separate rating task, we asked five participants (none of whom participated in the eye-tracking experiment) to rate how felicitous the target words (e.g., *perhetapahtuma* "family happening") were given the preceding context (*Iloinen ja jännittävä...* "The happy and exciting ...") using a scale from 1 (does not fit at all) to 5 (fits very well). The task included all target sentences from the eye-tracking experiment, as well as fillers. The mean rating for target words was 3.7, which shows that the target words were in general a good continuation of the preceding context. Compound-specific ratings were not significant predictors of reading times in our statistical models. Averages per suffix type were 3.8, 3.7 and 3.6 for bimorphemic compounds, compounds with *-stO* and compounds with *-Us*, respectively. Pairwise t-tests showed no difference in ratings between the different compound types.

Eighty filler sentences were added to the 150 target sentences. All sentences comprised 5-12 words and took up at most one line. The sentences were displayed one at a time starting at the central-left position on the computer screen. Stimuli were presented in fixed-width font Courier New size 12. With a viewing distance of about 65 cm, one character space subtended approximately 0.45^o of visual angle.

Sentences were presented in two blocks, while the order of sentences within the blocks was pseudo-randomized and the order of blocks was counterbalanced across participants. Approximately 14% of sentences were followed by a screen with a yes-no question pertaining to the content of the sentence. The experiment began with a practice session consisting of five filler sentences and two questions.

Procedure

Prior to the presentation of the stimuli, the eye-tracker was calibrated using a three-point grid that extended over the horizontal axis in the middle of the computer screen. Prior to each stimulus, correction of calibration was performed by displaying a fixation point in the central-left position. After calibration, a sentence was presented to the right of the fixation point.

Participants were instructed to read sentences for comprehension at their own pace and to press a "response" button on the button box. Upon presentation of a question, participants pressed either the "yes"-button or the "no"-button on the button box. If no response was registered after 3000 ms, the stimulus was removed from the screen and the next trial was initiated. Responses and response times of participants were recorded along with their eye movements. The experimental session lasted 50 minutes at most.

Dependent variables

In the analysis of the eye-tracking data, we considered as measures of early lexical processing the duration of the first fixation (*FirstDur*), as well as the subgaze duration for the left constituent of a compound (the summed duration of all fixations that landed on the left constituent of a compound before fixating away from that constituent, *SubgazeLeft*. As a measure of later lexical processing, we focused on the subgaze duration for the right constituent of a compound (the summed duration of all fixations that landed on the right constituent of a compound before fixating away from that constituent, *SubgazeRight*. As a global measure, we considered the gaze duration on the whole word (the summed duration of all fixations on the target word before fixating away from it, *GazeDur*). We obtained additional information from two other measures: the probability of a single fixation (*SingleFix*) and - in order to assess how smoothly compound processing proceeded - the probability of the second fixation landing to the left of the first fixation position (*Regress*)[3]. All durational measures were log-transformed to reduce the influence of atypical outliers.

Predictors

Trials were uniquely identified by the participant code (*Subject*) and item (*Word*). The type of affix used in the target words was coded by the factor *SuffixType* with values "stO", "Us" and "none" (for bimorphemic compounds).

Lexical distributional properties of morphological structure. We considered compound lemma frequency, *WordFreq*, while lemma frequency was defined as the summed frequency of all inflectional variants of a word (e.g., the lemma frequency of *cat* is the sum of the frequencies of *cat*, *cats*, *cat's* and *cats'*). As frequencies of compounds' constituents have been shown to codetermine the reading times along with compound frequency (e.g., Andrews *et al.*, 2004; Hyönä & Pollatsek, 1998; Juhasz *et al.*, 2003), we included lemma frequencies of the compound's left and right constituents as isolated words, *LeftFreq* and *RightFreq*. Additionally, for each derivational left constituent (e.g., *kirjasto* "library" in *kirjastokortti* "library card") we included the lemma frequency of its base word (e.g., *kirja* "book"), *BaseFreq*, as a predictor. All frequency-based measures in this study, including the ones reported in the remainder of this section, were (natural) log-transformed to reduce the influence of outliers.

The morphological family sizes and family frequencies of a compound's constituents are known to codetermine the processing of compounds (cf., e.g., De Jong, Schreuder & Baayen, 2000; Juhasz *et al.*, 2003; Krott & Nicoladis, 2005; Moscoso del Prado Martín, Bertram, Haikio *et al.*, 2004; Nicoladis & Krott, 2007; Pollatsek & Hyönä, 2005 and Chapter 3 of this dissertation). The larger the number

[3]Other considered dependent measures included the total number of fixations, durations of the second and third fixation, amplitude of the first and second within-word saccades, and the probability of eliciting more than two fixations. The measures did not provide additional insight into our research questions.

of members in such a family or the larger their cumulative frequency, the faster the identification of the constituent and the embedding compound proceeds, as shown in lexical decision and eye-tracking studies. The related measure, the family frequency of the left (right) constituent, failed to reach statistical significance in our models (even when the respective family size was not included in the models) and will not be further discussed.

Other variables.

To reduce variance in our models, we controlled for several variables that are known to modulate visual processing. Among many other predictors (see Appendix for the full list), we considered compound length (*WordLength*) and the length of the left constituent *LeftLength*. We also included as a predictor the position of trial N in the experimental list as a measure of how far the participant has progressed into the experiment. This measure, *TrialNum*, allows us to bring under statistical control longitudinal task effects such as fatigue or habituation.

Statistical considerations

Several of our measures showed strong pair-wise correlations. Orthogonalization of such variables is crucial for the accuracy of predictions of multiple regression models. Teasing collinear variables apart is also advisable for analytical clarity, as it affords better assessment of the independent contributions of predictors to the model's estimate of the dependent variable (see Baayen, 2008: 198). We orthogonalized every pair of variables for which the Pearson correlation index r exceeded the threshold of 0.5. Decorrelation was achieved by fitting a regression model in which one of the variables in the correlated pair, e.g., *LeftLength*, was predicted by the other variable, e.g., *WordLength*. We considered the residuals of this model, *ResidLeftLength*, as an approximation of the left constituent length, from which the effects of compound length were partialled out. Using the same procedure, we obtained *ResidLeftFreq* (orthogonalized with *WordFreq* and *LeftLength*), *ResidLeftFamSize* (orthogonalized with *LeftFreq*), *ResidBaseFreq* (orthogonalized with *LeftFreq*), and *ResidRightFamSize* (orthogonalized with *RightFreq*). All orthogonalized measures were very strongly correlated with the measures, from which they were derived (rs > 0.9, p < 0.0001). The collinearity between the resulting set of numerical predictors was low, as indicated by $\kappa = 1.44$.

Additionally, some of the predictors were centered, so that the mean of their distribution was equal to zero. This procedure is crucial to avoid spurious correlations between random slopes and random intercepts in mixed-effects regression models (cf., Baayen, 2008: 276).

Table 4.4 in the Appendix lists the distributions of the continuous variables used in this study, including statistics on their original values and (if different from the original values) the values actually used in the models.

In this study we made use of mixed-effects multiple regression models with *Subject* and *Word* as random effects. For predicting binary variables (e.g., indicators of whether the given fixation is word-final or regressive), we used generalized mixed-effects multiple regression models with a logistic link function and binomial variance. We coded the "Yes" values as successes and "No" values as failures.

The distribution of durational dependent measures was skewed even after the log transformation of durations. Likewise, residuals of the mixed-effects models for durations were almost always skewed. To reduce skewness, we removed outliers from the respective datasets, i.e., points that fell outside the range of -2.5 to to 2.5 units of SD of the residual error of the model. Once outliers were removed, the models were refitted, and we reported statistics for these trimmed models. Unless noted otherwise, only those fixed effects are presented below that reached significance at the 5%-level in a backwards stepwise model selection procedure.

The random effects included in our models significantly improved the explanatory value of those models. Improvement was indicated by the significantly higher values of the maximum likelihood estimate of the model with a given random effect as compared to the model without that random effect (all ps < 0.0001 using likelihood ratio tests).

Results and Discussion

The initial pool of data points comprised 13394 fixations. We log-transformed the fixation durations and removed from the dataset for each participant those fixations that exceeded 3.0 units of SD from that participant's mean log-transformed duration. The number of removed fixations was 397 (3%), and the resulting range of fixation durations was 60 to 892 ms. Subsequently, fixations that bordered microsaccades (fixations falling within the same letter) were removed (44 x 2 = 88 fixations, 0.6%). Finally, we only considered the fixations pertaining to the first-pass reading (i.e., the sequence of fixations made before the fixation is made outside of the word boundaries, 67% of the original dataset). As a result, we were left with a pool of 9023 valid fixations.

A negligible percent of the target words was skipped ($< 0.01\%$). Twenty-seven percent of the target words required only one fixation, 40% required exactly two fixations, 20% required exactly three fixations, and it took four or more fixations to read the remaining 13% of our compounds. The average number of fixations on a stimulus was 2.2 (SD = 1.2). Regressive fixations (i.e., fixations located to the left of the previous fixation within same word) constituted 14.2% of our data pool. The average fixation duration was 234 ms (SD = 84), and the average gaze duration was 455 ms (SD = 263).

We report in the Appendix full specifications of the models for the first fixation duration (3967 datapoints, Table 4.5), subgaze duration for the left constituent (3800 data points, Table 4.6), subgaze duration for the right constituent (2342 data points, Table 4.7), and gaze duration (3884 data points, Table 4.8).

Time-course of morphological effects

Table 4.1 summarizes effects of morphological predictors on reading of long, multiply complex Finnish compounds across statistical models for early and cumulative measures (see full specifications for the models in Appendix). The table provides effect sizes (see Appendix for the explanation as to how these were computed) and p-values for main effects, and it also indicates interactions between morphological and other predictors of interest. For clarity of exposition, we leave out in this section interactions between morphological predictors and the type of the suffix in the compound's left constituents: These interactions are presented in detail in the next section.

Results presented in Table 4.1 reveal the temporal pattern of how effects of morphological structure unfold in complex word recognition. First, characteristics pertaining to the compound's left constituent, such as left constituent frequency and family size, show effects in both the early measures of reading times (first fixation duration, subgaze duration on the left constituent), and in the later measure (subgaze duration of the right constituent). Conversely, characteristics of the compound's right constituent are not significant predictors at early stages of lexical processing and only yield significant effects (always modulated by interactions with other predictors) in the measures of right constituent subgaze duration and gaze duration. This sequence of effects corroborates previous findings that both constituents are activated during processing of compounds (cf., Hyönä, Bertram & Pollatsek, 2004). Moreover, the order of their activation goes hand in hand with the typical sequence of the visual uptake in long compounds that was observed previously in Hyönä *et al.* (2004) in Chapter 3 and again in the present study,

Table 4.1: Summary of morphological effects on durational measures

Predictor	FirstDur	SubgazeLeft	SubgazeRight	GazeDur
ResidLeftFreq	-13 ms (0.001)	-72 ms (<0.001)	ns	-72 ms (0.006)
ResidLeftFamSize	-9 ms (0.02)	-80 ms (<0.001)		-120 ms (0.001)
			interaction with *WordFreq* (0.004), Fig. 4.1	
RightFreq	ns	ns	ns	ns
			interaction with *WordLength* (<0.001)	
ResidRightFamSize	ns	ns	ns	ns
			interaction with *WordFreq* (0.022), Fig. 4.2	interaction with *WordFreq* (0.002), Fig. 4.1
WordFreq	-12 ms (0.010)	-110 ms (<0.001)	-44 ms (<0.001)	-136 ms (<0.001)
			interaction with family sizes	interaction with *ResidRightFamSize* (0.002)
			(left: 0.004; right: 0.022), Figs. 4.1, 4.2	

Numbers in columns 2-5 show sizes of statistically significant effects. Numbers in brackets provide p-values for the effects, estimated based on the MCMC method with 1000 simulations. "ns" stands for non-significant. Estimation of effect sizes is based on models that do not include interactions of morphological predictors by suffix type: those interactions are summarized below in Table 4.2

such that the first fixation tends to land on a compound's left constituent and the second fixation on its right constituent[4]. We also note that the influence of the frequency-based characteristics of the left constituent on the lexical processing of compounds is qualitatively stronger than the corresponding measures for the right constituent. Left constituent frequency and family size show main effects in the models for fixation durations and subgaze and gaze durations, whereas effects of the right constituent frequency and family size are qualified by the interaction with compound length and compound frequency, respectively. The dominant involvement of the left constituent in compound processing is in line with the findings of Taft and Forster (1976). It is at odds with the important role of the right constituent, which Juhasz *et al.* (2003) proposed due to the greater semantic similarity between the compound's meaning and the meaning of the right constituent (as opposed to the typically lower degree of semantic similarity between the compound and its left constituent).

Second, we observed effects of constituents' morphological families emerging simultaneously with the effects of the respective constituent frequencies. The early effect of the left constituent family size goes against the traditional interpretation, which holds that the semantic family size effect arises due to post-access spreading activation in the morphological family (cf., De Jong *et al.*, 2002). Surprisingly, the right constituent family (e.g., *vanilla cream*, *ice cream*, *shoe cream*) is activated even when the lexical processor might have begun identification of one member of that family (e.g., *vanilla cream*), the target compound itself (the left constituent of which was processed at the preceding fixation). It may be that this effect is driven by the cases in which a compound's left constituent is particularly difficult to recognize (e.g., due to its lexical properties or non-optimal foveal view). In such cases identification of the left constituent may not be complete at the first fixation and may continue even as the eyes move to the right constituent. It may also be that activation of morphological families is automatic and happens even when not fully warranted by the processing demands: This is an empirical question that requires further investigation. More generally, we argue in the General Discussion that characteristics of the compound's right constituent may provide a valuable source

[4]The size of perceptual span in reading (3-4 characters to the left and 10-15 characters to the right of the fixation position, see e.g., Rayner, 1998) suggests that at least some characters from the compound's right constituent are very likely to be identified either foveally or parafoveally. The absence of early effects stemming from the compound's right constituent implies, however, that the available orthographic information is apparently not sufficient for early activation of that morpheme (cf., Hyönä *et al.*, 2004).

of information that facilitates recognition of a complex word and its constituents, even when other such constituents have been activated and produced detectable effects on reading times.

Third, higher compound frequency came with a benefit in speed that was present as early as the first fixation, and extended over late measures of reading times[5]. Given the lengths of our compounds (10-18 characters), it is very likely that not all the characters of the compounds are identified at the first fixation. In fact, for nearly three quarters of our compounds, visual uptake is not completed at the first fixation. Importantly, the effect of compound frequency on fixation duration is still present when single-fixation cases are removed from the statistical model. We outline possible reasons for the very early and lingering effect of compound frequency in the General Discussion.

Fourth, the effect of compound frequency on cumulative reading times was weaker in compounds that had constituents with large families. In the compounds with very large left or right constituent families the effect of compound frequency vanished (see Figures 4.1 and 4.2).

The interactions of characteristics traditionally associated with the full-form representation (i.e., compound frequency) and characteristics of morphemes that imply decomposition (i.e., constituent family sizes) are not easily explained in the strictly sublexical and supralexical models that postulate temporally sequential activation of the full-forms and constituents of compounds and hence predict the effects of morphemes and compounds to reach their full magnitude independently of each other.

Additionally, we observe that higher right constituent frequency correlated with shorter *SubgazeRight*, and this effect was stronger in longer compounds. This implies that the strength of morphological effects can also be modulated by visual characteristics of the word, in line with the earlier report of Bertram & Hyönä (2003).

Differences across types of compounds

Recall that our data comprised three types of compounds: compounds with the left constituent ending in the relatively salient affix *-stO*, compounds with the left constituent ending in the less salient affix *-Us*, and bimorphemic compounds with

[5]There were no significant interactions of compound frequency with compound length (cf., Bertram & Hyönä, 2003). However, most our compounds fall into the category of "long" compounds (above 12 characters) in Bertram & Hyönä (2003). So the reported interaction across long and short compounds (8 or less characters) was unlikely to emerge here.

Figure 4.1: Interaction of compound frequency by (residualized) left constituent family size for right subgaze duration. The lines plot the effect of compound frequency for the quantiles of left constituent family size (quantile values provided at the right margin). Compound frequency comes with the strongest negative effect at the 1st quantile (solid line), the effect gradually levels off at the 2nd quantile (dashed line), the 3d quantile (dotted line) and the 4th quantile (dotdash line), and even reverses to the positive direction for the largest left constituent families, the 5th quantile (longdash line).

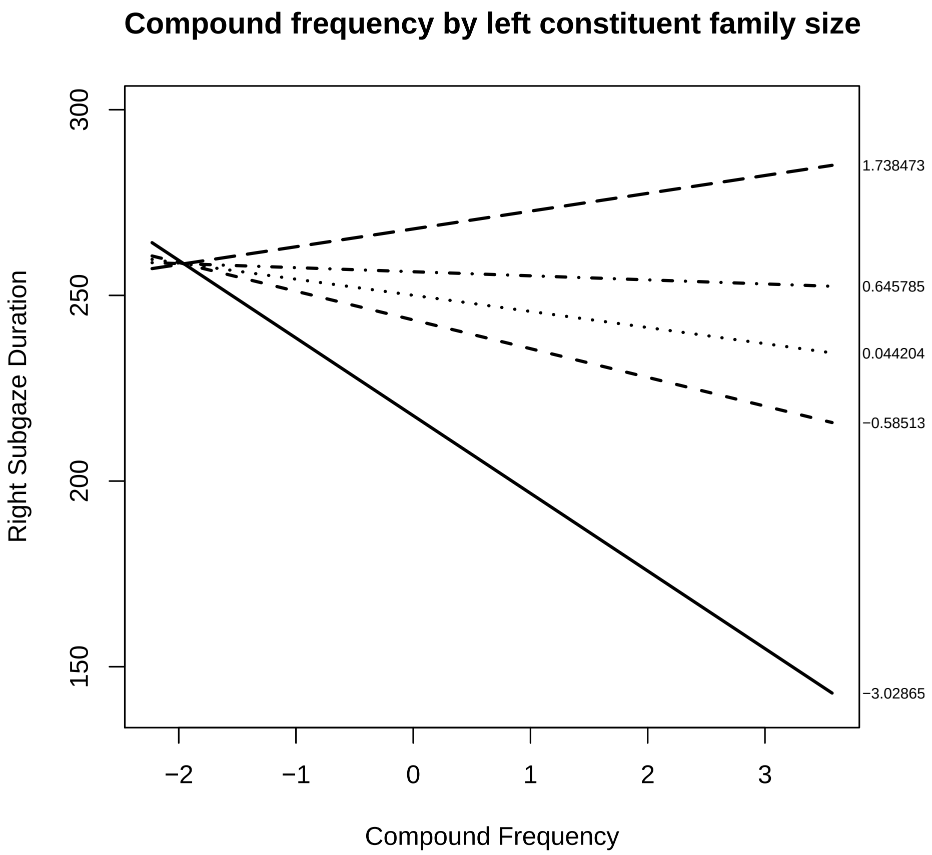
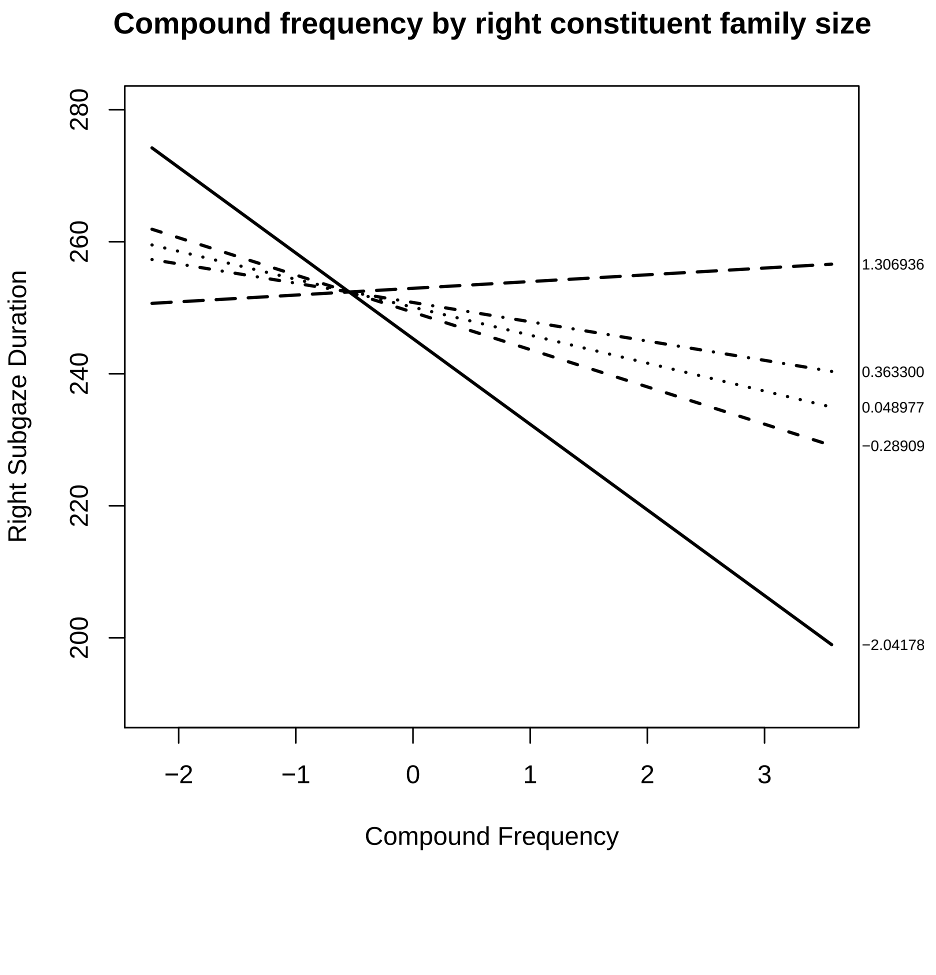


Figure 4.2: Interaction of compound frequency by (residualized) right constituent family size for right subgaze duration. The lines plot the effect of compound frequency for the quantiles of left constituent family size (quantile values provided at the right margin). Compound frequency comes with the strongest negative effect at the 1st quantile (solid line), the effect gradually levels off at the 2nd quantile (dashed line), the 3d quantile (dotted line) and the 4th quantile (dotdash line), and even reverses to the positive direction for the largest left constituent families, the 5th quantile (longdash line).



two simplex constituents. *SuffixType* did not reveal a simple main effect in our statistical models, but it qualified the effects of several morphological predictors, summarized in Table 4.2 across several statistical models. Table 4.2 provides a comparative overview of morphological effects across suffix types, including effect sizes and associated p-values per suffix, as well as p-values for interactions.

Measures of the early visual uptake (probability of a single fixation and probability of the regressive second fixation) suggest that bimorphemic compounds and especially compounds with the suffix *-Us* come with a higher processing load (i.e., require more fixations and elicit more regressive fixations) than words with the salient suffix *-stO*, which benefit most from the properties of the left constituent (i.e., require fewer fixations).

The cumulative measures of reading times demonstrate a straighforward pattern: Compounds with left constituents ending in the suffix *-stO* show much stronger effects of the left constituent frequency and family size than bimorphemic compounds and especially than compounds with the suffix *-Us*. We view this difference as evidence that this relatively salient suffix acts as a better segmentation cue for parsing out a compound's constituents than the suffix *-Us* with its many allomorphs, or the constituent boundary in bimorphemic compounds. Earlier identification of the left constituent ending in *-stO* may lead to easier recognition of that constituent and to earlier and larger effects of distributional characteristics pertaining to that constituent.

Surprisingly, bimorphemic compounds demonstrated stronger effects of the left constituent than compounds with the suffix *-Us* did. The three types of compounds can be ordered by the relative ease of processing (and, we argue, by the salience of their segmentation cues) as follows: (i) compounds with the suffix *-stO*, (ii) bimorphemic compounds and (iii) compounds with the suffix *-Us*. This finding is counterintuitive given that the bigram "Us" has a very high frequency of occurrence and a high productivity as a suffix in Finnish (see Table 1 in Järvikivi *et al.*, 2006). It represents the nominative case of two suffixes with high-frequency and high-productivity, deadjectival *-Us*, which we focus on in this study, and a homonymous deverbal *-(U)Us* (cf., Järvikivi *et al.*, 2006). That is, the character string "Us" would be a likely candidate for serving as a suffix and thus would be expected to perform as a better segmentation cue than the n-gram at the constituent boundary of a bimorphemic compound (we note that the frequency of a bigram straddling the constituent boundary was not a significant predictor in any of our models).

Table 4.2: Summary of interactions of morphological predictors with SuffixType

Predictor	Measure	-stO	-Us	None	p-value
ResidLeftFreq					
	single fixation probability	more likely single fixation 3.1 log odds units(<0.001)	ns	ns	p = 0.004
	probability of regressive fixation	ns	more likely regression 0.19 log odds units (0.025)	ns	p = 0.009
	left subgaze duration	shorter duration -148 ms (0.0001)	ns	shorter duration -48 ms (0.07)	p = 0.004
	gaze duration	shorter duration -120 ms (0.0001)	ns	shorter duration -15 ms (0.08)	p = 0.005
ResidLeftFamSize					
	left subgaze duration	shorter duration -204 ms (0.0001)	ns	shorter duration -80 ms (0.03)	p = 0.0045
	right subgaze duration	shorter duration -35 ms (0.0345)	ns	ns	p = 0.0045
	gaze duration	shorter duration -246 ms (<0.0001)	ns	ns	p = 0.0004

Numbers in columns 3-5 show sizes of statistically significant effects. Numbers in brackets provide p-values for the effects. "ns" stands for non-significant. Column 6 provides the estimate of statistical significance for the interactions with *SuffixType* based on the MCMC method with 1000 simulations.

One explanation for this finding is offered by Järvikivi *et al.* (2006) who argue that the identification of the suffix *-Us*, and subsequent parsing of the derived word, is impeded by the rich allomorphic paradigm that comes with that suffix. The two-level version of the dual-route model (Allen & Badecker, 2002) would predict that activation of competing allomorphic variants takes place as soon as access is attempted to any of the variants due to the lateral links between the different allomorphs. The early allomorphic competition for a structurally variant suffix may explain the worse performance of the suffix *-Us* as a segmentation cue in comparison to bimorphemic words, which indeed is noticeable from the first fixation onwards.

Another dimension of salience that differs across our suffixes is homonymy. The deverbal suffix *-Us* (analogous to the English *-ing*) is homonymous with the highly frequent deadjectival suffix *-(U)Us* (analogous to the English *-ness*), while the suffix *-stO* has no homonyms. Bertram, Laine and Kalvinen (1999) and Bertram, Schreuder and Baayen (2000) found that the presence of homonymy may create ambiguity as to the semantic/syntactic role that the suffix performs in the given word (in our case, the left constituent of a compound). Resolving this ambiguity might then come with slower processing of the homonymous suffix. This is unlikely to happen in our case, though, since the homonymous suffixes *-Us* and *-(U)Us* are very close in their meaning and syntactic function (cf., Järvikivi *et al.*, 2006).

A more important factor may be that the phonotactic rules of Finnish are such that the trigram "stO" only occurs in a word-initial position in a small number of borrowed words (26 word types, e.g., *stockman*). Thus, when embedded in complex words, this trigram serves as a clear cue of the constituent boundary, since it is much more probable to occur at the end of the left consituent than in the beginning of the right one. On the other hand, a substantial number of Finnish words begin with the bigram "Us" (509 word types, including highly frequent words like *ystävä* "friend" or *uskoa* "to believe"). The high positional probability of the bigram "Us" at the word's beginning may pave the way for misparsings that attribute the suffix *-Us* to the final constituent, rather than to the initial constituent in which the suffix is actually embedded. Due to a higher likelihood of misparsings, the suffix *-Us* would then figure as a less salient affix than its counterpart *-stO* in the situation when suffixes occupy a compound-medial position.

We find no effects of the morphological base of a compound's left constituent for any type of compound that we considered. This is at odds with the results of Järvikivi *et al.* (2006), who show significant effects of the base frequency for derivations with the relatively salient suffix *-stO*, as opposed to derivations with *-Us*. Clearly, in their data the identification of the suffix makes available two morphological sources of information, one provided by the base of the left constituent (e.g., *kirja* in *kirjastokortti*) and the other provided by the major constituent boundary between the left constituent *kirjasto* and the right constituent *kortti*. Our data only provides support for the detection of the immediate constituents. It appears that in trimorphemic compounds left constituent bases do not offer much information in addition to what information is carried by a compound's immediate constituents, and so the contribution of left constituent bases is too weak to be detected in our experiment.

We also report an interaction of *SuffixType* with *TrialNum*, such that the reading times for the right constituent were shorter towards the end of the experiment only for compounds including the suffix *-stO*, and not for other types of compounds ($p = 0.0015$ as estimated via the Monte Carlo Markov chain (MCMC) random-walk method using 1000 simulations). The suffix *-stO* is not too frequent in Finnish, so its presence in 22% of our stimuli sentences may have led to overrepresentation and easier recognition of this sequence of characters towards the end of the experimental list, more so than for the high-frequency suffix *-Us*. We note, however, that the covariance-analytical technique implemented in multiple regression models ensures that all other effects predicted by those models are observed over and above the impact of overrepresentation on eye movements.

Below we offer a formal, model-based view of the role that affixes structurally and orthographically embedded in compounds play in activation of other morphological constituents.

General Discussion

The key issue that we investigated in this chapter is the time-course of morphological effects in the lexical processing of long, multiply complex Finnish compounds.

We found evidence for the activation of most morphological cues (i.e., morphemes, sequences of morphemes and morphological paradigms) that are available in our compounds. These cues create opportunities for recognition of complex words. Moreover, there is a temporal flow of morphological information during reading of our compounds, which is roughly as follows. Typically the first fixation on a compound lands on its left immediate constituent. As early as the first fixation, we observe simultaneous effects of compound frequency, compound length, left constituent frequency and left constituent family size. The second and subsequent fixations usually land further into the word, such that the right constituent comes under foveal inspection and a new source of morphological information becomes available for recognition of compounds. Consequently, the effects of right constituent frequency and right constituent family size emerge late, and their effects are weaker than those of the left constituent. Finally, we observe interactions between compound frequency and both the left and the right constituent family sizes.

Perhaps the most intriguing of our findings is that the early effect of compound frequency apparently precedes the complete identification of all characters and of the right constituents of our long compounds. This effect suggests that readers make inferences about the compound's identity as soon as they have available any (potentially incomplete) information about the word. Information about formal compound properties, such as its initial characters or length, may be available from the parafoveal preview and from the earliest stages of foveal inspection of the word (see Rayner, Well, Pollatsek & Bertera, 1982). Readers may match the visual pattern consisting of several initial characters in combination with word length against words stored in memory long before the compound as a whole is scanned. The more frequent matches to such patterns may boost the identification of that compound. Compound frequency may also be considered as the combinatorial strength of association between the morphemes of a compound and its full-form representation. Activation of one morpheme may then lead to activation of combinations with that morpheme, which will be stronger for higher-frequency combinations. Thus, identification of the left constituent, potentially enhanced by the information about word length, may also lead to early identification of compounds that embed that constituent (for the length constraint hypothesis, see O'Regan, 1979; Clark & O'Regan, 1999; for the opposing view, see Inhoff & Eiter, 2003). We note that the effect of compound frequency lingers on throughout the entire course of reading a compound, which implies that the full-form representation of a compound keeps being actively involved in the recognition process as other morphological and orthographic cues to identification become available to the reader.

Observed effects of left and right constituent frequency, like the effect of compound frequency, may gauge both the ease of access to the morpheme in the mental lexicon, and, at the level of form, the reader's experience with identifying a character string that represents the constituent as a word pattern within a larger word. Additionally, left and right constituent family sizes may be measures of the semantic resonance following activation of a constituent, but also a measure of experience that the reader has with parsing that constituent out of compound words.

We explain qualitatively stronger effects pertaining to the compound's left constituent (as compared to those pertaining to the compound's right constituent) by the time-course of visual uptake. As a result of its later availability for the visual system, identification of a compound's right constituent may proceed against the backdrop of existing knowledge gleaned from the left constituent. Since the informational value carried by a compound's right constituent is attenuated by the information obtained earlier, the contribution of that constituent to the comprehension of a compound is smaller than the contribution of the left constituent.

We note that most of the morphological measures that we have described so far can be argued to tap both into the formal properties of a compound or its morphemes, and into their semantic representations and semantic integration of morphemes in a whole: This duality is quite in line with recent findings that morphological effects imply at least two processing stages, that of form-based decomposition and that of semantic integration (e.g., Meunier & Longtin, 2007). However, the finding of Pollatsek and Hyönä (2005) that there is no semantic transparency effect on encoding of Finnish compounds in reading indicates that the role of formal properties in compound recognition may be stronger than that of semantics.

The present findings show remarkable convergence with the findings in Chapter 3 of this dissertation, which included the early effect of compound frequency, early effects of left constituent frequency and family size, late effects of right constituent frequency and family size, and interactions between compound frequency and frequency-based measures of the left constituent. In other words, the findings are robust to language (Dutch vs. Finnish), the experimental task (lexical decision vs. reading), the experimental technique (single word reading vs. sentential reading), or the range of word lengths (8-12 vs. 10-18 characters). Below we discuss implications of these findings for current models of morphological processing, and propose a formal model, the PRObabilistic Model of Information SourcEs (henceforth, PROMISE) to account for the present results and results of Chapter 3.

Our set of findings has far-reaching consequences for current theories of morphological processing. While eye-movements (like any other known experimental paradigm) cannot exhaustively access the time course of compound processing in absolute terms, they certainly give us insight in some crucial aspects of the processing time-flow. The fact that we are using long compounds allows for naturalistic separation of information sources into those that are available (and used) early in the processing and those that come into play only relatively late. For instance, the early effect of compound frequency is problematic for approaches that require prelexical decomposition of full-forms prior to identification of complex words (e.g., Taft, 1991; Taft, 2004). A pure decompositional model proposed for inflections and derivations assumes access to both morphological constituents before full-form representations are activated. More specifically, Taft and Ardasinski (2006) argue that in the case of inflections, full-form representations are not activated at all, while in the case of derivations, full-form representations are activated at the lemma level after activation of both constituents. Our results go against these assumptions, since we find evidence for activation of the full-form representation before the activation of the right constituent. The kind of a decompositional feed-forward model, advanced by Taft and Forster (1976) for compounds, assumes that the compound's full-form is activated by and after access to the left constituent. It does not predict any effect of the right constituent at all, contrary to our results (see also Lima & Pollatsek, 1983 and Bertram & Hyönä, 2003).

For supralexical models, there is a logical possibility that the full-form representation of the compound is activated and, in sequence, this activation spreads to the compound's left constituent, such that the effects of both the compound as a whole and its left constituent are detectable within the short duration span of the first fixation. A problem for this class of models, however, is that activation of the right constituent of a compound is predicted to be simultaneous with that of the left constituent, but we observed no effect pertaining to characteristics of right constituents in either first nor second fixation measures. Also for short compounds we predict, on the the basis of the temporal shift in the effects of compound frequency and right constituent frequency, that accessing the compound's full-form does not automatically imply lexical access to properties of the right constituent.

Another finding that is not easy to reconcile with several current models of morphological processing is the interactions between the characteristics of a full-form (e.g, compound frequency) and the characteristics of a compound's constituents (left and right constituent family sizes), such that compound frequency has little or no effect on the reading time for the words with very large constituent families. As we argued above, in the strictly sublexical models and in supralexical models, activation of full-forms and that of morphemes are separated in time (i.e., are not parallel), so the effects of full-forms and of those morphemes are expected to fully develop on their own. In other words, these models do not predict the full-form effects to modulate, or be modulated by, the effects of morphemic properties.

Our statistical models show that the effects of compound frequency and the effects of constituent frequencies and family size unfold in parallel throughout the entire time-course of compound recognition. This observation even holds for most compounds with large constituent families or high constituent frequencies, of which we may assume that their processing is dominated by decomposition. However, the fact that both whole words and morphemes contribute to word recognition, attests that the winner-takes-it-all principle as advocated by some dual-route models (Schreuder & Baayen, 1995) can be questioned. Rather, the processing routes seem to be more co-operative than previously assumed, that is, the processing of complex words appears to draw information from multiple routes, even when one of them is more favorable.

Our results show that the patterns of morphological effects in compound processing are not captured in their entirety by current models of morphological processing. Moreover, with the exception of Pollatsek, Reichle and Rayner (2003), computational models of morphological processing make no provision about the temporal unfolding of reading, as if complete identification of the word would always require a single fixation. In Chapter 3 we suggest that theoretical assumptions such as instant access to full visual information, obligatory sequentiality or independence of processing stages need to be reconsidered in order to account for the readers' interactive use of multiple morphological cues (see Libben, 2005; Libben, 2006). In fact, most current models have been developed on the basis of experiments with relatively short compounds, i.e., those where the visual uptake is not stretched over time and the order of activation of morphemes and full-forms is difficult to establish empirically. From this perspective, it is not surprising that their predictions do not generalize to long morphologically complex words. Below we present the model of morphological processing that is based on the reading data from long words, yet it makes explicit predictions about the patterns of morphological processing expected for short complex words.

Towards a Probabilistic Model of Information Sources

We have documented a broad range of lexical distributional properties of morphological structure that codetermine the uptake of information (as gauged by durational measures in the eye-movement record). In what follows, we sketch a framework for understanding and modeling these lexical effects.

The mental lexicon is a long-term memory store for lexical information. We view an incoming visual stimulus as a key for accessing this lexical information. The information load of a stimulus is defined by the lexical information in long-term memory. Without knowledge of English, words like *work* or *cat* carry no information for the reader. It is the accumulated knowledge of words and their paradigmatic and syntagmatic properties that define a word's information load, and hence the speed with which information can be retrieved from lexical memory.

Our Probabilistic Model of Information Sources (PROMISE) takes as its point of departure the perhaps most basic statement of information theory, that information (I) can be quantified as minus log probability (P):

$$I = -\log_2 P \tag{4.1}$$

As P decreases, I increases: less probable events are more informative. A fundamental assumption of our model is that the time spent by the eye on a constituent or word is proportional to the total amount of lexical information available in long-term memory for identification of that constituent or word at that timepoint (cf., Moscoso del Prado Martín, Kostić & Baayen, 2004). Events with small probability and hence a large information load require more processing resources and more processing time (see Levy, 2008 for a similar probabilistic approach to processing demands in online sentence comprehension)[6].

Seven lexical probabilities are fundamental to our model. First, we have the probability of the compound itself. We construe this probability as a joint probability,

[6]While most of the measures considered below are traditionally considered as semantic (e.g., degree of compatability of constituents in a compound, degree of connectivity in a morphological paradgim, etc.), we remain agnostic in the present chapter to whether information originates from the level of form or the level of meaning. In all likelihood, formal properties of words reach the lexical processing system earlier than their semantic properties. Yet, as argued in e.g., Meunier and Longtin (2007) and in the present chapter, most morphological effects take place at both the level of form and that of meaning. The model is able to capture informations originating at either level as long as they can be represented numerically: as frequency measures, as the Latent Semantic Analysis scores, or as a number of members in a morphological family, of words of a given length, of synonyms, of orthographic or phonological neighbors, etc.

the probability of the juxtaposition of two constituents, μ_1 and μ_2: $\Pr(\mu_1, \mu_2)$. In what follows, subscripts refer to the position in the complex word. We estimate this probability by the relative frequency of the complex word in a large corpus with N tokens. Similar frequency-based estimates are done for all other probabilities used in PROMISE. Alternatively, the estimates of probabilities may be obtained from norming studies, e.g., Cloze sentence completion tasks, where participants are asked to guess what the next word is given the preceding sentential context and, possibly, some cues about the upcoming word. The ratio of correct guesses and total guesses serves as an estimate of the word's probability in its context. With F_{12} denoting the absolute frequency of the complex word in this corpus, we have that

$$\Pr(\mu_1, \mu_2) = \frac{F_{12}}{N}. \tag{4.2}$$

This is an unconditional probability, the likelihood of guessing the complex word without further contextual information from sentence or discourse. Two further unconditional probabilities that we need to consider are the probability of the left constituent and that of the right constituent:

$$\Pr(\mu_1) = \frac{F_1}{N} \tag{4.3}$$

$$\Pr(\mu_2) = \frac{F_2}{N}. \tag{4.4}$$

The remaining four probabilities are all conditional probabilities. The first of these is the probability of the right constituent (μ_2) given that the left constituent (μ_1) has been identified: $\Pr(\mu_2|\mu_1)$. Using Bayes' theorem, we rewrite this probability as

$$\Pr(\mu_2|\mu_1) = \frac{\Pr(\mu_1, \mu_2)}{\Pr(\mu_{1+})}, \tag{4.5}$$

where μ_{1+} denotes the set of all complex words that have μ_1 as left constituent. Hence, $\Pr(\mu_{1+})$ is the joint probability mass of all words starting with μ_1. We estimate $\Pr(\mu_2|\mu_1)$ with

$$\Pr(\mu_2|\mu_1) = \frac{\Pr(\mu_1, \mu_2)}{\Pr(\mu_{1+})} = \frac{\frac{F_{12}}{N}}{\frac{F_{1+}}{N}} = \frac{F_{12}}{F_{1+}}, \tag{4.6}$$

where F_{1+} denotes the summed frequencies in the corpus of all μ_1-initial words. This probability comes into play when the left constituent has been identified and the right constituent is anticipated, either by the end of the information uptake from the left constituent, or during the processing of the right constituent.

The next conditional probability mirrors the first: It addresses the likelihood of the left constituent given that the right constituent is known. Denoting the set of words ending in the right constituent μ_2 by μ_{+2}, the summed frequencies of these words by F_{+2}, and the corresponding probability mass by $\Pr(\mu_{+2})$, we have that

$$\Pr(\mu_1|\mu_2) = \frac{\Pr(\mu_1,\mu_2)}{\Pr(\mu_{+2})} = \frac{\frac{F_{12}}{N}}{\frac{F_{+2}}{N}} = \frac{F_{12}}{F_{+2}}. \qquad (4.7)$$

This probability is relevant in any situation where the right constituent is identified before the left, for instance, because the left constituent was skipped or only partly processed[7].

The preceding two probabilities are conditioned on the full availability of the left or the right constituent. The final two probabilities are more general in the sense that they condition on the presence of some unspecified right or left constituent, without narrowing this constituent down to one specific morpheme. The unspecified left constituent stands for the subset of all morphemes or words in a language that can appear in the word-initial position. Essentially, this subset is equal to full vocabulary with the exception of suffixes (e.g., *-ness, -ity*) and of those compounds' constituents that can only occur word-finally. Suppose that the reader has an intuition that the word under inspection, say *blackberry*, is potentially morphologically complex (based, for example, on its length or the low probability of the bigram "kb"). While the left constituent of such a compound is unspecified, combinations like **nessberry* or **ityberry* will never be part of the lexical space, which needs to be considered for identification of the full compound. Likewise, the unspecified right constituent is the set of morphemes that excludes prefixes (e.g., *un-, anti-*) or compounds' constituents (e.g., *cran-*) that can only occur word-initially.

Denoting the presence of such an unspecified left constituent by M_1 and that of such an unspecified right constituent by M_2, we denote these more general conditional probabilities as $\Pr(\mu_1|M_2)$ and $\Pr(\mu_2|M_1)$ respectively, and estimate them as follows:

$$\Pr(\mu_1|M_2) = \frac{\Pr(\mu_1,M_2)}{\Pr(M_2)} = \frac{\Pr(\mu_{1+})}{\Pr(M_2)} = \frac{F_{1+}}{F_{M_2}} \qquad (4.8)$$

$$\Pr(\mu_2|M_1) = \frac{\Pr(M_1,\mu_2)}{\Pr(M_1)} = \frac{\Pr(\mu_{+2})}{\Pr(M_1)} = \frac{F_{+2}}{F_{M_1}} \qquad (4.9)$$

[7] μ_{1+} and μ_{+2} denote the left and right constituent families. In the present formulation of the model, we estimate the corresponding probabilities and informations using the summed frequencies of these families. It may be more appropriate to estimate the amount of information in the morphological family using Shannon's entropy, the *average* amount of information (cf. e.g., Moscoso del Prado Martín, Kostić & Baayen, 2004), or, under the simplifying assumption of a uniform probability distribution for the family members, by the (log-transformed) family size, which is the measure we used for our experimental data.

In these equations, F_{M_2} denotes the summed frequencies of all words that can occur as a right constituent. Likewise, F_{M_1} denotes the summed frequencies of all words that can occur as a left constituent in a complex word. The probabilities $\Pr(M_1)$ and $\Pr(M_2)$ are independent of μ_1 and μ_2 and hence are constants in our model. $\Pr(\mu_2|M_1)$ comes into play when the left constituent is not fully processed and the likelihood of the right constituent is nevertheless evaluated. $\Pr(\mu_1|M_2)$ becomes relevant when length information or segmentation cues clarify that there is a right constituent, and this information is used to narrow down the set of candidates for the left constituent. To keep the presentation simple, here we build a model for compounds with only two morphemes: Extension to trimorphemic cases, however, is straightforward.

The basic model. We introduce our model with only three of the seven probabilities defined in the preceding section. For each of the probabilities

$$\begin{aligned}\Pr(\mu_2|\mu_1) &= \frac{F_{12}}{F_{1+}} \\ \Pr(\mu_1,\mu_2) &= \frac{F_{12}}{N} \\ \Pr(\mu_1|M_2) &= \frac{F_{1+}}{F_{M_2}}\end{aligned} \tag{4.10}$$

we calculate the corresponding weighted information using (5.1),

$$\begin{aligned}\mathsf{I}_{\mu_2|\mu_1} &= w_1(\log F_{1+} - \log F_{12}) \\ \mathsf{I}_{\mu_1,\mu_2} &= w_2(\log N - \log F_{12}) \\ \mathsf{I}_{\mu_1|M_2} &= w_3(\log F_{M_2} - \log F_{1+})\end{aligned} \tag{4.11}$$

with positive weights $w_1, w_2, w_3 > 0$. A crucial assumption of our model is that the time t spent by the eye on a constituent or word is proportional to the total amount of information available at a given point in time:

$$\begin{aligned} t &= I_{\mu_2|\mu_1} + I_{\mu_1,\mu_2} + I_{\mu_1|M_2} \\ &= w_1(\log F_{1+} - \log F_{12}) + w_2(\log N - \log F_{12}) + w_3(\log F_{M_2} - \log F_{1+}) \\ &= w_1 \log F_{1+} - w_1 \log F_{12} + w_2 \log N - w_2 \log F_{12} + w_3 \log F_{M_2} - w_3 \log F_{1+} \\ &= w_2 \log N + w_3 \log F_{M_2} - (w_1 + w_2) \log F_{12} - (w_3 - w_1) \log F_{1+}.\end{aligned} \tag{4.12}$$

Equation (4.12) states that processing time linearly covaries with $\log F_{12}$ and $\log F_{1+}$, with facilitation for compound frequency and facilitation or inhibition for left constituent family frequency, depending on the relative magnitude of w_1 and w_3. In other words, starting from simple probabilities and using information theory, we have derived a model equation the parameters of which can be directly estimated from the data using multiple (linear) regression models. Note that these parameters are simple sums of our weights *w*.

We now bring the remaining probabilities

$$\begin{aligned} \Pr(\mu_1|\mu_2) &= \frac{F_{12}}{F_{+2}} \\ \Pr(\mu_2|M_1) &= \frac{F_{+2}}{F_{M_1}} \\ \Pr(\mu_1) &= \frac{F_1}{N} \\ \Pr(\mu_2) &= \frac{F_2}{N} \end{aligned} \tag{4.13}$$

into the model as well. For each of these probabilities we have a corresponding weighted amount of information, again with positive weights:

$$\begin{aligned} I_{\mu_1|\mu_2} &= w_4(\log F_{+2} - \log F_{12}) \\ I_{\mu_2|M_1} &= w_5(\log F_{M_1} - \log F_{+2}) \\ I_{\mu_1} &= w_6(\log N - \log F_1) \\ I_{\mu_2} &= w_7(\log N - \log F_2) \end{aligned} \tag{4.14}$$

We can now define the general model as

$$\begin{aligned} t = {} & (w_2 + w_6 + w_7)\log N + w_3 \log F_{M_2} + w_5 \log F_{M_1} \\ & -(w_1 + w_2 + w_4)\log F_{12} - (w_3 - w_1)\log F_{1+} - (w_5 - w_4)\log F_{+2} \\ & -w_6 \log F_1 - w_7 \log F_2. \end{aligned} \tag{4.15}$$

This equation, as well as equations in (4.11) and (4.14), sheds light on some of the intriguing findings reported above. Compound frequency contributes to probabilities (and respective amounts of information) that readers can start estimating even before all characters may be scanned: for instance, as a term in the conditional information of the right constituent $I_{\mu_2|\mu_1}$ given the (partial) identification of the left constituent, defined in the first equation in (4.11). Also recall that the property of the right constituent family plays a role even though activation of this family would seem dysfunctional given that the only relevant right constituent family member is the compound itself. This seemingly unwarranted contribution of the right constituent family originates, however, from the fact that the family contributes to the estimate of the conditional probability $I_{\mu_2|M_1}$ of the right constituent and to the conditional probability $I_{\mu_1|\mu_2}$ of the left constituent. In other words, the family is used to narrow down the lexical space from which both constituents are selected, and thus it offers a larger amount of information about the compound and its morphemes.

Equation (4.16) in its present form treats all information sources as if they are simultaneously available to the processing system. This describes cases when the visual uptake of the word is complete in one fixation (typical of shorter and more frequent words). The formulation, however, is easily adjustable to the cases where multiple fixations are required to read the word, like in the long compounds used in the current study and in the study reported in Chapter 3. Information sources that are available early in the time-course of the visual uptake are demonstrably more important in compound recognition (cf. the weaker role of right constituent measures as compared to properties of the left constituent). In the equation, weights *w* for "early" information sources can be multiplied by a time-step coefficient α_1, such that $\alpha_1 > 1$. For "late" information sources, the value of α_2 is equal to or smaller than 1. As with weights w, the value of α can be directly estimated from comparing regression coefficients of a predictor in the models for early measures of the visual uptake (cf., *SubgazeLeft*) vs. the models for later measures (e.g., *SubgazeRight*). For the sake of exposition, we restrict our further discussion to a simpler, temporally indiscriminate, model (4.16).

There are several falsifiable predictions that follow straightforwardly from the properties of (4.16).

- The frequency of the whole compound, as well as the frequencies of its constituents as isolated words, have negative coefficients in the equation. This predicts that higher *a priori*, unconditional, frequencies of complex words and their morphemes always come with facilitation of processing (e.g., shorter reading times or lexical decision latencies).

- Three corpus constants contribute to the intercept: the token size of the corpus/lexicon (N), the number of tokens in the corpus/lexicon that can occur as a left constituent (F_{M_1}), and the number of tokens in the corpus/lexicon that can occur as a right constituent (F_{M_2}). The larger the size of a corpus/lexicon, the higher the values of all three constants and the higher the intercept. Given the positive weight coefficients, the model predicts a longer processing time for a word in a larger corpus/lexicon. This is hardly surprising, since we use absolute frequencies in (4.16). So a word with 100 occurrences per corpus would be recognized slower in a corpus of 100 million word forms that in a corpus of 1000 word forms.

- All coefficients, with the exception of w_1, occur in more than one term of equation (4.16). This expresses various trade-offs in lexical processing. For instance, w_3 appears with a positive sign for the intercept ($w_3 \log F_{M_2}$) and with a negative sign for the left constituent family frequency ($-w_3 \log F_{1+}$). We predict that the stronger facilitation compounds receive due to their higher family frequency, the higher the intercept (i.e., average processing time) across compounds is.

In the remainder of this section we apply PROMISE to the key statistical models that we fitted to our experimental data. Since most results of the model for first fixation duration are also found in the model for left subgaze duration, and most results of the model for gaze duration are also attested in the model for right subgaze duration, in what follows we concentrate on the two models for subgaze durations (cf., Tables 4.6 and 4.7 in Appendix).

Left subgaze duration. The effects of right constituent frequency and family size do not reach significance in the model for the left subgaze duration (see Table 4.6). We conclude that those information sources defined in (13) that require identification of the right constituent ($\mathrm{I}_{\mu_1|\mu_2}$, and I_{μ_2}), as well as the information source conditioned on the presence of some unspecified left constituent ($\mathrm{I}_{\mu_2|M_1}$), play no role when the left constituent is being processed. In other words, respective coefficients w_4, w_5 and w_7, are all equal to zero in (4.16).

The effect of compound frequency $logF_{12}$ on reading times is weighted in (4.16) by the sum $-(w_1 + w_2 + w_4)$. Since $w_4 = 0$ and since the regression coefficient for the predictor *WordFreq* in Table 4.6 is -0.0471, we infer that $w_1 + w_2 = 0.0471$. Given that the expression $-(w_3 - w_1)$ qualifies the effect of the left constituent family frequency, F_{1+}, and that the regression coefficient for left constituent family size *ResidFamSizeL* in Table 4.6 is -0.0431, we infer that $w_3 - w_1 = 0.0431$. It follows that 0.0471 is an upper bound for w_1 and that 0.0431 is a lower bound for w_3. Following definitions in (4.11), we state that $I_{\mu_1|M_2}$ receives greater weight than $I_{\mu_2|\mu_1}$. Apparently, the identification of the left constituent given the knowledge that there is some right constituent plays a more important role at that timepoint than anticipating the right constituent given the identity of the left constituent. Anticipation of the right morpheme probably is a process that only starts up late in the uptake of information from the left morpheme.

Interestingly, the importance of the *a priori*, context-free probability of the left constituent (I_{μ_1}) is much smaller than the contribution of that constituent recognized as part of a compound. Recall that 0.0431 is a lower bound for w_3 (the coefficient for the left constituent family frequency effect). Since $-w_6$, the coefficient for the effect of left constituent frequency as defined in (4.14), is estimated at -0.0219 from the regression coefficient for *ResidLeftFreq* in Table 4.6, the weight of the *a priori* probability w_6 is at best roughly half of that of the contextual probability of the left constituent.

An important finding for the left subgaze durations is that the effects of the left constituent frequency and left constituent family size were greater for those left constituents ending in the suffix *-stO*, cf., Table 4.2. Within the present framework, this implies that the weights w_6 (for the left constituent frequency) and w_3 (for the left constituent family size) have to be greater for left constituents with *-stO* compared to left constituents with *-Us* or simplex left constituents. Since w_6 and w_3 are used with positive signs as weights for $logN$ and $logF_{M_2}$ in (4.16), greater values for these coefficients for *-stO* imply that the intercept should be larger as well for left constituents with this suffix. As can be seen in Table 4.6, this is indeed the case: The main effect for *-stO* is positive (see the regression coefficient 0.045 for *SuffixTypeSt* in Table 4.6) and is more than twice the main effect for *-Us* (see the regression coefficient 0.0245 for *SuffixTypeUs* in Table 4.6). This suggests that a better segmentation cue helps narrowing down the set of candidates for the left constituent and hence affords better facilitation from the properties of the left constituent. Yet processing of compounds with a good segmentation cue always comes with a price of an increased intercept (i.e., longer mean processing time), the price of 'spurious' lexical co-activation. For instance, a large family may raise the resting activation level of its members (thus making easier lexical access to the target compound), and at the same time it brings along a larger number of competitors (thus inhibiting the recognition of the actual target via, for instance, lateral inhibition). Similarly, higher constituent frequency implies easier access to the compound's constituent in the mental lexicon, but stronger activation of a constituent also makes it a stronger competitor with the compound. Higher constituent frequency may also more strongly activate orthographic neighbors of the constituent and words semantically related to the constituent, all of which may enter into a competition with the target compound and thus inhibit its recognition.

Right subgaze duration. Left constituent frequency does not reach a significant effect in the regression model for the subgaze for the right constituent (Table 4.7). This indicates that $w_6 = 0$ when (4.16) is applied to this model: the unconditional information source for the left constituent, I_{μ_1}, no longer plays a role.

The regression model for the subgaze durations for the right constituent presents us with the familiar and expected facilitation for compound frequency. The facilitation for the right constituent frequency and family size are also in line with (4.16).

For left constituents in *-Us*, there is no effect of left constituent family size ($\hat{\beta} = -0.028; p = 0.18$; see *SuffixTypeUs:ResidFamSizeL* in Table 4.7). Since the effect of left constituent family $logF_{1+}$ has as its weight $-(w_3 - w_1)$ in (4.16), we conclude that here $w_1 \approx w_3$.

For left constituents in *-stO*, by contrast, we have facilitation ($\hat{\beta} = -0.055; p = 0.035$, see *SuffixTypeSt:ResidFamSizeL* in Table 4.7), indicating that $w_1 > w_3$, while for simplex left constituents there is some evidence for inhibition ($\hat{\beta} = 0.025; p = 0.085$, see *ResidFamSizeL* in Table 4.7). It follows from our model that the intercept must be greatest for *-stO*, and Table 4.7 shows that this is indeed the case. The intercept for bimorphemic compounds is the model's intercept (5.44 log units); the intercept is not significantly different for compounds with *-Us* (the model's intercept plus the regression coefficient for *SuffixTypeUs*, -0.004); and the intercept is higher for compounds with *-stO* (the model's intercept plus the regression coefficient for *SuffixTypeSt*, $5.44 + 0.12 = 5.56$ log units). Compared to the model for the left subgaze durations, this balance between increased intercept and increased facilitation emerges more clearly, with unambiguous support from the significance levels.

The right subgaze durations are characterized by (multiplicative) interactions of compound frequency by left constituent family size and compound frequency by right constituent family size that are absent for the left subgaze durations (see Figures 4.2 and 4.1). Within the present framework, an interaction such as that of compound frequency by left constituent family size implies a more complex evaluation of $I_{\mu_2|\mu_1}$, which we weighted above simply by a scalar weight w_1.

First note that the equation for $I_{\mu_2|\mu_1}$ defined in (4.11) can be re-written as follows:

$$I_{\mu_2|\mu_1} = w_1(\log F_{1+} - \log F_{12}) = \log(\frac{F_{1+}}{F_{12}})^{w_1} \quad (4.16)$$

In other words, both cues $logF_{1+}$ and $logF_{12}$ are assumed to contribute to this information source to the same extent, quantified as the coefficient w_1. We have to revise information $I_{\mu_2|\mu_1}$ in such a way that the magnitude of one cue contributing to an information source modulates the extent to which another cue contributes to that information source (see also Chapter 3). We achieve this by assigning the weight to one term in the equation (e.g., F_{12}) so that it is proportional to another term (e.g.,

F_{1+}). The weight adjusted for another cue can be defined then as $w_1 + C_1 \log F_{1+}$ for F_{12}, and as $w_1 + C_2 \log F_{12}$ for $logF_{12}$. Equation (16) can be re-written as:

$$I_{\mu_2|\mu_1} = \log \frac{F_{1+}^{w_1 + C_1 \log F_{12}}}{F_{12}^{w_1 + C_2 \log F_{1+}}} = w_1 \log F_{1+} - w_1 \log F_{12} + (C_1 - C_2) \log F_{12} \log F_{1+}, \quad (4.17)$$

$(w_1, w_2, C_1, C_2 > 0)$.

Notably, this new weighting of terms in the information source introduces into our model the desired multiplicative interaction between compound frequency and left constituent family size[8].

The interaction of compound frequency with right constituent family size can be modeled in terms of $I_{\mu_1|\mu_2}$ in the same way $(w_4, K_1, K_2 > 0)$:

$$I_{\mu_1|\mu_2} = \log \frac{F_{+2}^{w_4 + K_1 \log F_{12}}}{F_{12}^{w_4 + K_2 \log F_{+2}}} = w_4 \log F_{+2} - w_4 \log F_{12} + (K_1 - K_2) \log F_{12} \log F_{+2}. \quad (4.18)$$

Inclusion of adjusted weights in our definitions of information sources leads to the emergence of multiplicative interactions in the model, and allows to reformulate (4.16) and obtain the following model for the right subgaze durations:

$$\begin{aligned} t \;=\; & (w_2 + w_7) \log N + w_3 \log F_{M_2} + w_5 \log F_{M_1} \\ & -(w_1 + w_2 + w_4) \log F_{12} \\ & -(w_3 - w_1) \log F_{1+} - (w_5 - w_4) \log F_{+2} - w_7 \log F_2 \\ & +(C_1 - C_2) \log F_{12} \log F_{1+} + (K_1 - K_2) \log F_{12} \log F_{+2}. \end{aligned} \quad (4.19)$$

Figure 4.3 illustrates the geometry of the interactions in 4.19 by example of the interaction $(C_1 - C_2) \log F_{12} \log F_{1+}$.

The upper panels illustrate the difference between a model without (left) and with (right) an interaction with a positive coefficient ($C_1 > C_2$). The right panel illustrates how facilitation can be reversed into inhibition depending on the value of the other predictor. Crucially, the interactions predicted by our statistical model for right subgaze duration in Figure 4.1 and Figure 4.2 are two-dimensional representations of the shape shown in the right panel of Figure 4.3.

[8] Other estimates of weights are also possible. For instance, the amount of information I_{μ_1,μ_2} can be derived from probability equation (2) using the same weight, rather than different weights for the numerator and denominator: $log[F_{12}/N]^{w_2 + logF_{12}} = w_2 logN - logF_{12}(logN + w_2) + logF_{12}^2$. Note that I_{μ_1,μ_2} becomes a polynomial with F_{12} as a negative linear term and a positive quadratic term. This equation predicts the L-shape or the U-shape functional relationship between processing time and compound frequency. The L-shape frequency effect is indeed observed in comprehension (Baayen, Feldman & Schreuder, 2006) and the U-shape effect in production (Bien, Levelt & Baayen, 2005).

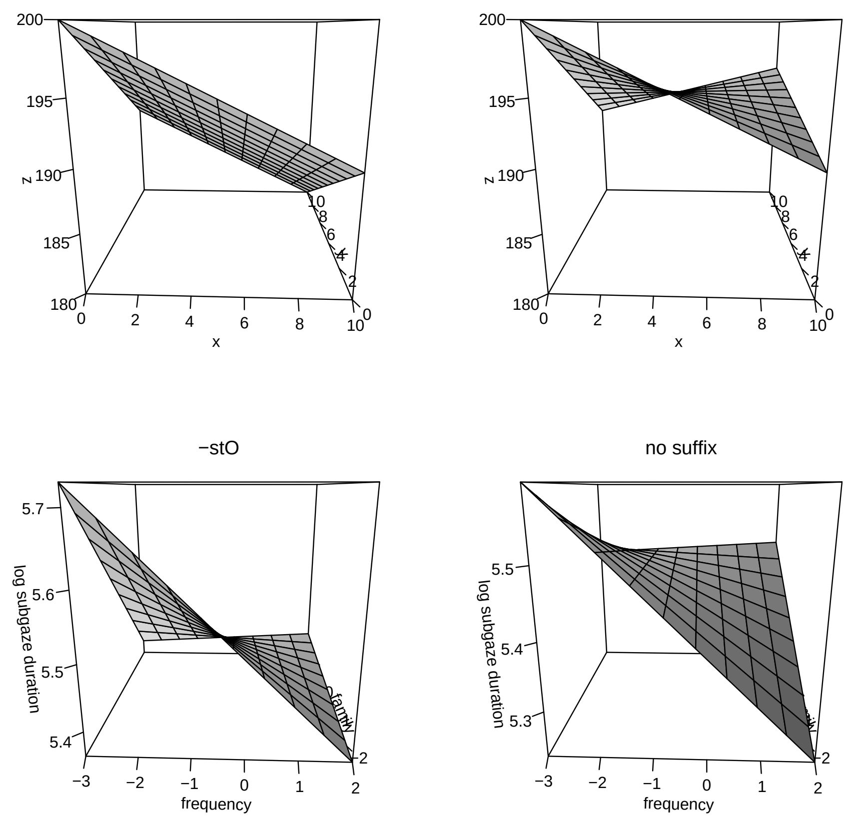


Figure 4.3: Perspective plots for (upper left panel) a linear model with additive main effects and no interaction and for (upper right panel) a linear model with a multiplicative interaction ($\beta_0 = 200, \beta_1 = -1, \beta_2 = -1$, for the left panel, $\beta_3 = 0$, for the right panel, $\beta_3 = 0.2$). The lower panels show the interaction of left constituent family size and compound frequency for the right subgaze durations for compounds with left constituents ending in the suffix *-stO* (left panel) and compounds with simplex left constituents (right panel).

The coefficients for the interactions listed in Table 4.7 are all positive, which implies that $C_1 > C_2$ and $K_1 > K_2$. Apparently, the left (and right) family measures receive greater weight from compound frequency than compound frequency from the family measures. In other words, the compound's own probability has priority. The more C_1 (or K_1) increases with respect to C_2 (or K_2), the greater the inhibitory force of the interaction. The bottom panels of Figure 4.3 visualize the interactions of of compound frequency by left constituent family size, for compounds with left constituents ending in *-stO* (lower left panel) and compounds with simplex left constituents (lower right panel). For the compounds in *-stO*, we effectively have a floor effect, with a maximum for the amount of facilitation that never exceeds the maximum for any of the marginal effects. For the bimorphemic compounds, maximum facilitation is obtained only when compound frequency is large and family size is **small**. In terms of morphological processing, the observed interaction may receive the following interpretation. There is a balance between the contributions of compound frequency and left constituent family size to the ease of compound recognition. The effect of the family size may differ from facilitatory (as in the compounds with *-stO*) to slightly inhibitory (as in the bimorphemic compounds), see the lower panels of Figure 4.3. As we argued above, this may reflect the potentially dual impact of constituent families: A large family may come with easier access to the target compound due to the increased resting activation level of the family members, but it also brings along a larger number of competitors, which need to be inhibited in order for the target compound to be recognized. Crucially, regardless of the direction of the left constituent family size effect, the larger the morphological family, the more processing resources are allocated to it and the less impact is elicited by compound frequency. Again, we witness how the magnitude of some processing cues modulates the utility of the cues for compound recognition.

Since we focus on lexical distributional predictors in this version of the model, our formulation in (4.16) leaves out the interaction of right constituent frequency by word length attested for the right subgaze duration. The effect of length might be brought into the model, however, by conditioning on lexical subsets of the appropriate length. In particular, PROMISE is expected to support the finding of Bertram and Hyönä (2003) that the left constituent frequency effect becomes weak for short Finnish compounds. We leave this issue to future research.

The PROMISE model is a formalization of the idea that readers and listeners maximize their opportunities for recognition of complex words (see Libben, 2006 and Chapter 3 of this dissertation). Parameters of PROMISE can be directly estimated from the regression coefficients of statistical models. As we have shown, estimated values of parameters do not only shed light on which sources of information are preferred over others, but also specify at what timesteps of the visual uptake and at what cost to the processing system. Importantly, PROMISE is not restricted to compounding as a type of morphological complexity, nor to long polymorphemic words. The model allows dealing with word length and morphological complexity (e.g., simplex, inflected, derived or compound words) in a principled probabilistic way. As a research perspective, a series of experiments involving a broad spectrum of languages and word lengths would be desirable to quantify the range of opportunities that morphological structure offers for efficient recognition of complex forms. We also believe that PROMISE can be easily incorporated into general models of eye-movement control in reading, such as E-Z Reader or SWIFT, extending the line of research of Pollatsek, Reichle and Rayner (2003). Consideration of parameters of PROMISE along with other visual and lexical parameters may improve predictions of such models for the processing of complex morphological structures.

References

Allen, M. and Badecker, W. (2002). Inflectional regularity: Probing the nature of lexical representation in a cross-modal priming task. *Journal of Memory and Language*, 46:705–722.

Andrews, S., Miller, B., and Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, 16(1/2):285–311.

Baayen, R. H. (1994). Productivity in language production. *Language and Cognitive Processes*, 9:447–469.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge University Press, Cambridge.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *In press*.

Baayen, R. H., Feldman, L. B., and Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55:290–313.

Baayen, R. H. and Schreuder, R. (1999). War and peace: morphemes and full forms in a non-interactive activation parallel dual route model. *Brain and Language*, 68:27–32.

Baayen, R. H. and Schreuder, R. (2000). Towards a psycholinguistic computational model for morphological parsing. *Philosophical Transactions of the Royal Society (Series A: Mathematical, Physical and Engineering Sciences)*, 358:1–13.

Bates, D. M. and Sarkar, D. (2005). The lme4 library. *[On-line], Available: http://lib.stat.cmu.edu/R/CRAN/*.

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

Bien, H., Levelt, W., and Baayen, R. (2005). Frequency effects in compound production. *PNAS*, 102:17876–17881.

Clark, J. and O'Regan, J. (1999). Word ambiguity and the optimal viewing position in reading. *Vision Research*, 39:842–857.

de Almeida, R. and Libben, G. (2005). Changing morphological structures: The effect of sentence context on the interpretation of structurally ambiguous English trimorphemic words. *Language and Cognitive Processes*, 20:373–394.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and 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., and Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, 15:329–365.

Duñabeitia, J. A., Perea, M., and Carreiras, M. (2007). The role of the frequency of constituents in compound words: Evidence from Basque and Spanish. *Psychonomic Bulletin & Review*, 14:1171–1176.

Giraudo, H. and Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review*, 8:127–131.

Hyönä, J., Bertram, R., and Pollatsek, A. (2004). Are long compound words identified serially via their constituents? Evidence from an eye-movement-contingent display change study. *Memory and Cognition*, 32:523–532.

Hyönä, J. and 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.

Hyönä, J. and Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental*

Psychology: Human Perception and Performance, 24:1612–1627.

Inhoff, A. and Eiter, B. (2003). Knowledge of word length does not constrain word identification. *Psychological Research*, 67:1–9.

Inhoff, A. W., Radach, R., and Heller, D. (2000). Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory and Language*, 42:23–50.

Järvikivi, J., Bertram, R., and Niemi, R. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and Cognitive Processes*, 21:394–431.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., and Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, 94:223–244.

Krott, A., Baayen, R. H., and Schreuder, R. (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics*, 39(1):51–93.

Krott, A., Libben, G., Jarema, G., Dressler, W., Schreuder, R., and Baayen, R. H. (2004). Probability in the grammar of German and Dutch: Interfixation in tri-constituent compounds. *Language and Speech*, 47:83–106.

Krott, A. and Nicoladis, E. (2005). Large constituent families help chidren parse compounds. *Journal of Child Language*, 32(1):139–158.

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2007). Reading of polymorphemic Dutch compounds: Towards a multiple route model of lexical processing. *Manuscript submitted for publication*.

Laine, M. and Virtanen, P. (1999). *WordMill Lexical Search Program*. Center for Cognitive Neuroscience, University of Turku, Finland.

Laudanna, A. and Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 345–364. Lawrence Erlbaum Associates, Hillsdale, N. J.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177.

Libben, G. (2005). Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *Canadian Journal of Linguistics*, 50:267–283.

Libben, G. (2006). Why study compound processing? In Libben, G. and

Jarema, G., editors, *The representation and processing of compound words*, pages 1–23. Oxford University Press, Oxford.

Lima, S. D. and Pollatsek, A. (1983). Lexical access via an orthographic code? The basic orthographic syllabic structure (BOSS) reconsidered. *Journal of Verbal Learning and Verbal Behavior*, 22:310–332.

Meunier, F. and Longtin, C. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, 56:457–471.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94:1–18.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., and Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30:1271–1278.

Nicoladis, E. and Krott, A. (2007). Word family size and French-speaking children's segmentation of existing compounds. *Language Learning*, 57(2):201–228.

O'Regan, J. (1979). Saccade size control in reading: Evidence for the linguistic control hypothesis. *Perception and Psychophysics*, 25:501–509.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Statistics and Computing. Springer, New York.

Pollatsek, A. and Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, 20:261–290.

Pollatsek, A., Hyönä, J., and Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26:820–833.

Pollatsek, A., Reichle, E., and Rayner, K. (2003). Modeling eye movements in reading: Extensions of the E-Z Reader model. In Hyönä, Y., Radach, R., and Deubel, H., editors, *The mind's eye: Cognitive and applied aspects of eye movement research*, pages 361–390. Elsevier, Amsterdam.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422.

Rayner, K., Well, A., Pollatsek, A., and Bertera, J. (1982). The availability of useful information to the right of fixation in reading. *Perception and Psychophysics*, 31:537–550.

Schreuder, R. and Baayen, R. H. (1995). Modeling morphological processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 131–154. Lawrence Erlbaum, Hillsdale, New Jersey.

Sereno, J. and Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, 25:425–437.

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*. Lawrence Erlbaum, Hove, U.K.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology*, 57A:745–765.

Taft, M. and Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. *Mental Lexicon*, 1:183–189.

Taft, M. and Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14:638–647.

Taft, M. and Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, 15:607–620.

Van Jaarsveld, H. J. and Rattink, G. E. (1988). Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of Psycholinguistic Research*, 17:447–473.

Virtanen, P. and Pajunen, A. (2000). *ContextMill Computer Software*. General Linguistics, University of Turku, Finland.

Appendix

Table 4.3: Item characteristics per compound type

Predictor	No suffix	*-stO*	*-Us*
WordLength	12.2 (1.4)	13.9 (1.7)	12.5 (1.5)
WordFreq	51.4 (66.0)	17.7 (16.1)	88.0 (121.5)
LeftFreq	3253.6 (4362.3)	925.2 (1091.1)	1494.0 (1949.4)
RightFreq	3008.0 (2615.1)	5246.2 (5407.7)	9917.5 (12578.9)
LeftFamSize	195.2 (165.9)	88.4 (156.3)	104.1 (95.8)
RightFamSize	243.9 (199.1)	384.8 (361.5)	522.9 (389.3)

Numbers in columns 2-4 show mean values and standard deviations (in brackets) for predictors per compound type.

Key to Table 4.4: Predictors of primary interest for this study are presented in the main body of paper. Additional control variables that show significant effects in our statistical models are as follows: *NextLength*, length of the word to the right of the target word; *NextSkipped*, indicator of whether the word following the target is skipped during reading; *LeftLength*, length of the compound's left constituent; *InitTrigramFreq*, token-based frequency of the word-initial trigram (based on 22.7 million corpus of written Finnish); *AverageBigramFreq*, average bigram frequency across the target word (based on 22.7 million corpus of written Finnish); *LastSaccade*, amplitude of the saccade preceding the fixation; *NextSaccade*, amplitude of the saccade following the fixation; *FixPos* and *FixPos2*, first fixation position and its squared value; *Nomore*, indicator of whether the fixation is word-final; and *Sex*, participants' gender. Table 4.4 summarizes continuous (dependent and independent) variables, which show significant effects in our statistical models. In addition to these, we have considered a large number of control variables that were not significant predictors of reading times or probabilities. These included: transitional probabilities of word pairs N-1 and N and words N and N+1 (computed with the help the ContextMill software, Virtanen & Pajunen, 2000); frequencies of words N-1 and N+1; length of word N-1; frequency of the word-final trigram; word position in the sentence; and the total number of words in the sentence.

Key to Tables 4.5-4.9 and to estimating effect sizes for the models' predictors: Throughout the tables, the second column shows estimates of the regression coefficients for the model's predictors. Columns 3-6 provide information on the distributions of those estimates obtained via the Monte Carlo Markov chain

Table 4.4: Summary of Continuous Variables Reported in Statistical Models.

Variable	Range (Adjusted Range)	Mean(SD)	Median
FixPos	0.1:16 characters (1:160 pixels)	37.1(21.8)	35.1
FirstDuration	67:735 ms (4.2:6.6 log units)	5.4(0.3)	5.4
SubgazeLeft	60:1808 ms (4.1:7.5 log units)	5.8(0.5)	5.7
SubgazeRight	81:812 ms (4.4:6.7 log units)	5.5(0.4)	5.5
GazeDuration	60:1998 ms (4.2:7.6 log units)	6.1(0.6)	6.2
LastSaccade	1:15 characters (10:151 pixels)	70.8(27.9)	70.5
NextSaccade	-12:19 characters (-112:189 pixels)	46.3(55.2)	54.7
NextLength	2:13 characters	4.9(3.1)	4
WordLength	10:18 characters (-3.1:4.9)	0.0(1.7)	-0.12
LeftLength	4:14 characters	7.5(1.4)	8
InitTrigramFreq	3:601 (1.1:6.4 log units)	4.3(1.0)	4.5
AverageBigramFreq	2:151 (0.7:5.0 log units)	4.1(0.9)	4.3
WordFreq	2:665 (-2.2:3.6 log units)	0.1(1.4)	0.1
ResidLeftFreq	11:1.8*10^4 (-4.1:3.1 log units)	0.0(1.5)	0.1
RightFreq	33:8.1*10^4 (-4.5:3.3 log units)	0.0(1.4)	0.14
ResidLeftFamilySize	2:812 (-3.0:1.7)	0.0(0.9)	0.1
ResidRightFamilySize	3:1808 (-2.0:1.3)	0.0(0.6)	-0.1
ResidBaseFreq	49:3.3*10^4 (-2.8:4.0)	0.0(1.2)	-0.2
TrialNum	11:272	142.1(76.3)	143

Numbers in the second column show original value ranges for predictors. If any transformations have been made to the original values for statistical reasons (i.e., natural log transformation, decorrelation with other predictors or centering), the numbers in the brackets show the ranges actually used in statistical models. Means, standard deviations and median values refer to the predictor values used in the models. Values for frequency and family size measures are based on the corpus with 22.7 million word-forms.

(MCMC) random-walk method using 1000 simulations: this information is useful for evaluating stability of the models' predictions. The third column shows the MCMC estimate of the mean for each predictor, while the fourth and the fifth columns show highest posterior density intervals, which are a Bayesian measure for the lower and upper bounds of the 95% confidence interval, respectively. The sixth column provides a p-value obtained with the help of MCMC simulations; and the final column provides less conservative p-values obtained with the t-test using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom.

For the predictors of primary interest for this study we report effect sizes, either in the body of the paper or in Tables 4.1 and 4.2. These were obtained as follows. Our models used contrast coding for discrete variables. Therefore, the effect size for factors was calculated as the difference between (i) the (exponentially-transformed) sum of the intercept value and the contrast regression coefficient, $\hat{\beta}$, and (ii) the (exponentially-transformed) intercept value. Exponential transformation was only applied, when the dependent variable had log-transformed values, i.e. fixation or gaze duration. For instance, the effect size of the indicator of whether the word after the target word is skipped (*NextSkipped*) on gaze duration, after log gaze duration is back-transformed to original values in milliseconds, is:

exp$(\text{Intercept}+\hat{\beta})-exp(\text{Intercept}) =$ exp$(5.9+0.105)-$exp$(5.9) = 40$ ms,

where *Intercept* is the intercept of the model for gaze duration $(= 5.9)$ and $\hat{\beta}$ is the contrast coefficient for *NextSkipped* $(= 0.105)$.

Effect sizes for simple main effects of numeric variables were calculated as the difference between the (exponentially-transformed) model's predictions for the minimum and maximum values of a given variable. For instance, the regression coefficient, $\hat{\beta}$, associated with compound frequency, *WordFreq*, in the model for first fixation duration is -0.0111, while the range of values, *Min*:*Max*, used in that model for *WordFreq* and obtained via the operation of centering, is $-2.2 : 3.6$, see Table 4.4. To compute the effect size for log-transformed dependent measures, like first fixation duration, we used the following formula:

exp$(\text{Intercept}+\hat{\beta}*\text{Max})-exp(\text{Intercept}+\hat{\beta}*\text{Min})$.

The effect of *WordFreq* (i.e., the difference between the model's predictions for the lowest-frequency and the highest-frequency target words) on first fixation duration is then:

exp$(5.2+-0.0111*3.6)-$exp$(5.2+-0.0111*-2.2) = -11.6$ ms

Computation of effect sizes for interactions involved obtaining model predictions for the extreme values of one term in the interaction of interest, while holding all other terms in that model (and in that interaction) constant at their median values. Again, the estimate of the effect size for an interacting variable was calculated as a difference between the (exponentially-transformed) values of the regression function corresponding to the minimum and the maximum values of that variable. To estimate the effect sizes for interactions we also used conditioning plots that are not explained here (for detailed treatment, see Baayen, 2008).

Table 4.5: First Fixation Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.2048	5.2060	5.1153	5.3001	0.001	0.0000
SuffixTypeSt	-0.0131	-0.0131	-0.0500	0.0207	0.458	0.4269
SuffixTypeUs	0.0143	0.0137	-0.0204	0.0463	0.428	0.3549
ResidLeftLength	-0.0099	-0.0095	-0.0196	0.0016	0.088	0.0533
NextSaccade	0.0010	0.0010	0.0008	0.0013	0.001	0.0000
LastSaccade	0.0013	0.0013	0.0009	0.0017	0.001	0.0000
WordFreq	-0.0111	-0.0109	-0.0179	-0.0033	0.008	0.0019
TrialNum	-0.0001	-0.0001	-0.0002	0.0000	0.158	0.1303
FixPos	0.0025	0.0025	0.0014	0.0036	0.001	0.0000
FixPos2	0.0000	0.0000	0.0000	0.0000	0.001	0.0000
NomoreTRUE	0.1194	0.1173	0.0718	0.1633	0.001	0.0002
RightFreq	-0.0080	-0.0079	-0.0161	-0.0010	0.044	0.0286
WordLength	-0.0066	-0.0064	-0.0137	-0.0003	0.062	0.0316
InitTrigramFreq	0.0072	0.0069	-0.0035	0.0177	0.190	0.1276
NextLen	0.0010	0.0009	-0.0022	0.0041	0.602	0.5148
ResidLeftFreq	-0.0129	-0.0128	-0.0196	-0.0057	0.002	0.0001
ResidFamSizeL	-0.0138	-0.0142	-0.0262	-0.0043	0.012	0.0062
SubjectSexM	-0.0069	-0.0085	-0.1112	0.0916	0.876	0.8958
SuffixTypeSt:ResidLeftLength	0.0229	0.0223	-0.0008	0.0466	0.068	0.0356
SuffixTypeUs:ResidLeftLength	0.0007	0.0000	-0.0235	0.0260	0.962	0.9526
SuffixTypeSt:NextSaccade	0.0000	0.0000	-0.0004	0.0003	0.888	0.8410
SuffixTypeUs:NextSaccade	-0.0002	-0.0002	-0.0006	0.0002	0.276	0.2698
RightFreq:WordLength	0.0016	0.0015	-0.0026	0.0057	0.494	0.4475
NomoreTRUE:SubjectSexM	-0.0620	-0.0758	-0.1403	-0.0070	0.026	0.2254

Table 4.6: Model for for Subgaze Duration for the Left Constituent

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.7703	5.7719	5.6822	5.8638	0.001	0.0000
WordLength	0.0219	0.0221	0.0072	0.0376	0.004	0.0046
WordFreq	-0.0471	-0.0469	-0.0646	-0.0283	0.001	0.0000
ResidLeftLength	0.0594	0.0600	0.0406	0.0802	0.001	0.0000
ResidFamSizeL	-0.0431	-0.0431	-0.0887	-0.0016	0.044	0.0529
SuffixTypeSt	0.0456	0.0451	-0.0206	0.1095	0.188	0.1796
SuffixTypeUs	0.0247	0.0242	-0.0328	0.0788	0.426	0.4044
ResidLeftFreq	-0.0219	-0.0216	-0.0460	0.0037	0.096	0.0713
SuffixTypeSt:ResidLeftFreq	-0.0384	-0.0396	-0.0804	0.0033	0.068	0.0608
SuffixTypeUs:ResidLeftFreq	0.0152	0.0148	-0.0220	0.0484	0.408	0.3948
ResidFamSizeL:SuffixTypeSt	-0.0814	-0.0835	-0.1526	-0.0136	0.008	0.0227
ResidFamSizeL:SuffixTypeUs	0.0316	0.0321	-0.0308	0.0821	0.250	0.2792

Table 4.7: Model for Subgaze Duration for the Right Constituent

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.4395	5.4387	5.3463	5.5407	0.001	0.0000
WordLength	0.0187	0.0189	0.0082	0.0295	0.002	0.0005
WordFreq	-0.0230	-0.0225	-0.0347	-0.0084	0.001	0.0006
TrialNum	0.0000	0.0000	-0.0003	0.0004	0.798	0.8069
ResidLeftLength	-0.0489	-0.0490	-0.0653	-0.0330	0.001	0.0000
SuffixTypeSt	0.1177	0.1208	0.0420	0.2107	0.001	0.0063
SuffixTypeUs	-0.0040	-0.0023	-0.0783	0.0811	0.950	0.9232
ResidFamSizeL	0.0259	0.0257	-0.0023	0.0554	0.084	0.0850
RightFreq	-0.0439	-0.0435	-0.0653	-0.0213	0.001	0.0001
NextSkipped	0.0777	0.0782	0.0329	0.1226	0.001	0.0003
NextLen	0.0079	0.0079	0.0007	0.0146	0.020	0.0180
ResidFamSizeR	-0.0024	-0.0022	-0.0303	0.0257	0.886	0.8711
TrialNum:SuffixTypeSt	-0.0008	-0.0009	-0.0013	-0.0004	0.001	0.0007
TrialNum:SuffixTypeUs	-0.0003	-0.0003	-0.0008	0.0001	0.228	0.2583
SuffixTypeSt:ResidFamSizeL	-0.0545	-0.0538	-0.1023	-0.0009	0.044	0.0345
SuffixTypeUs:ResidFamSizeL	-0.0282	-0.0277	-0.0679	0.0135	0.180	0.1808
WordLength:RightFreq	-0.0155	-0.0156	-0.0220	-0.0081	0.001	0.0000
WordFreq:ResidFamSizeL	0.0210	0.0210	0.0076	0.0367	0.004	0.0055
RightFreq:NextLen	0.0085	0.0084	0.0042	0.0123	0.001	0.0000
WordFreq:ResidFamSizeR	0.0242	0.0244	0.0051	0.0478	0.028	0.0222

Table 4.8: Model for Gaze Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.8979	5.9073	5.6691	6.1598	0.001	0.0000
WordLength	0.0540	0.0538	0.0376	0.0687	0.001	0.0000
TrialNum	-0.0001	-0.0002	-0.0003	0.0001	0.140	0.1633
WordFreq	-0.0303	-0.0302	-0.0514	-0.0123	0.004	0.0018
ResidLeftFreq	-0.0130	-0.0133	-0.0355	0.0122	0.268	0.2833
ResidFamSizeL	-0.0201	-0.0198	-0.0633	0.0261	0.376	0.3745
SuffixTypeSt	0.3112	0.3046	0.0512	0.5812	0.018	0.0227
SuffixTypeUs	0.3682	0.3636	0.0781	0.6204	0.010	0.0077
AverageBigramFreq	0.0638	0.0616	0.0158	0.1056	0.006	0.0063
ResidFamSizeR	-0.0079	-0.0087	-0.0543	0.0271	0.708	0.7075
SubjectSexM	-0.0385	-0.0370	-0.2782	0.2251	0.778	0.7580
NextSkipped	0.1051	0.1047	0.0711	0.1362	0.001	0.0000
SuffixTypeSt:AverageBigramFreq	-0.0623	-0.0604	-0.1257	0.0029	0.066	0.0636
SuffixTypeUs:AverageBigramFreq	-0.0821	-0.0810	-0.1442	-0.0171	0.010	0.0114
ResidLeftFreq:SuffixTypeSt	-0.0538	-0.0538	-0.0896	-0.0109	0.006	0.0076
ResidLeftFreq:SuffixTypeUs	0.0230	0.0228	-0.0186	0.0575	0.228	0.2028
ResidFamSizeL:SuffixTypeSt	-0.1233	-0.1239	-0.1987	-0.0574	0.002	0.0007
ResidFamSizeL:SuffixTypeUs	0.0206	0.0206	-0.0419	0.0760	0.452	0.4881
WordFreq:ResidFamSizeR	0.0535	0.0533	0.0257	0.0854	0.002	0.0005
TrialNum:SubjectSexM	-0.0007	-0.0007	-0.0010	-0.0003	0.001	0.0001

Table 4.9: Random effects for *FirstFixDur*, *SubgazeLeft*, *SubgazeRight* and *GazeDur*

A. First fixation duration				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.015	0.025	0.011	0.045
Subject	0.106	0.114	0.084	0.156
Subject by Nomore	0.068	0.025	0.083	0.156
Residual	0.265			
B. Subgaze duration for the left constituent				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.104	0.104	0.085	0.130
Subject	0.195	0.198	0.151	0.271
Residual	0.446			
C. Subgaze duration for the right constituent				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.009	0.012	0.003	0.044
Subject	0.168	0.171	0.129	0.227
Residual	0.368			
D. Gaze duration				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.113	0.114	0.095	0.139
Subject	0.298	0.303	0.233	0.398
Residual	0.394			

Affixal Salience at Work: Morphological Processing of Derived Words in Dutch

Chapter 5

This chapter has been submitted as a paper to *Journal of Memory and Language* as Victor Kuperman, Raymond Bertram and R. Harald Baayen, Affixal Salience at Work: Morphological Processing of Derived Words in Dutch.

Abstract

This eye-tracking study explores effects of morphological structure on the lexical processing of Dutch suffixed words in sentential reading. We show that affixal salience crucially moderates the use of morphological properties. In words with shorter suffixes, we observe a strong negative effect of derived word frequency and a weak negative effect of suffix family size (i.e. the number of words ending in the given suffix) on reading times. As suffix length increases, the former effect vanishes and the latter increases, which points at the interplay between storage and computation in complex word recognition. We also observe higher processing costs if there is an imbalance in the family sizes of the base word and the suffix. We model this effect using relative entropy and interpret it as an independent dimension of parsing complexity. Finally, we explore which aspects of affixal salience show effects on reading times when considered against the backdrop of other predictors and across many Dutch suffixes, and we provide explanations for discrepancies with earlier reports.

Introduction

One of the aims of psycholinguistic research in morphological processing is to pin down the characteristics of complex words that bias the lexical processing system towards storage (i.e., memorizing and recognizing complex words as unstructured units) or computation (i.e., decomposition of complex words into morphemes, from which the meaning of whole words is computed on-line). The present chapter addresses the interplay of storage and computation during the visual recognition of derived words. Each of the options, storage and computation, has been instantiated in models of morphological processing. Obligatory decomposition of complex words characterizes the sublexical models of morphological processing proposed, for instance, in Taft and Forster (1975) and Pinker (1999). These models predict a central role for bases and affixes during lexical access, and a peripheral role of full-forms, in the recognition of derived words (see also Fiorentino & Poeppel, 2007; Taft, 2004). Conversely, obligatory full-form access is a hall-mark of supralexical models (e.g., Giraudo & Grainger, 2000). Supralexical models do not expect morphemes to influence substantially the costs of lexical access, since the activation of morphemes is only attributed to the post-access processing stage. Finally, dual-route parallel models (e.g., Baayen & Schreuder, 1999; Baayen & Schreuder, 2000; Schreuder & Baayen, 1995) allow morphological decomposition and full-form access to operate jointly. On these models, bases and affixes of the derived words are activated to the extent that lexical processing makes use of the decomposition route.

Recent eye-tracking experiments on Dutch and Finnish compounds (Chapters 3 and 4 of this dissertation: Kuperman, Bertram & Baayen, 2008a; Kuperman, Schreuder, Bertram & Baayen, 2008b) have shown that the complexity of morphological processing may not be fully captured by any of the models outlined above. One of the crucial findings in both studies is the early effect of frequency of the compound (e.g., Dutch *oorlogsverklaring* "declaration of war") on reading times, which emerges as early as the first fixation, that is, before the whole long word can be visually inspected. The compound frequency effect is simultaneous with the effect of the lexical properties of the compound's left constituent (*oorlog* "war") and precedes the effects pertaining to the compound's right constituent (*verklaring* "declaration"). Taken together, these results are problematic both for sublexical and supralexical models of morphological processing. Full-form access before the right constituent is processed is at odds with the former models, while activation of morphological constituents simultaneously with (and not after) full-form activation goes against the latter models. Crucially, the compound frequency effect is lingering throughout the entire recognition process and overlaps in time with frequency-based effects of both compound's constituents. This shows that the use of the decompositional route and the full-form route in lexical access does not follow the categorical "winner-takes-it-all" principle of the dual-route parallel processing advocated in Schreuder and Baayen (1995). Yet another piece of evidence pointing at the intricate balance between processing routes in morphological processing is the finding of interactions between properties of the full-form and those of the morphological constituents (see e.g., interactions of compound frequency by left and right constituent family size in Chapter 4), which are not expected in either the sublexical or the supralexical theoretical framework.

To account for the complex pattern of results in reading of compounds, we proposed in Chapter 4 a PRObabilistic Model of Information SourcEs (henceforth, PROMISE). The core assumption of the model is that all sources of morphological information (i.e., morphemes, combinations of morphemes, morphological paradigms and morphological selectional constraints) may be accessed in the course of complex word recognition, as soon as (partial) information about these sources becomes available. The model further argues that reading times are proportional to the amount of information present in the system at any given point of time (cf. Levy, 2008; Moscoso del Prado Martín, Kostić & Baayen, 2004). PROMISE quantifies the amount of information carried by morphological stucture by considering various probabilities of morphemes, including conditional probabilities of morphemes given their morphological and orthographic context. Importantly, the PROMISE model allows for the parallel interactive processing of multiple sources of morphological information. The relative contributions of information sources to the speed of processing are estimated directly from coefficients of regression models for reading times. The present chapter sets as its first goal the application of the PROMISE model developed for the processing of compounds to a different type of word-formation, namely, morphological derivation.

Derivation is different from compounding in several crucial ways that challenge predictions of PROMISE about the balance of storage and computation in morphological processing. For instance, morphological families of affixes (i.e., all words that share an affix in the same position) are much larger and arguably more semantically diverse than morphological families of compounds' constituents (i.e., sets of compounds that share a constituent). Compare, for instance, base words in the morphological family of the suffix *-ness* (e.g., *good-ness*, *mad-ness*, and *abrupt-ness*) and modifiers in the morphological family of the noun *cream* (e.g., *vanilla cream*, *ice cream*, and *chocolate cream*). Effects of constituent families of compounds on processing times are traditionally argued to reflect the amount of semantic resonance that the word that is being recognized receives from morphologically and semantically related words (e.g., Schreuder & Baayen, 1997; De Jong, Schreuder & Baayen, 2000; 2003). Given the size and diversity in morphological families of affixes, we expect distributional properties of those families to elicit effects that qualitatively differ from the facilitatory effects of constituent families shown for lexical processing of compounds. Furthermore, most derivational affixes are bound morphemes (e.g., *-ing* in *reading*), unlike most right constituents of compounds (e.g., *way* in *railway*). Thus, affixes are invariably recognized in the context of their morphological bases. Since the PROMISE model quantifies conditional (contextual) probabilities of constituents of derived words with the help of morphological families of those constituents, we expect the constituent families to play a bigger role in the processing of derived words than frequencies of occurrence of constituents. This study aims to establish the validity of predictions of the PROMISE model for derived words, based on a large-scale regression study on a broad range of Dutch suffixed words embedded in sentential contexts.

The second goal of the present chapter is to contribute to the current knowledge of what properties of derived words, their bases and affixes codetermine visual recognition by shifting the balance between storage and computation. Laudanna and Burani (1995) introduced the concept of affixal salience, defined as the likelihood of recognizing the affix as a processing unit in its own right. The idea is that the more perceptual salience an affix has, the more it stands out of its embedding word, and the more biased lexical processing is towards morphological decomposition and towards using the properties of the base and the affix for the identification of the complex word. A wide range of studies have proposed dimensions that would increase affixal salience: orthographic properties of affixes (e.g., affix length, affixal confusability and transitional probabilities of n-grams near the morphemic boundary, e.g., Andrews & Davis, 1999; Laudanna & Burani, 1995), their phonological and phonotactic properties (e.g., co-occurrence probabilities of n-phones and of discontinuous patterns across the morphemic boundary, e.g., Bertram, Pollatsek & Hyönä, 2004; Hay & Baayen, 2003), and their lexical properties (e.g., word formation type of the affix, existence of inflectional allomorphs or homonyms for the affix, cf. Baayen, 1994; Bertram, Schreuder

& Baayen, 2000; Bertram, Laine, Karvinen, 1999; Bertram, Laine, Baayen, Schreuder, & Hyönä, 2000; Järvikivi, Bertram & Niemi, 2006; Sereno & Jongman, 1997). A related line of research explores the separability of the base word and the affix. For instance, Hay (2001) argues that the likelihood of the affix being separated from its base (i.e., likelihood of decomposition in the dual-route parallel processing) is lower, if the frequency of the derived word is higher than the frequency of the base word. Corpus-based examinations of distributional properties of affixes showed that affixes are more salient, and hence more separable from bases of derived words, if they are productive in language (e.g., occur in a larger number of different, frequent or new words), or if they occur outside of other affixes in derived words ending in two affixes (cf., Baayen, 1993; Baayen, 1994; Burzio, 1994; Hay, 2001; Hay & Baayen, 2002; Hay & Baayen, 2003; Hay & Plag, 2004; Plag, 1999).

Importantly, many dimensions of affixal salience have been proposed on the basis of experiments that considered only a small number of affixes and a small number of predictors at a time (often experimenting on pairs of affixes differing in only one dimension). Using step-wise multiple regression mixed-effects modeling with participants and items as crossed random effects (cf., Baayen, 2008; Bates & Sarkar, 2007; Pinheiro & Bates, 2000), we will consider many predictors simultaneously and will pit dimensions of affixal salience against a variety of other dimensions and control variables in the sentence reading task. As argued in Gries (2003), the burden of interpretation for this kind of research is two-fold: First, we need to explain how the important contributors to visual recognition of derived words affect the effort of processing of such words, and second, we need to show why some of the proposed predictors of affixal salience play no role in our study.

Method

Participants

Twenty-eight students of the Radboud University Nijmegen (21 females and 7 males) participated in this experiment for the reward of 6 euros. All were native speakers of Dutch and had normal or corrected-to-normal vision.

Apparatus

Eye movements were recorded with an EyeLink II eyetracker manufactured by SR Research Ltd. (Canada). The eyetracker is an infrared video-based tracking system combined with hyperacuity image processing. The eye movement cameras are mounted on a headband (one camera for each eye), but the recording was monocular (right eye) and in the pupil-only mode. There are also two infrared LEDs for illuminating the eye. The headband weighs 450 g in total. The cameras sample pupil location and pupil size at the rate of 500 Hz. Recording is performed by placing the camera and the two infrared light sources 4-6 cm away from the eye. Head position with respect to the computer screen is tracked with the help of a head-tracking camera mounted on the center of the headband at the level of the forehead. Four LEDs are attached to the corners of the computer screen, which are viewed by the head-tracking camera, once the participant sits directly facing the screen. Possible head motion is detected as movements of the four LEDs and is compensated for on-line from the eye position records. The average gaze position error of EYELINK II is $<0.5^o$, while its resolution is 0.01^o. The stimuli were presented on a 17-inch computer screen, which had a refresh rate of 60 Hz.

Materials

The set of target words included 156 Dutch bimorphemic words (e.g., *president+schap* "presidency") ending in one of the following derivational suffixes: *-sel*, *-nis*, *-ig*, *-te*, *-er*, *-ing*, *-es*, *-schap*, *-ster*, *-baar*, *-zaam*, *-lijk*, *-vol*, *-dom*, *-erig*, *-erij*, *-loos*, *-achtig*, and *-heid* (3 to 12 words per suffix). These nineteen suffixes were selected for inclusion in our study since they are reasonably productive in modern Dutch and they belong to the Germanic stratum[1]. To raise the likelihood that our target words are fixated during reading, we set the minimum length of those words to 8 characters (range = 8-14, mean = 9, SD = 1.3). Each target word was embedded without further inflectional suffixes into a separate sentence, and it never occupied the sentence-initial or sentence-final position. The experimental list also included 136 filler sentences with a different experimental manipulation: Analyses of these sentences are not reported here. All sentences comprised 6-17 words (mean = 11.2 words, SD = 2.2) and took up at most one line. The sentences were displayed one at a time starting at the central-left position on the computer screen. Stimuli were presented in fixed-width font Courier New size 12. With a viewing distance of about 80 cm, one character space subtended approximately 0.36^o of visual angle.

Sentences were presented in two blocks, while the order of sentences within the blocks was pseudo-randomized and the order of blocks was counterbalanced across participants. Approximately 15% of sentences were followed by a yes-no question pertaining to the content of the sentence. The experiment began with a

[1]Latinate affixation is marginal in Dutch as compared to English, and has been argued to be unproductive (cf. Van Marle, 1985).

practice session consisting of five filler sentences and two questions.

Procedure

Prior to the presentation of the stimuli, the eye-tracker was calibrated using a three-point grid that extended over the horizontal axis in the middle of the computer screen. Prior to each stimulus, correction of calibration was performed by displaying a fixation point in the central-left position. After calibration, a sentence was presented to the right of the fixation point.

Participants were instructed to read sentences for comprehension at their own pace and to press a "response" button on the button box. Upon presentation of a question, participants pressed either the "yes"-button or the "no"-button on the button box. If no response was registered after 3000 ms, the stimulus was removed from the screen and the next trial was initiated. Responses and response times of participants were recorded along with their eye movements. The experimental session lasted 50 minutes at most.

Dependent variables

The duration of the single fixation landing on the target word (*SingleDur*) and gaze duration (the summed duration of all fixations on the target word before fixating away from it, *GazeDur*) provided most insight into the naturalistic reading of derived words. Other dependent variables considered in this study included: initial fixation position, duration of the first fixation, amplitude of the first within-word saccade, probability of a single fixation on the word, as well as the total number of fixations on the word. All durational measures were log-transformed to reduce the influence of atypical outliers.

Predictors

The full list of predictors considered in this study is presented in Appendix, along with ranges, means and median values for numerical predictors (see Table 5.1). In what follows we describe predictors of primary interest for this study.

Distributional predictors: Previous research has shown that lexical processing of derived words is codetermined by the distributional properties of those words, as well as by the properties of their bases and affixes (cf., e.g., Baayen, 1994; Hay & Baayen, 2002, 2004; Niswander-Klement & Pollatsek, 2006). There is a considerable number of similar lexical-statistical measures that attempt to operationalize the intuition that more productive suffixes are easier to parse out; these are surveyed in detail in Hay and Baayen (2002). Since many of these predictors are highly correlated, we opted for considering only those variables the effects of which are well attested in the literature, namely suffix productivity (*SuffixProd*, number of word types in which the suffix occurs) and suffix frequency (*SuffixFreq*, number of word tokens in which the suffix occurs). All frequency-based measures described here and in the remainder of the section were (natural-)logarithmically transformed to decrease the influence of atypical outliers.

Higher frequencies and better morphological connectivity of constituents in compounds and derived words tend to increase the speed of visual recognition (Andrews, Miller & Rayner, 2004; De Jong, Schreuder & Baayen, 2000; Hyönä, Bertram & Pollatsek, 2004; Juhasz, Starr, Inhoff & Placke, 2003; Pollatsek, Hyönä & Bertram, 2000). Distributional properties of base words and derivations as whole words were estimated using the following variables: base word frequency, *BaseFreq* (lemma frequency of *president* in *presidentschap*), family size of the base, *BaseFamilySize* (the type-based count of derived words in which the base (*president*) occurs in the word-initial position)[2], and frequency of the whole derived word, *WordFreq* (e.g., lemma frequency of *presidentschap*). Computation of these distributional measures was based on the combined pool of roughly 120 million tokens, obtained from the CELEX lexical database (Baayen, Piepenbrock & Gulikers, 1995) and from the newspapers in the Twente News Corpus (Ordelman, 2002). All lemma frequency measures were collapsed over inflectional variants (*cat*, *cats*, *cat's* and *cats'*).

Other predictors: The lengths of derived words and suffixes, as measured in characters, phonemes and syllables, were taken into account. We also considered a broad range of predictors that were proposed in the literature as codeterminers of affixal salience, including affixal homonymy, confusability, structural invariance, as well as frequencies of bigrams preceding, following and straddling the morphemic boundary in the derived words. None of these predictors reached significance. We provide possible explanations for the discrepancies between our findings and the previously reported role of these predictors in the Results and Discussion section.

Statistical considerations: Several of our measures showed strong pair-wise correlations. Orthogonalization of such variables is crucial for the accuracy of predictions of multiple regression models. Teasing collinear variables apart is also advisable for analytical clarity, as it affords better assessment of the independent contributions of predictors to the model's estimate of the dependent variable (see Baayen, 2008). We orthogonalized every pair of variables

[2]Base family frequency (the token-based count of derived words in which the base *president* occurs in the word-initial position) did not reach significance in any of our models.

for which the Pearson correlation index r exceeded the threshold of 0.5. Decorrelation was achieved by fitting a regression model in which one of the variables in the correlated pair, e.g., suffix length in characters (*SuffixLength*) was predicted by the other variable, e.g., derived word's length in characters (*WordLength*). We considered the residuals of this model, *ResidSuffixLength*, as an approximation of the number of characters in the suffix, from which the effects of word length in characters were partialled out. Using the same procedure, we obtained *ResidSuffixLengthPhoneme* and *ResidSuffixLengthSyl* (both orthogonalized with *SuffixLength*), and *ResidSuffixProd* (orthogonalized with *SuffixFreq*). All orthogonalized measures were highly correlated with the measures from which they were derived (rs $>$ 0.5, p $<$ 0.0001). The index of collinearity between all numerical predictors before decorrelation was applied yielded an astronomic value of $\kappa = 9.3 * 10^{15}$. The index of collinearity was reduced dramatically for the set of decorrelated predictors in the final model for single-fixation duration ($\kappa = 13.9$).

In this study we made use of mixed-effects multiple regression models with participant and word as random effects (cf., Baayen, 2008; Baayen, Davidson & Bates, 2008; Bates & Sarkar, 2007; Pinheiro & Bates, 2000).

Unless noted otherwise, only those fixed effects are presented below that reached significance at the 5%-level in a backwards stepwise model selection procedure. The distribution of durational dependent measures was skewed even after the log-transformation of durations. Likewise, residuals of the mixed-effects models for durations were almost always skewed. To reduce skewness, we removed outliers from the respective datasets, i.e., points that fell outside the range of -2.0 to to 2.0 units of SD of the residual error of the model. Once outliers were removed, the models were refitted.

The random effects included in our models significantly improved the explanatory value of those models. Improvement was indicated by the significantly higher values of the maximum log likelihood estimate of the model with a given random effect as compared to the model without that random effect (all ps < 0.0001 using likelihood ratio tests).

Results and Discussion

The initial pool of data points comprised 6672 fixations. We removed fixations that were shorter than 50 ms and longer than 1000 ms (201 fixation, 3%). Subsequently, fixations that bordered microsaccades (fixations falling within the same letter) were removed (28 x 2 = 56 fixations, 0.8%). Finally, we only considered the fixations pertaining to the first-pass reading (i.e., the sequence of fixations made before the fixation is made outside of the word boundaries, 77% of the original dataset). As a result, we were left with a pool of 4916 valid fixations.

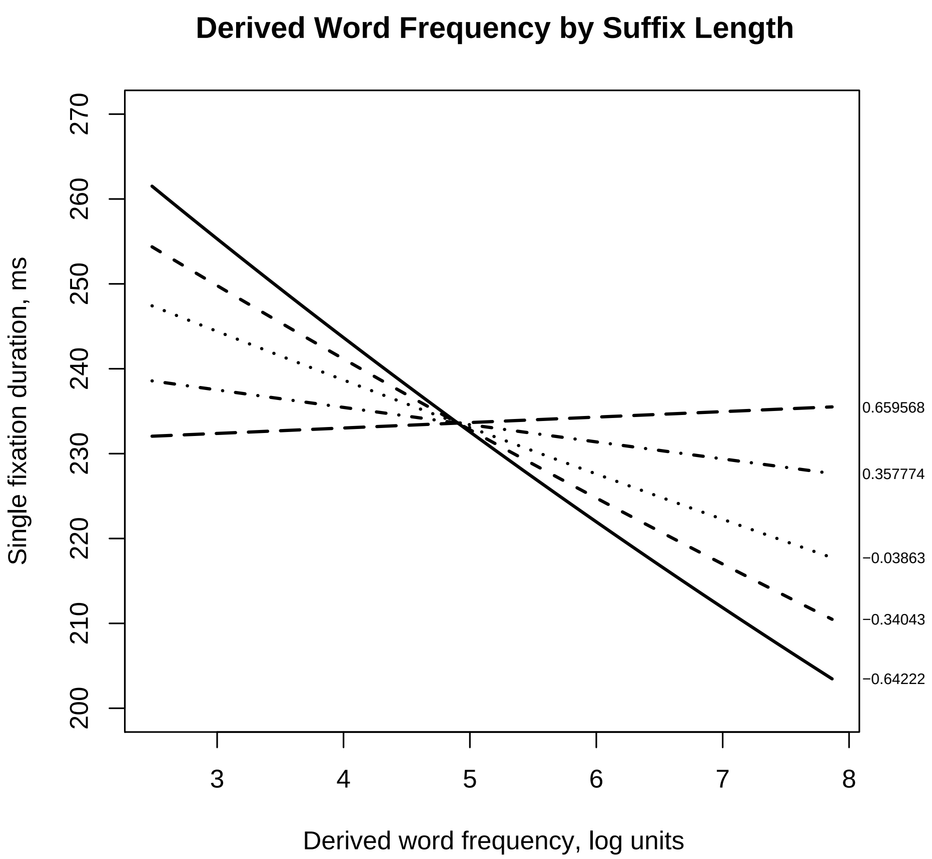
A negligible percent of the target words was skipped ($< 0.1\%$). Eighty-three percent of the target words required exactly one fixation, 16% required exactly two fixations, and only 1% required more than two fixations. The average number of fixations on a stimulus was 1.2 (SD = 0.4). The average fixation duration was 229 ms (SD = 64), and the average gaze duration was 262 ms (SD = 93).

Since the majority of target words elicited exactly one fixation, the analyses for gaze duration, first fixation duration and single-fixation duration yielded very similar results. We opted for providing in Appendix full specifications of the models for single-fixation duration (3267 data points, see Tables 5.2 and 5.3) and for gaze duration (3950 datapoints, Tables 5.4 and 5.5).

Specifications for all models include estimates of the regression coefficients; highest posterior density intervals (HPDs), which are a Bayesian measure of confidence intervals; p-values estimated by the Monte Carlo Markov chain (MCMC) method using 10000 samples; and p-values obtained with the t-test for fixed effects using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom (for the detailed treatment of the method, see Baayen, 2008; Baayen, Davidson & Bates, 2008; Pinheiro & Bates, 2000). For the effects reported in the body of the chapter we provide beta coefficients and p-values also estimated by the MCMC method using 10000 samples. Random effects for the final models for single-fixation duration and gaze duration are summarized in Table 5.6.

Main predictors of interest: In line with earlier robust findings in the eye-movements literature (e.g., Rayner, 1998), higher-frequency derived words elicited shorter single-fixation and gaze durations, with a main effect of about 4 ms decrease in single-fixation duration (6 ms decrease in gaze duration) per one log unit of frequency. Moreover, *WordFreq* entered into a significant interaction with the length of the suffix in phonemes, *ResidSuffixLenPhoneme* ($p_{mcmc} = 0.008$ for gaze durations and $p_{mcmc} = 0.007$ for gaze durations), such that the effect of *WordFreq* on single-fixation and gaze durations was strongest for derived words with shortest suffixes, gradually weakened in derived words with longer suffixes and virtually vanished in words with the longest suffixes, see the conditioning plot

Figure 5.1: Interaction of derived word frequency by suffix length for single-fixation duration. The lines plot the effect of derived word frequency for the quantiles of suffix length in phonemes (quantile values provided at the right margin). Derived word frequency comes with the strongest negative effect at the 1st quantile (solid line), the effect gradually levels off at the 2nd quantile (dashed line), the 3d quantile (dotted line) and the 4th quantile (dotdash line), and virtually vanishes in words with very long suffixes, the 5th quantile (longdash line).



for single-fixation durations in Figure 5.1. Apparently, the more complexity there is in the phonological representation of the suffix, the more salient that suffix is in the derived word and the more biased readers are towards using properties of the word's morphemes for lexical processing, rather than using as processing cues the properties of the complex word as a whole.

Both the frequency and the family size of the derivation's base codetermined reading times of derived words. These effects, however, were significantly qualified in the models for single-fixation duration and gaze duration by interactions of both variables with the measure of suffix productivity, i.e., the type count of words in which the suffix occurs, i.e., *ResidAffixProd*[3].

The effect of the interaction *BaseFreq* by *ResidSuffixProd* lost significance ($p_{mcmc} > 0.1$) when included in the models for single-fixation and gaze duration together with the interaction *BaseFamilySize* by *ResidSuffixProd*. Thus we only kept in our models the latter interaction, which was either well below or just above the 0.05-level of significance in both models ($p_{mcmc} = 0.02$ for single-fixation durations and $p_{mcmc} = 0.06$ for gaze durations), see the interaction plot for single-fixation durations in Figure 5.2. We note that the stronger effect of base family size, as compared to the base frequency effect, is in line with our anticipation that distributional properties of constituent families have more weight in the lexical processing of derived words than unconditional constituent frequencies of occurrence (for discussion see Balling & Baayen, 2008).

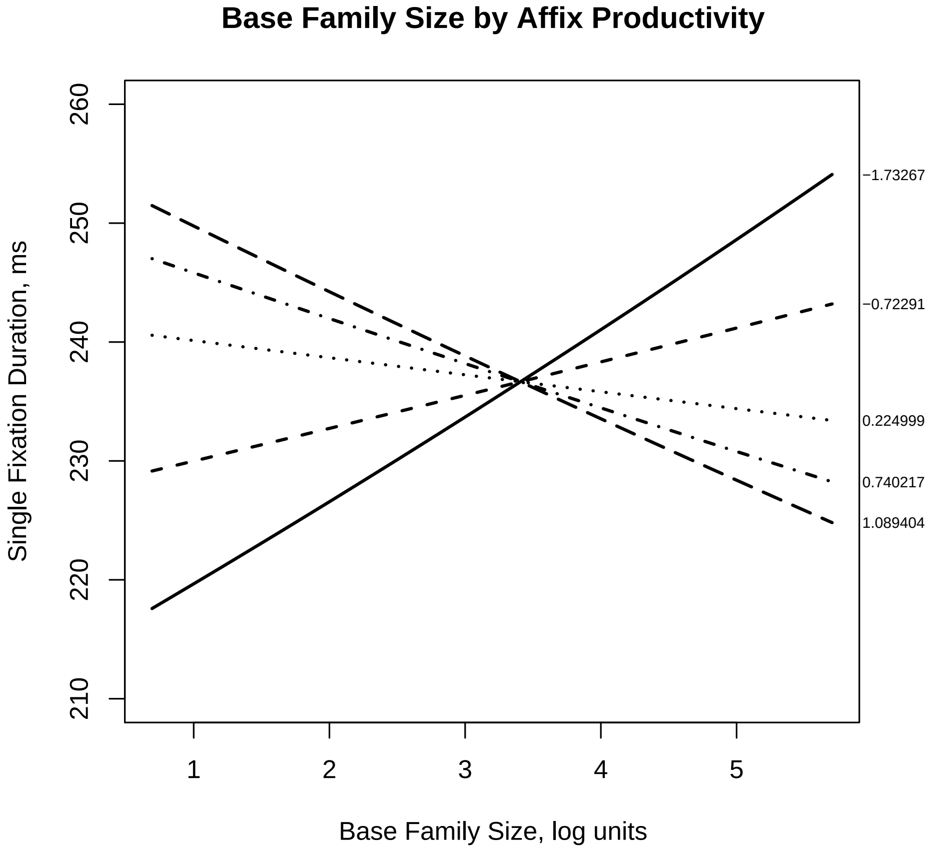
The interaction indicates that a large base family size came with longer reading times in those derived words that embed low-productivity suffixes (i.e., suffixes that can only combine with a small number of bases). Furthermore, the base family size effect reversed in words with suffixes that are relatively productive (i.e., can combine with a large number of bases). In other words, we observed an interaction of two morphological families, the one reflecting combinability of the base word (e.g., *happy*) with suffixes (*-ly, -ness, -less, -lessness*) and the other reflecting the ability of a given suffix (e.g., *-able*) to attach to a range of base words (e.g., *love, dispense, expand*). When both families are similar in size (both are small or both are large), the processing costs are minimal. The costs increase, however, if there is a substantial discrepancy in the sizes of the two families. In the General Discussion, we offer an interpretation for this interaction in the framework of PROMISE model.

Other predictors: Whole word length showed a strong effect in the model for gaze duration (with about a 21 ms increase for one additional character) and no significant effect in the model for single-fixation duration.

Longer words preceding the fixated word, *PrecLength*, induced longer single-fixation durations and gaze durations (a 2 ms and a 3 ms increase for one additional character, respectively). Additionally, the initial fixation position (*InitFixPos*) shows a well-attested inverse-U shape relationship with single-fixation duration, which we attribute to the Inverted Optimal Viewing Position effect

[3]We use the word type count as the measure of suffix productivity here. However, the interactions with base family size retain significance, even if we use — as alternative measures of suffix productivity — the count of hapax legomena in which the suffix occurs, or the growth rate of the lexicon (for detailed definitions, see Hay & Baayen, 2004). We opted for reporting only one of alternative measures.

Figure 5.2: Interaction of base family size by suffix productivity. The lines plot the effect of base family size for the quantiles of suffix productivity (quantile values provided at the right margin). Base family size comes with the strongest positive effect at the 1st quantile (solid line, lowest suffix productivity), the effect levels off at the 2nd quantile (dashed line), it reverses at the 3d quantile (dotted line) and the 4th quantile (dotdash line), and it is strongly negative in words with very large suffix families, the 5th quantile (longdash line).



discussed, for example, in Vitu, Lancelin & Marrier d'Unienville (2007).

Discrepancies with earlier reports

Our results demonstrate that the processing of derived words in reading is codetermined by a constellation of interacting phonological, distributional, and orthographic properties of derivations and their morphological constituents. These findings allow us to delimit the large number of proposed dimensions of affixal salience to only those dimensions (i.e., frequency-based characteristics of derived words, their morphological bases and suffixes, as well as lengths of derived words and suffixes) that show robust effects when considered against the backdrop of multiple control variables and of multiple competing dimensions.

We see it as an important task to explain why we find no evidence for some of predictors of affixal salience proposed in literature (cf., Gries, 2003). Specifically, we address below such predictors as homonymy, confusability and structural invariance of suffixes.

Several studies of derivation in English, Dutch and Finnish reported that derived words with homonymous suffixes (i.e., suffixes that can serve multiple syntactic or semantic functions, e.g., the English suffix *-er* in *warmer* and *builder*) tend to be processed as full-forms, rather than via their morphemes (e.g., Bertram, Laine, Baayen *et al.*, 2000a; Bertram, Schreuder & Baayen, 2000b; Sereno & Jongman, 1997). In Dutch, two derivational suffixes from our list exhibit homonymy, *-er* and *-te*. Both suffixes form different word classes in their respective syntactic functions: adjectives in the comparative form versus agentive nouns, for the suffix *-er* (cf., *warmer* "warmer" and *werker* "worker"), and verbs in the past tense versus nouns, for the suffix *-te* (cf., *hoopte* "hoped" and *lengte* "length"). Experiments on Dutch derivations established that derived words ending in the homonymous suffixes *-er* and *-te* showed effects of whole word frequency, but no effects of base frequency on lexical decision latencies (Bertram, Schreuder & Baayen, 2000). Yet we found no interaction between homonymy and either whole word frequency or base frequency in our derived words. Possibly, the number of words ending in the suffixes *-er* and *-te* in our experimental list was too small to offer sufficient statistical power to the test. Another, perhaps more likely, explanation for the lack of a homonymy effect may arise from the presence of sentential context in our experimental stimuli, and its absence from the lexical decision studies. The sentential context preceding a complex word may offer strong syntactic cues as to what the expected class is for the word under identification (a noun or an adjective in the comparative; a noun or a past-tense verb) and, consequently, may allow the reader to anticipate the morphosyntactic function of the suffix. No such disambiguating cues are available in experimental paradigms where words are presented in isolation. This lack of contextual constraint may have given rise to ambiguities in word identification and to task-specific differences in frequency effects reported for complex words with homonymous and non-homonymous suffixes.

Suffix confusability (the ratio of word types in which the character string functions as a suffix and all word types ending in that character string) has been argued to affect the balance between storage and computation in complex words, such that more confusable suffixes are less salient and their processing is biased towards storage (Laudanna & Burani, 1995). We observe no effect of confusability and argue that previously reported effects may also be artefacts of the experimental presentation of words in isolation. Syntactic cues provided by the sentential context preceding the target word (for instance, word class) may greatly reduce the ambiguity of whether the word-final characters represent a suffix or not. To test this hypothesis, we considered the four Dutch suffixes with the largest confusability ratios (*-es*, *-te*, *-er* and *-nis*). We conditioned by the word class the number of word types in which those character strings occurred in the word-final position. The resulting confusability ratios were reduced on average by a factor of 6.5. If readers anticipate word class given the preceding sentence fragment, the chances of confusing suffixes with non-morphological word endings are drastically reduced. This may explain the lack of the earlier reported effect in our data.

Finally, structural (in)variance, i.e., whether or not affixes change their orthographic form across inflectional paradigms, has been shown to influence reading of derived words in Finnish. The more allomorphs the suffix has, the slower its recognition proceeds (Järvikivi, Bertram & Niemi, 2006). The inflectional system of Dutch is much simpler than that of Finnish, and for nouns and adjectives considered here the main inflectional categories are Number (for nouns, e.g., sg. *werk-er* "worker" - pl. *werk-er-s* "workers") and Gender (for adjectives, common *bereikbare* - neuter *bereikbaar*). The most common change in the spelling of affixes across inflectional forms is fully determined by regular spelling conventions for representing short and long vowels in different syllable types (cf. doubling of consonants in common gender *succesvolle* vs. neuter *succesvol*, or a vowel loss in the example with *bereikbare* above). The only two suffixes that are structurally variable are *-heid* (pl. *-heden*) and *-loos* (pl. *-lozen*). Since words with these suffixes do not come with slower reading times, we conclude that inflectional paradigms and the number of structurally variant suffixes are too small to elicit the effect observed in Finnish.

General Discussion

The present data give rise to two insights into the lexical processing of morphologically complex words. First, access to the full-form (diagnosed by the derived word frequency effect) is modulated by the salience of the word's suffix, i.e., suffix length in phonemes (see Figure 5.1). Words with longer affixes show a weaker effect of word frequency than those with shorter suffixes. This finding is not easy to reconcile with any model that requires an obligatory temporary order in accessing the full-forms and morphemes of complex words. The single full-form route postulated by supralexical models suggests that readers directly access the central representation of the whole word, so that the subsequent, post-access activation of morphemes cannot modulate the full unfolding of the derived word frequency effect. On the other hand, sublexical models require pre-access obligatory decomposition of the complex words into morphemes, so that morphemic properties, such as suffix length, can only modulate the ease of access to those morphemes and are not predicted to interfere with the post-access full-form activation. Importantly, the observed interaction of derived word frequency and suffix length follows quite naturally from the premises of dual- or multiple-route models of parallel processing. The longer (hence, more salient) suffix shifts the balance between storage and computation towards computation and makes the full-form route less beneficial for the lexical processor than decomposition. The stronger the bias towards decomposition is, the less use readers make of the properties of the full-form properties in the recognition of derived words (cf., Bertram & Hyönä, 2003).

Second, we observe an interaction between morphological families of the base and suffix of the derived word. Figure 5.2 suggests that the time spent fixating derived words is minimal when the two families are of a similar size. Intuitively, this implies that processing is optimal when both morphological families are similarly strong as cues for complex word identification, or when both are similarly weak cues. Possibly, activation of morphological families gives rise to two subprocesses involved, recognition of one of family members (the constituent actually realized in the complex word) and inhibition of competing members of the family. The first subprocess ties in with the experience of segmenting constituents out of the embedding word and the larger the families the easier such segmentation proceeds. Yet there are fewer competitors to inhibit when families are small, so the inhibition may come with less processing costs when the families are small. We speculate that the optimal processing coincides with either the easiest segmentation task (when both families are large) or with the easiest inhibition task (when both families are small). At any rate, this interaction demonstrates that the balance in using storage and computation as ways of lexical processing is more intricate and interactive than the one hypothesized in single-route processing models. Moreover, it is problematic for computational dual-route models of parallel morphological processing like MATCHEK (Baayen & Schreuder, 2000). The MATCHEK model predicts greater competition if representations of both constituents have the same strength, so these conditions should come with slower processing times, contrary to fact. Evidently, a large spectrum of models cannot fully account for the present results. In what follows we discuss our findings, including the two interactions, in the framework of PROMISE, a probabilistic model for the processing of morphologically complex words (Chapter 4 of this dissertation).

The conceptual background of PROMISE is the view of the mental lexicon as a long-term memory storage for lexical information. The visual or auditory uptake of a stimulus triggers access of this lexical information. The ease of access, and generally of lexical processing, depends in part on the amounts of information carried by words, which are defined by the accumulated knowledge of words and their paradigmatic and syntagmatic connectivity in the mental lexicon. To quantify the amount of lexical information, PROMISE uses tools of information theory, including its statement that the amount of information (I) of a (linguistic) unit can be defined as a negative binary log probability (p) of that unit in its context:

$$I = -\log_2 p. \tag{5.1}$$

As p decreases, I increases: less probable units carry more information for word identification. PROMISE postulates that the time t spent by the eye on a morpheme or word is proportional to the total amount of lexical information available in long-term memory for identification of that morpheme or word at that timepoint, i.e., $t \propto I$ (cf., Levy, 2008; Moscoso del Prado Martín *et al.*, 2004). Units with small probability and hence a large information load require more processing resources and more processing time.

The interpretation of the observed main effect of derived word frequency F_d in the framework of PROMISE is straightforward. The probability of a derived word can be approximated as the relative frequency of that word in the corpus, $p_d = \frac{F_d}{N}$, where N is corpus size. The amount of information I_d carried by derived word frequency is

$$I_d = -\log_2 p_d = -\log_2 F_d + \log_2 N. \tag{5.2}$$

The corresponding processing time t is then proportional to the amount of information:

$$t \propto I_d = w_1(-\log_2 F_d + \log_2 N), \tag{5.3}$$

where w_1 is a positive weight coefficient. I_d is negatively correlated with log derived word frequency, so when that frequency is high, the amount of information in the system is small and the word is processed relatively fast. Also, N is the token size of the corpus and, as a corpus constant, it codetermines the average processing time for a word in the corpus, i.e., it modulates the intercept of the model equation. Crucially, starting from a simple probability and using information theory, we have derived a model equation the parameters of which can be directly estimated from the data using multiple (linear) regression models. For instance, based on the beta coefficient for $WordFreq$ (log frequency of the derived word) in Table 5.2, we obtain the estimate for $w_1 = -0.02$. What PROMISE does is develop model equation (5.3) further to incorporate multiple information sources that can be linked to the coefficients of regression models for the processing of complex words. For instance, one way of modeling the interaction between derived word frequency and suffix length in the PROMISE framework is through conditioning the probability estimates for the derived word on subsets of words that have suffixes with different lengths. Since the focus of the present study is on the effects of morphological structure, we leave the implementation of this interaction to future research.

The other interaction that is the focus of our interest here involves morphological families of both constituents, the family of the base (e.g., *happi-less, happi-ly, happi-lessness*) and that of the suffix (or suffix productivity, e.g., *happi-ly, marri-ly, sad-ly*). The two families interact such that words with similarly small or similarly large families are processed fastest, while the words in which one of the families prevails in size come with increased processing costs (see Figure 5.2). The PROMISE model in its current form introduces families of morphological constituents into the model equation by estimating conditional probabilities of one such constituent given that the other one has been identified. That is, the full model equation for the processing of compound words in Chapter 4 includes weighted probabilistic estimates for the left constituent family and the right constituent family, just like (5.3) incorporates the weighted estimate of the whole word probability. Inclusion of the family-based estimates allows PROMISE to account for the main effects of the constituent family sizes reported for compound processing (e.g., Chapters 3 and 4). Crucially, however, PROMISE currently does not predict the multiplicative interaction between constituent families, such as we observe for derived words here, nor is there a conditional probability that would straightforwardly lead to such an interaction.

The theoretical and modeling framework developed for PROMISE forces us to identify the real cause of this multiplicative interaction. This interaction appears to reflect a conflict between base productivity (i.e., family size) and suffix productivity. If the two constituents diverge in their productivity, they also diverge in the amount of information they carry. This imbalance in the informativeness of morphological constituents may slow down the parsing route and result in inflated reading times. Information theory offers a measure, relative entropy (RE), that allows quantifying this imbalance. Relative entropy is a measure of how much information is gained from the probability distribution P when the probability distribution Q is taken as the reference, or how well Q approximates P (for an overview of applications of RE in morphological processing see Chapter 6 of this dissertation: Milin, Kuperman, Kostić & Baayen, 2008; for syntactic processing see Levy, 2008). Below we shall define the P and Q distributions for the two morphological families. In what follows, we first explain in what way RE as a predictor differs from the multiplicative interaction between the two families in our regression model.

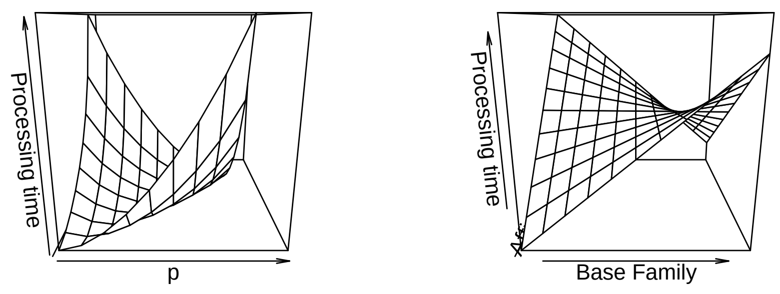
The formal definition of RE is

$$RE(P||Q) = \sum p \log_2 \frac{p}{q}. \tag{5.4}$$

Relative entropy equals zero when the two probability distributions P and Q are identical, and it increases if those distributions diverge. If our interpretation of the interaction of base family size by suffix productivity is correct, we expect to observe a positive correlation of relative entropy with reading times. We initially test this intuition by simulating the effect of relative entropy on the total amount of information in the processing system across the full ranges of $p \in P$ and $q \in Q$, see Figure 5.3a. We independently vary probabilities p and q in the interval [0; 1], and we compute the information gain estimated by the measure of relative entropy defined in (5.4). By implication, the resulting information gain would be proportional to the processing time. For the sake of qualitative comparison, we also show in Figure 5.3b the multiplicative linear interaction between base family size and suffix family size, using the empirical ranges of values (see Table 5.1), as well as the regression coefficients for the respective main effects and the interaction in the model for single-fixation duration (see Table 5.2). Note that the interaction plot in Figure 5.2 is the two-dimensional representation of the perspective plot in Figure 5.3b.

There is one important difference between the two subplots in Figure 5.3. The entropy-based simulation predicts that as long as the base and the suffix family-based probabilities are similar, processing will be maximally fast, while the multiplicative interaction of two family sizes shows a less clear indication of what constitutes the optimal processing conditions. For medium-sized equal families, costs are higher than for small- and large-size families of the equal size. We conclude from this simulation that our intuition about the conflicting behavior of the two morphological families is better captured by the measure of relative entropy than by the multiplicative interaction.

Figure 5.3: a. Simulation: the effect of relative entropy on processing times. b. Interaction plot: processing time as the function of the interaction of suffix family size by base family size.



a. Relative Entropy

b. Interaction of constituent families

We therefore proceed to model the interaction between morphological families, *BaseFamilySize* and *ResidAffixProd*, with the help of the relative entropy measure. First, we have to define the probability distributions for suffix productivity P and base family size Q (we discuss below the reasons for selecting base family size as the reference distribution). The probability that a derived word type carries a specific suffix (i.e., belongs to the suffix family) can be estimated as the number of types in which the suffix occurs, divided by the total number of types in CELEX (roughly, 40000). We exponentially transform the log counts[4] to get an estimate of the number of types in which a given suffix occurs (e.g., 233 for the Dutch suffix *-ing* as in *plaatsing* "placing"). The (simplest) probability distribution P for this suffix is a tuple $(p_1; p_2)$, which includes the probability of that suffix $p_1 = \frac{233}{40000}$ and the probability of its complement $p_2 = 1 - p_1$, that is, here $P = (0.006; 0.994)$.

Likewise, we can estimate the probability of the base family q_1 as the ratio of the number of members in that family (e.g., 121 for the base *plaats* "place" in *plaatsing* "placing") and the total number of types, 40000. The reference probability distribution Q is again a tuple $(q_1; q_2)$ of the probability q_1 and the probability of its complement, $q_2 = 1 - q_1$, that is, $Q = (0.003; 0.997)$. The relative entropy RE for the base and suffix families of *plaatsing* is now estimated from (5.4) as follows:

$$\begin{aligned} RE(P||Q) &= 0.006 * (\log_2 \frac{0.006}{0.003}) + 0.994 * (\log_2 \frac{0.994}{0.997}) \\ &= 0.006 - 0.004 = 0.002. \end{aligned} \quad (5.5)$$

Using the procedure described above, we compute RE for each target word in our dataset. We further multiply the values of relative entropy by 100 to bring them to a similar scale as other predictors in the model.

As the next step, we include RE as a predictor in our statistical model for single-fixation duration. We observe that RE is indeed a highly significant predictor of the reading time ($\hat{\beta} = 0.014, p_{mcmc} = 0.005$), and its regression coefficient is positive, as we anticipated. Moreover, the interaction of base family size by (residualized) suffix productivity loses its significance in the new model, and so do the main effects of base family size and suffix productivity (all ps > 0.1). This suggests that the relative entropy measure absorbs the variance in the data previously explained by the other predictors. Furthermore, RE retains significance when the non-significant interaction and the main effects are removed from the model. Importantly, the resulting model with RE as a predictor has a better

[4]Our interaction of interest includes as one of its terms the residualized (rather than raw) suffix productivity counts, as we de-correlated suffix productivity from the highly collinear measure of suffix frequency. We scale the residualized values of productivity back to log counts by multiplying those values by $\frac{a*b}{c}$, where a is the range of residualized counts, b is the raw productivity count (the number of word types occurring with the suffix), and c is the range of raw counts.

(greater) value of log-likelihood (which is a measure of the model's fit to the data), -883.2, than that for the original model in Table 5.2, -890.5, whereas it uses less parameters. The effect size of RE in that model is 16.8 ms, with a 1.6 ms increase in single-fixation duration for one bit of information. Similarly, relative entropy outperforms the interaction between two family sizes in the model for gaze duration, leading to a model with the greater value of log-likelihood and less parameters, in which RE shows a 9 ms effect. We conclude that RE allows us to fit superior models to the data.

Since relative entropy is measured in bits, we can directly incorporate it in PROMISE as a reflection of the trade-off between relative strengths of information sources:

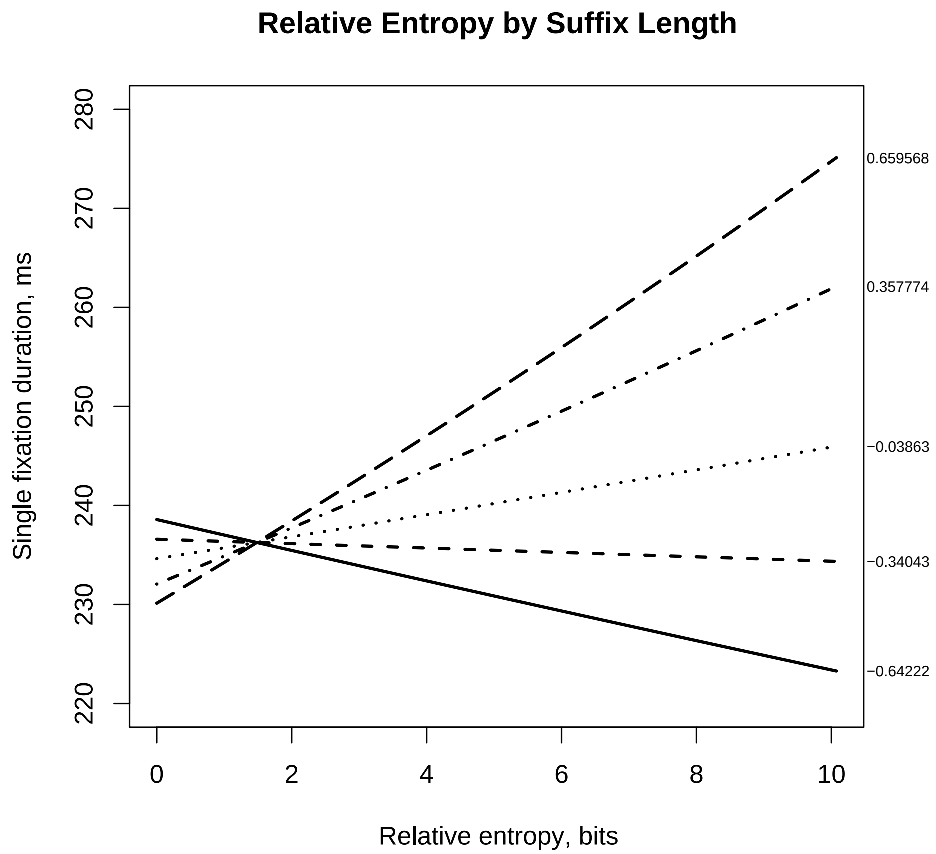
$$t = w_1 I_d + w_2 RE = w_1(-logF_d + logN) + w_2 RE. \tag{5.6}$$

The inhibitory effect of RE shows that a mismatch in the family size of base and suffix is detrimental to the speed of lexical processing. At present, we can only speculate about why this might be so. Possibly, the RE effect bears witness to a competition between morphological families, with the resonance between the family members of the morpheme with the larger family swamping the resonance of the family of the morpheme with the smaller family. (For a computational model implementing family size effects as the result of resonance in the mental lexicon, see De Jong, Schreuder & Baayen, 2003). This imbalance in the amount of lexical support would then delay the integration of the two morphemes into a coherent representation of the derived word as a whole.

We further hypothesize that the conflict between morphemic properties (estimated by the relative entropy measure) may codetermine lexical processing only to the extent that morphemes are accessed during the recognition of complex words. It is plausible that words in which morphological decomposition is more likely (e.g., due to a higher affixal salience) would show a stronger effect of relative entropy than those words which are preferrably accessed via their full-forms. Indeed, the predictor RE enters into a significant interaction with (residualized) suffix length in phonemes ($p_{mcmc} = 0.003$ for single-fixation durations and $p_{mcmc} = 0.010$ for gaze durations)[5]. The interaction is such that RE elicits longer single fixations and gazes in the words with longer (hence, more salient) suffixes, see Figure 5.4. That the effect of RE depends on the likelihood of parsing supports our

[5] Suffix length in characters also interacts significantly with the relative entropy measure, but we report the phoneme-based measure of suffix length to keep our list of predictors consistent within models.

Figure 5.4: Interaction of relative entropy by suffix length for single-fixation durations. The lines plot the effect of relative entropy for the quantiles of suffix length in phonemes (quantile values provided at the right margin). Relative entropy comes with the weak negative effect in words with short suffixes at the 1st quantile (solid line), the effect gradually increases at the 2nd quantile (dashed line), the 3d quantile (dotted line) and the 4th quantile (dotdash line), and it reaches its maximum in words with very long suffixes, the 5th quantile (longdash line).



idea that RE is a measure of parsing complexity. Again, we leave to future research the issue of how orthographic information, such as the suffix length, modulates the weight coefficients in the PROMISE model (e.g., w_2 in 5.6).

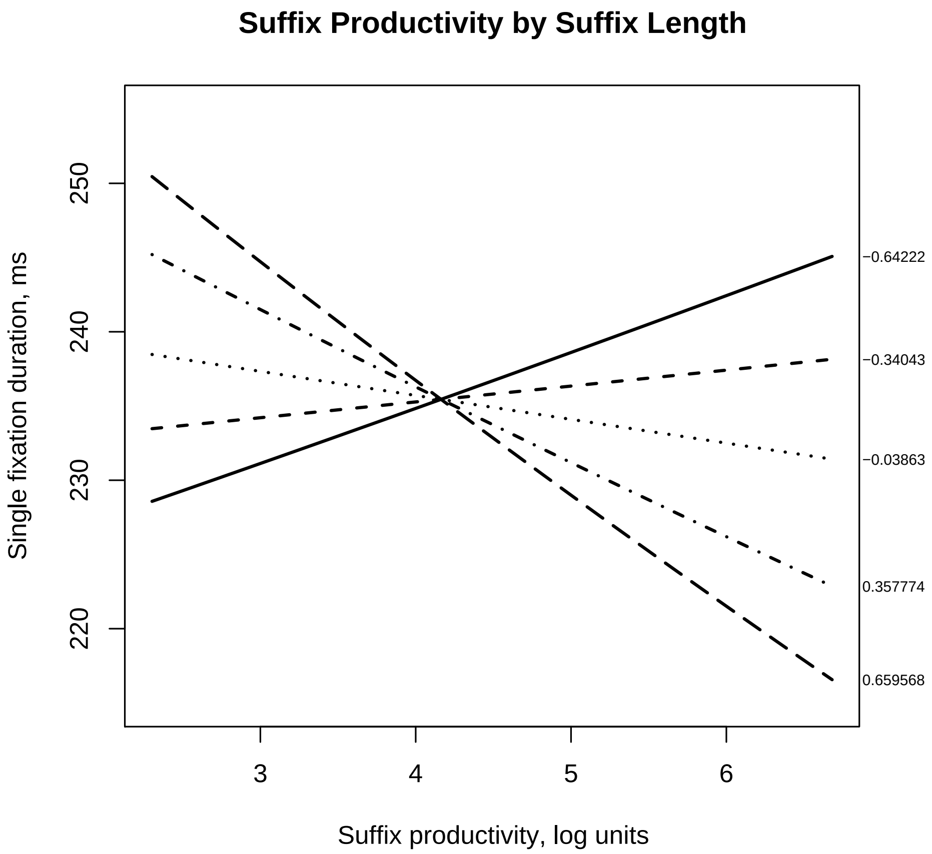
Importantly, relative entropy is sensitive to which of the probability distributions is selected as the reference distribution, that is, $RE(P||Q) \neq RE(Q||P)$. In the analysis outlined above we opted for measuring how much information is gained from the probability distribution of the suffix family given that the reference probability distribution associated with the base family is already known. One reason for this decision was that research on long complex words showed that the order of activation of morphological constituents ties in with the typical left-to-right sequence of visual uptake (cf. Hyönä, Bertram & Pollatsek, 2004). Most of the present stimuli were read in a single fixation, so both morphemes were likely to become available to the visual system simultaneously during this first fixation. However, one might argue on the basis of earlier findings that the typical direction of reading of morphemes in long words and of words in sentences leads to increased attention to word-initial morphemes (and more generally, characters) of complex words rather than to word-final ones, cf. Inhoff (1989a, 1989b). Second, the landing position of the single fixation in the word (mean = 3.9 characters, SD = 2.5) and the constraints of visual acuity in reading (e.g., Rayner, 1998) suggest that the view of the base is more accurate than that for the suffix in most target words in the length range of 8-12 characters. The better quality of the visual information about the base may also translate into a certain advantage that the base has over the suffix in the recognition of the derived word, so that typically the information provided by the suffix is processed against the backdrop of the (partly) available information about the base and its morphological family. We also examined the alternative measure of relative entropy, the one in which the probability distribution of the suffix family serves as the reference distribution. This predictor did not reach significance in any of our models as a main effect, nor in the interaction with suffix length.

Our hypothesis of the temporal relationship of constituents in complex word recognition offers a straightforward, testable prediction for the case of prefixed words (e.g., *re-build*). We expect that a conflict between productivities of the prefix and the base word will also be reflected in the eye-movements record (e.g., in longer processing times) and that it can be modelled as an effect of the relative entropy computed over probability distributions of respective morphological families. Crucially, we expect that taking the probability distribution of the prefix family (and not that of the base family, as in the present study) as the reference distribution will lead to a better approximation of the observed data on reading of prefixed words.

Apart of the interaction of relative entropy by suffix length, the statistical models for single-fixation durations and gaze durations reveal another interaction of interest, that of suffix productivity by (residualized) suffix length in phonemes ($p_{mcmc} = 0.013$ for single-fixation durations and $p_{mcmc} = 0.044$ for gaze durations), see Figure 5.5.

Suffix productivity comes with shorter single-fixation durations and gaze durations in words with larger suffixes, while the effect becomes weaker and

Figure 5.5: Interaction of suffix productivity by suffix length for single-fixation durations. The lines plot the effect of suffix productivity for the quantiles of suffix length in phonemes (quantile values provided at the right margin). Suffix productivity comes with the weak positive effect in words with short suffixes at the 1st quantile (solid line), the effect gradually decreases at the 2nd quantile (dashed line), the 3d quantile (dotted line) and the 4th quantile (dotdash line), and it reaches its minimum in words with very long suffixes, the 5th quantile (longdash line).



positive in the words with shorter suffixes[6]. We checked whether the positive effect of suffix family size was significant for words with short suffixes by fitting a statistical model to the subset of words with median or shorter suffix length: There was no significant effect ($p_{mcmc} > 0.1$). We conclude that the positive effect for words with very short suffixes is an artefact of the statistical model, which overestimates the influence of these extreme data points on the dependent variable. Simply put, the interaction suggests that the longer the suffix is, the better it is parsed, and suffix productivity elicits more facilitation of processing. Tables 5.3 and 5.5 present the final statistical models for single-fixation durations and gaze durations after removal of outliers, with the interactions of relative entropy by suffix length and suffix productivity by suffix length as predictors.

We introduce the suffix family size into the PROMISE model by considering the conditional probability of the right constituent (suffix), μ_2, conditioned on the presence of some unspecified left constituent (base), M_1. The unspecified left constituent stands for the subset of all morphemes or words in a language that can appear word-initially. This subset is then equal to the full vocabulary, except for suffixes (e.g., *-ness, -ity*) and those compounds' constituents that can only occur word-finally. Suppose that the reader has an intuition that the word under inspection, say *blackberry*, as potentially morphologically complex (based, for example, on its length or the low probability of the bigram "kb"). While the left constituent of such a compound is unspecified, combinations like **nessberry* or **ityberry* will never be part of the lexical space which needs to be considered for identification of the full compound. This probability becomes relevant in situations when the base is not fully processed and the likelihood of the suffix is nevertheless evaluated, for instance, because eyes landed on that suffix or because the suffix is perceptually salient.

We denote this conditional probability as $\Pr(\mu_2|M_1)$, and estimate it using the Bayes' theorem as follows:

$$\Pr(\mu_2|M_1) \quad = \quad \frac{\Pr(M_1,\mu_2)}{\Pr(M_1)} = \frac{\Pr(\mu_{+2})}{\Pr(M_1)} = \frac{F_{+2}}{F_{M_1}}. \qquad (5.7)$$

In these equation, F_{M_1} denotes the summed frequencies of all words that can occur as a left constituent in a complex word, μ_{+2} stands for the set of words ending in the right constituent μ_2 (suffix family), and F_{+2} denotes the summed frequency of all words ending in the right constituent μ_2 (suffix frequency[7]). The probability

[6]The interaction of suffix family frequency by suffix length is also significant, but for internal consistency we report the interaction with suffix productivity as a term.

[7]At present the model estimates the corresponding probabilities and informations using the log of

$\Pr(M_2)$ is independent of μ_2 and hence is a constant in our model. The amount of information carried by the probability in (5.7) is

$$I_{\mu_2|M_1} = \log F_{M_1} - \log F_{+2}. \quad (5.8)$$

We can now introduce this amount of information into our model equation:

$$\begin{aligned} t &= w_1 I_d + w_2 RE + w_3 I_{\mu_2|M_1} \quad (5.9) \\ &= w_1(-logF_d + logN) + w_2 RE + w_3 logM_1 - w_3 logF_{+2}. \end{aligned}$$

Since all weight coefficients w are positive, equation (5.9) predicts the negative correlations of both derived word frequency and suffix family size with reading times, as well as the positive correlation of the relative entropy measure with the reading times.

To sum up, using the distributional properties of the full-form, such as derived word frequency, can make the process of complex word recognition easier (faster) for the reader, and so can using the distributional properties of the word's morpheme, such as suffix productivity. Relative entropy represents a difficulty that the lexical parser encounters under the imbalance in the informativeness of the word's morphemes. The relative entropy effect emerges side by side but independent of the effects of derived frequency (as a measure of how easily the full-form can be retrieved from the mental lexicon) or suffix family size (as a measure of support from the family). Hence, we interpret relative entropy as a separate dimension of processing complexity that emerges during decomposition of words into morphemes, and yet is distinct and not reducible to either parsing or full-form lexical access.

Crucially, all the effects of morphological structure that we observe in the present study are significantly qualified by the measure of suffix length, and hence, by affixal salience, see Figures 5.1, 5.4 and 5.5. That is, suffix length appears to serve as a key parameter that regulates the allocation of cognitive resources over available processing routes. In particular, it fine-tunes the share of storage and computation in the processing of derived words. Words with extremely short and hence non-salient suffixes do not provide a clear pointer to the reader that a complex word is at stake. This makes the full-form processing a preferred recognition route, which may be why the negative effect of derived word frequency

the summed frequency of these families, see footnote 6. It may be more appropriate to alternatively estimate the amount of information in the morphological family using Shannon's entropy (cf. e.g., Moscoso del Prado Martín, Kostić & Baayen, 2004), or by the log the family size, which is the measure we used in our statistical models.

on reading times is at its strongest in such words, while the effects of relative entropy and suffix productivity are only weak. As affixal salience increases in words with longer suffixes, the effect of derived word frequency is attenuated and virtually vanishes, while the effects pertaining to the parsing of derived word's morphemes (negative for suffix productivity and positive for relative entropy) increase in size.

To illustrate the simultaneous modulation of morphological effects by affixal salience in terms of reading times, we consider two hypothetical words with extreme distributional properties. Suppose word A has the highest frequency, the largest suffix family and the largest value of the relative entropy measure among all words in our experimental list. Our word B is the opposite of word A, with the lowest word frequency, the smallest suffix family size and an identical base family size (i.e., zero relative entropy). We further assume that words A and B carry suffixes of the same length. If both A and B have the shortest suffixes (represented by the solid black lines in Figures 5.1, 5.4 and 5.5), our model for single-fixation duration makes the following predictions. Word A will come with a 55 ms reduction of reading time due to higher word frequency, a 10 ms reduction due to lower relative entropy and a 15 ms inflation of the reading time due to higher suffix productivity, as compared to word B. In total, word A would be processed some 50 ms faster than word B, given that suffixes in A and B are short and hence full-form access is likely. If A and B are words with median suffix length (represented by the dotted lines in Figures 5.1, 5.4 and 5.5), a single fixation on word A is predicted to be only 30 ms shorter due to its higher word frequency. Also the reading time for word A increases by 10 ms due to relative entropy and reduces again by some 5 ms due to higher suffix productivity. In total, the processing advantage for word A over word B is only 25 ms, if the word's suffix is in the mid-range of length. Finally, if words A and B both carry the longest suffixes (longdash lines in Figures 5.1, 5.4 and 5.5), word A has no advantage due to its higher word frequency, its higher relative entropy comes with the cost of 45 ms, but the reading time for word A is further reduced by 35 ms due to its larger suffix family. In total, the processing time for word A is predicted to be 10 ms longer than that for word B, given extremely long suffixes and hence preferred morphological decomposition.

Naturally, most words are not as extreme in their distributional properties as the words in our example and the benefits and costs of morphological processing may not be as drastic. Still, this example highlights two points with respect to the dynamics of lexical processing for complex words. First, the effect of any single information source (e.g., suffix family size) on the speed of word recognition can range from facilitatory to negligibly small to detrimental depending on the effects of other such sources (e.g., suffix length) and the likelihoods of available processing routes. There are several ways in which PROMISE implements these trade-offs in processing. For instance, information sources introduced into PROMISE come with two terms carrying weight coefficients with opposite signs (see w_1 in equation (3) for the probability of the derived word), so that the increase of one equation term comes with the decrease of the other. Methodologically, this implies that considering any one information source in isolation from others (by, say, keeping the values of other information sources constant through matching of stimuli) is bound to miss the essentially interactive use of bits and pieces of morphological structure in complex word recognition.

Second, in our comparison between hypothetical words A and B we followed the assumption of PROMISE that the total costs of processing are codetermined by the linear sum of costs associated with specific information sources, see (5.9). Moreover, we showed that even strong morphological effects can be cancelled out or outweighted by the sums of other effects. This may raise a general question of whether using morphological sources of information is a viable alternative to the recognition of words as unstructured units, in line with the full-listing hypothesis of Butterworth (1983). The fact that we, along with the long tradition of morphological research, observe readers making use of morphemic properties does not necessarily imply that on average readers benefit from such use in terms of the processing speed. Morphological cues may merely impose themselves on the recognition system and be followed automatically, even to the disadvantage of the reader. However, there is a growing evidence from word comprehension studies that on average complex words are processed faster than simplex words with similar values of frequency, length and several other characteristics. Thus, Bertram *et al.* (1999) observed that Finnish derived words elicited shorter visual lexical decision latencies than monomorphemic words, Fiorentino and Poeppel (2007) replicated this finding comparing English compounds and simplex words, while Balling and Baayen (2008) reported the processing advantage for Danish derived and inflected words in the auditory lexical decision task. That is, the fine-tuned balance between multiple processing routes, modeled in PROMISE, may impose conditions on what counts as the optimal processing strategy and what the costs of suboptimality are, but it also offers an overall processing advantage unavailable to simpler recognition systems.

Appendix

Key to Table 5.1: This table provides statistics on continuous variables, which show significant effects in our statistical models, including ranges of their original values, and (where applicable) ranges of the values after (logarithmic or orthogonalization) transformations. Column Variable lists predictors of interest. Numbers in the second column show original value ranges for predictors. If any transformations have been applied to the original values for statistical reasons (i.e., natural log transformation, decorrelation with other predictors or scaling), the numbers in the brackets show the ranges actually used in statistical models. Means, standard deviations (Column 3) and median values (Column 4) refer to the predictor values used in the models. Computation of these distributional measures was based on the combined pool of roughly 120 million tokens, obtained from the CELEX lexical database and from the newspapers in the Twente News Corpus.

Table 5.1: Summary of Continuous Variables

Variable	Range (units)	Mean(SD)	Median
InitFixPos	0.0:13.68 characters	3.56(2.4)	3.81
SingleDur	74:899 ms (4.3:6.8 log units)	5.5(0.3)	5.5
GazeDur	74:812 ms (4.3:6.7 log units)	5.6(0.4)	5.5
WordLength	8:14 characters	9.1(1.3)	9
ResidSuffixLenPhoneme	2:5 phonemes (-0.6:0.7)	0.0(0.4)	0.06
WordFreq	12:2607 (2.5:7.9 log units)	4.6(1.2)	4.6
SuffixFreq	1813:1.9$10^5$ (7.5:12.2 log units)	9.8(1.5)	9.3
ResidSuffixProd	10:812 (-1.7:1.1)	0.02(0.7)	0.03
RE	0.0:10.1 bits	1.9(2.4)	0.7
BaseFreq	2:59874 (0.7:11.1 log units)	6.9(1.9)	6.9
BaseFamilySize	2:300 (0.7:5.7 log units)	2.9(1.2)	2.8
PrecLength	2:17 characters	4.1(2.5)	3

In addition to the variables reported in Table 5.1, we considered a large number of control variables that were not significant predictors of reading times or probabilities. These included such distributional predictors as complexity-based ranking of suffixes; number of word types in which the ratio of base frequency and whole word frequency is above/below the mean ratio for the affix, and their ratio; number of hapax legomena in which the suffix occurs; growth rate and type/token ratio for the affix; cumulative base frequency, the relative frequency of the base, average relative base frequency, and the ratio of word types in which whole word frequency exceeds base frequency and word types in which whole word frequency is lower than base frequency. Orthographic and phonological factors included: whether or not the first or the last syllable of the word was stressed; whether stress falls on any of the suffix's syllables; whether or not a suffix began with a vowel; as well as type-based frequencies of the word-initial and word-final trigram and frequency of occurrence of the bigrams straddling, preceding and following the morphemic boundary. Lexical predictors included: whether or not affixes change their orthographic form across the inflectional paradigm; whether the word class of the derivation as a whole differs from the word class of the base word; and whether affixes in target words were homonymous with Dutch inflectional affixes; and word class of the target word. We also considered the number of word types in which the character string occurs as a suffix and the number of word types in which it occurs in the word-final position in any other non-morphemic capacity. Contextual control variables included: joint probabilities of words N-1 and N, and of words N and N+1; lengths, frequencies and word classes of words N-1 and N+1; and word position in a sentence. Visual control variables included: amplitudes of saccades preceding and following the fixation. Also, including affix as a random effect did not significantly improve the statistical models.

Key to Tables 5.2-5.5: The first column lists the intercept as well as all predictors and interactions that reached significance in the model. The second column shows estimates of the regression coefficients for the model's predictors. Columns 3-6 provide information on the distributions of those estimates obtained via the Monte Carlo Markov chain (MCMC) random-walk method using 10000 simulations: this information is useful for evaluating stability of the model's predictions. The third column shows the MCMC estimate of the mean for each predictor, while the fourth and the fifth columns show highest posterior density intervals, which are a Bayesian measure for the lower and upper bounds of the 95% confidence interval, respectively. The sixth column provides a p-value obtained with the help of MCMC simulations; and the final column provides less conservative p-values obtained with the t-test using the difference between the number of observations and the number of fixed effects as the upper bound for the degrees of freedom.

Table 5.2: Original Model for Single-Fixation Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.4894	5.4890	5.3845	5.5887	0.0001	0.0000
PrecLength	0.0081	0.0081	0.0044	0.0121	0.0001	0.0001
InitFixPos	0.0109	0.0107	-0.0014	0.0226	0.0772	0.0745
InitFixPos^2	-0.0021	-0.0021	-0.0033	-0.0009	0.0012	0.0008
WordFreq	-0.0204	-0.0203	-0.0293	-0.0118	0.0001	0.0000
ResidSuffixLenPhoneme	-0.1315	-0.1317	-0.2317	-0.0273	0.0122	0.0124
BaseFamilySize	-0.0051	-0.0050	-0.0136	0.0040	0.2704	0.2584
ResidAffixProd	0.0600	0.0600	0.0197	0.1024	0.0040	0.0041
WordFreq:ResidSuffixLenPhoneme	0.0332	0.0333	0.0112	0.0552	0.0030	0.0033
BaseFamilySize:ResidAffixProd	-0.0163	-0.0163	-0.0289	-0.0038	0.0130	0.0115

Table 5.3: Final Model for Single-Fixation Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.5655	5.5650	5.4875	5.6471	0.0001	0.0000
PrecLength	0.0077	0.0077	0.0040	0.0114	0.0001	0.0001
InitFixPos	0.0114	0.0115	-0.0003	0.0233	0.0604	0.0598
InitFixPos^2	-0.0022	-0.0022	-0.0034	-0.0009	0.0006	0.0004
WordFreq	-0.0238	-0.0238	-0.0317	-0.0153	0.0001	0.0000
ResidPhonemeAffix	-0.0438	-0.0466	-0.2012	0.1132	0.5600	0.5883
RE	0.0070	0.0069	0.0015	0.0126	0.0174	0.0171
AffixProd	-0.0083	-0.0081	-0.0200	0.0038	0.1750	0.1813
WordFreq:ResidSuffixLenPhoneme	0.0396	0.0399	0.0187	0.0620	0.0008	0.0005
ResidSuffixLenPhoneme:RE	0.0252	0.0253	0.0121	0.0388	0.0001	0.0002
ResidSuffixLenPhoneme:AffixProd	-0.0377	-0.0374	-0.0672	-0.0081	0.0156	0.0129

Table 5.4: Original Model for Gaze Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.3875	5.3877	5.2681	5.4983	0.0001	0.0000
PrecLength	0.0080	0.0080	0.0033	0.0127	0.0006	0.0010
WordLength	0.0520	0.0520	0.0426	0.0616	0.0001	0.0000
InitFixPos	-0.0656	-0.0656	-0.0767	-0.0536	0.0001	0.0000
InitFixPos^2	0.0028	0.0028	0.0016	0.0040	0.0001	0.0000
WordFreq	-0.0291	-0.0292	-0.0394	-0.0183	0.0001	0.0000
ResidSuffixLenPhoneme	-0.1325	-0.1327	-0.2555	-0.0091	0.0364	0.0379
BaseFamilySize	-0.0013	-0.0013	-0.0120	0.0092	0.8074	0.8110
ResidAffixProd	0.0500	0.0499	0.0015	0.0973	0.0416	0.0451
WordFreq:ResidSuffixLenPhoneme	0.0349	0.0350	0.0087	0.0621	0.0086	0.0111
BaseFamilySize:ResidAffixProd	-0.0143	-0.0142	-0.0293	0.0003	0.0596	0.0641

Table 5.5: Final Model for Gaze Duration

	Estimate	MCMCmean	HPD95lower	HPD95upper	pMCMC	Pr(>|t|)
(Intercept)	5.4537	5.4532	5.3052	5.6005	0.0001	0.0000
WordLength	0.0485	0.0485	0.0379	0.0587	0.0001	0.0000
PrecLength	0.0066	0.0066	0.0014	0.0120	0.0160	0.0172
InitFixPos	-0.0630	-0.0630	-0.0768	-0.0487	0.0001	0.0000
InitFixPos^2	0.0026	0.0026	0.0011	0.0041	0.0004	0.0007
WordFreq	-0.0311	-0.0311	-0.0425	-0.0191	0.0001	0.0000
ResidSuffixLenPhoneme	-0.0290	-0.0298	-0.2506	0.2010	0.7960	0.8022
RE	0.0021	0.0022	-0.0057	0.0104	0.6006	0.6102
AffixProd	-0.0060	-0.0060	-0.0236	0.0112	0.5054	0.5015
WordFreq:ResidSuffixLenPhoneme	0.0479	0.0477	0.0171	0.0805	0.0028	0.0032
ResidSuffixLenPhoneme:RE	0.0252	0.0250	0.0065	0.0451	0.0110	0.0099
ResidSuffixLenPhoneme:AffixProd	-0.0436	-0.0432	-0.0849	-0.0003	0.0484	0.0444

Table 5.6: Random effects of the final models for *SingleDur* and *GazeDur*

Single fixation duration				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.031	0.026	0.014	0.050
Subject	0.101	0.103	0.078	0.135
Residual	0.269	0.269	0.262	0.275
B. Gaze duration				
Estimate	St. Deviation	MCMCmean	HPD95lower	HPD95upper
Word	0.057	0.055	0.043	0.072
Subject	0.112	0.114	0.084	0.156
Residual	0.33			

References

Andrews, S. and Davis, C. (1999). Interactive activation accounts of morphological decomposition: Finding the trap in mousetrap? *Brain and Language*, 68:355–361.

Andrews, S., Miller, B., and Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, 16(1/2):285–311.

Baayen, R. H. (1993). On frequency, transparency, and productivity. In Booij, G. E. and van Marle, J., editors, *Yearbook of Morphology 1992*, pages 181–208. Kluwer Academic Publishers, Dordrecht.

Baayen, R. H. (1994). Productivity in language production. *Language and Cognitive Processes*, 9:447–469.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge University Press, Cambridge.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *In press*.

Baayen, R. H. and Schreuder, R. (1999). War and peace: morphemes and full forms in a non-interactive activation parallel dual route model. *Brain and Language*, 68:27–32.

Baayen, R. H. and Schreuder, R. (2000). Towards a psycholinguistic computational model for morphological parsing. *Philosophical Transactions of the Royal Society (Series A: Mathematical, Physical and Engineering Sciences)*, 358:1–13.

Balling, L. and Baayen, R. (in press). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*.

Bates, D. M. and Sarkar, D. (2007). The lme4 library. *[On-line], Available: http://lib.stat.cmu.edu/R/CRAN/*.

Bertram, R. and 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., Baayen, R. H., Schreuder, R., and Hyönä, J. (2000a). Affixal homonymy triggers full-form storage even with inflected words, even in a morphologically rich language. *Cognition*, 74:B13–B25.

Bertram, R., Laine, M., and 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., and 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., Schreuder, R., and 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.

Burzio, L. (1994). *Principles of English stress*. Cambridge Studies in Linguistics, Cambridge, UK.

Butterworth, B. (1983). Lexical representation. In Butterworth, B., editor, *Language production (Vol. II): Development, Writing and Other language Processes*, pages 257–294. Academic Press, London.

De Jong, N. H., Schreuder, R., and Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, 15:329–365.

De Jong, N. H., Schreuder, R., and Baayen, R. H. (2003). Morphological resonance in the mental lexicon. In Baayen, R. H. and Schreuder, R., editors, *Morphological structure in language processing*, pages 65–88. Mouton de Gruyter, Berlin.

Diependaele, K., Sandra, D., and Grainger, J. (2005). Masked cross-modal morphological priming: unravelling morpho-orthographic and morpho-semantic inluences in early world recognition. *Language and Cognitive Processes*, 20:75–114.

Fiorentino, R. and Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, 22:953–1000.

Giraudo, H. and Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review*, 8:127–131.

Gries, S. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement.* Continuum International Publishing Group Ltd., New York.

Hay, J. (2001). Lexical Frequency in Morphology: Is Everything Relative?

Linguistics, 39:1041–1070.

Hay, J. and Baayen, H. (2002). Parsing and productivity. In Booij, G. and Van Marle, J., editors, *Yearbook of Morphology 2001*, pages 203–235. Kluwer Academic Publishers, Dordrecht.

Hay, J. and Baayen, H. (2003). Phonotactics, parsing and productivity. *Italian Journal of Linguistics*, 1:99–130.

Hay, J. B. and Plag, I. (2004). What constrains possible suffix combinations? on the interaction of grammatical and processing restrictions in derivational morphology. *Natural language and linguistic theory*, 22:565–596.

Hyönä, J., Bertram, R., and Pollatsek, A. (2004). Are long compound words identified serially via their constituents? Evidence from an eye-movement-contingent display change study. *Memory and Cognition*, 32:523–532.

Inhoff, A. (1989a). Lexical access during eye fixations in reading: Are word access codes used to integrate lexical information across interword fixations? *Journal of Memory and Language*, 28:444–461.

Inhoff, A. (1989b). Parafoveal processing of words and saccade computation during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 15:544–555.

Järvikivi, J., Bertram, R., and Niemi, R. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and Cognitive Processes*, 21:394–431.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., and Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, 94:223–244.

Kuperman, V., Bertram, R., and Baayen, R. H. (2008a). Morphological dynamics in compound processing. *In press, Language and Cognitive Processes*.

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2008b). Reading of multimorphemic Dutch compounds: Towards a multiple route model of lexical processing. *Manuscript submitted for publication*.

Laudanna, A. and Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 345–364. Lawrence Erlbaum Associates, Hillsdale, N. J.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177.

Marle, J. v. (1985). *On the Paradigmatic Dimension of Morphological Creativity*. Foris, Dordrecht.

Milin, P., Kuperman, V., Kostić, A., and Baayen, R. (in press). Words and paradigms bit by bit: An information-theoretic approach to the processing of inflection and derivation. In Blevins, J. and Blevins, J., editors, *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94:1–18.

Niswander-Klement, E. and Pollatsek, A. (2006). The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Memory and Cognition*, 34:685–702.

Ordelman, R. (2002). Twente Nieuws Corpus (TwNC). Report 25, Parlevink Language Techonology Group. University of Twente.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Statistics and Computing. Springer, New York.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Weidenfeld and Nicolson, London.

Plag, I. (1999). *Morphological productivity: structural constraints in English*. Mouton de Gruyter, Berlin.

Pollatsek, A., Hyönä, J., and Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 26:820–833.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422.

Schreuder, R. and Baayen, R. H. (1995). Modeling morphological processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 131–154. Lawrence Erlbaum, Hillsdale, New Jersey.

Schreuder, R. and Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, 37:118–139.

Sereno, J. and Jongman, A. (1997). Processing of English inflectional morphology. *Memory and Cognition*, 25:425–437.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *The Quarterly Journal of Experimental Psychology*, 57A:745–765.

Taft, M. and Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14:638–647.

Vitu, F., Lancelin, D., and Marrier d'Unienville, V. (2007). A perceptual-economy account for the inverted-optimal viewing position effect. *Joural of Experimental Psychology: Human Perception and Performance*, 33:1220–1249.

Words and paradigms bit by bit: An information-theoretic approach to the processing of inflection and derivation

Chapter 6

This chapter is in press as a book chapter: Petar Milin, Victor Kuperman, Aleksandar Kostić & R. Harald Baayen. In: Blevins, James P. and Juliette Blevins (eds.), *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press.

Introduction

Syntagmatically oriented theories of word structure have inspired most of the experimental work on morphological processing. The way inflection is modeled by Levelt, Roelofs and Meyer (1999), for instance, comes close to the theory of distributed morphology proposed by Halle and Marantz (1993). In Levelt's model of speech production, nodes at the lemma stratum (what would be the lexeme stratum in the terminology of Aronoff (1994)) are marked for features such as tense, aspect, number and person. For a given set of feature values, a node at the form stratum will be activated, e.g., *-ed* for the past tense in English. Paradigmatic relations do not have a place in this model, in fact, it is a design feature of the model that paradigmatic relations at the level of word forms are predicted to be irrelevant.

A syntagmatic bias is also visible in the comprehension model proposed by Schreuder and Baayen (1995). In this model, there is no principled difference between stems or words on the one hand, and affixes (whether inflectional or derivational) on the other hand. In their three-layered network, with access units, lemma units, and semantic and syntactic feature units, the organization of nodes within a layer is arbitrary. Paradigmatic relations do not play a role, they are simply deemed to be irrelevant. The same holds for the dual mechanism model of Pinker (1991, 1999).

In this chapter, we present a line of research that departs from the syntagmatic orientation of mainstream experimental psycholinguistics, and that is close in spirit to Word and Paradigm (WPM) morphology (Anderson, 1992; Aronoff, 1994, Beard, 1995, Blevins, 2003, 2006; Hockett, 1954; Matthews, 1974). WPM questions the morphemic status of lexical formatives, and assumes that words (both simple and complex) are the basic units in the lexicon. Furthermore, in WPM, inflected words are organized into paradigms, which are further organized into inflectional classes[1].

From a processing perspective, the central tenets of WPM imply, first, that complex words, including regular inflected words, leave traces in long-term lexical memory, and second, that the processing of a given word is co-determined by paradigmatically related words.

A central diagnostic for the presence of memory traces in long-term memory has been the word frequency effect. A higher frequency of use allows for shorter processing latencies in both visual and auditory comprehension (cf., Baayen, Feldman & Schreuder, 2006; Baayen, McQueen, Dijkstra & Schreuder, 2003; New, Brysbaert, Segui *et al.*, 2004; etc.), and lower rates of speech errors in production (Stemberger & MacWhinney, 1986). The effect of word frequency tends to be stronger for irregular complex words than for regular complex words, and stronger for derived words than for inflected words. But even for regular inflected words, the effect of prior experience clearly emerges (Baayen, Wurm & Aycock, 2008), contrary to the claims of the dual mechanism model. The ubiquitous effect of word frequency shows that large numbers of complex words are indeed available in the (mental) lexicon, as claimed by WPM.

The focus of this chapter is on the second central processing consequence of WPM, namely, that paradigmatic organization should co-determine lexical processing. For derivational morphology, work on the morphological family size effect (see, e.g., Moscoso del Prado Martín, Bertram, Häikiö *et al.*, 2004) has clarified that processing of a given word is co-determined by other morphologically related words. This constitutes evidence for paradigmatic organization in the mental lexicon. However, morphological families are very heterogeneous, and do not readily allow words to be grouped into higher-order sets similar to inflectional classes. Therefore, the morphological family size effect provides at best

[1] In what follows, we will use the term *inflectional paradigm* to refer to the set of inflected variants of a given lexeme, and the term *inflectional class* to refer to a set of lexemes that use the same set of exponents in their inflectional paradigms.

circumstantial evidence for the central ideas of WPM.

In the remainder of this chapter, we first review a series of recent experimental studies which explore the role of paradigmatic structure specifically for inflected words. We then present new experimental results showing how the principles that structure inflectional paradigms can be generalized to subsets of derived words.

The approach to morphological organization and morphological processing that we describe in this chapter departs significantly from both theoretical morphology and mainstream of experimental psycholinguistics in that it applies central concepts from information theory to lexical processing. The greater the amount of information carried by an event (e.g., a word's inflected variant, an exponent, or an inflectional class), the smaller the probability of that event, and the greater the corresponding processing costs (see for a similar approach to syntax, Levy, 2008). We believe that information theory offers exactly the right tools for studying the processing consequences of paradigmatic relations. Furthermore, we do believe that the concepts of information science provide us with excellent tools to probe the *functional organization of the mental lexicon*, but we shall remain agnostic about how paradigmatic structures are implemented in the brain.

We begin this chapter with an introduction to a number of central concepts from information theory and illustrate how these concepts can be applied to the different levels of paradigmatic organization in the mental lexicon. We then focus on three key issues: (i) the processing cost of an exponent given its inflectional class, (ii) the processing cost associated with inflectional paradigms and inflectional classes, and (iii) the processing cost that arises when the probabilistic distributional properties of paradigms and classes diverge.

In what follows, we first provide a comprehensive review of previous experimental findings that use information-theoretic measures of lexical connectivity. We then present some new results that provide further empirical support for the relevance of paradigmatic organization for lexical processing, and for the importance of information-theoretic measures for gauging the processing consequences of paradigmatic structure. As we proceed through our discussion of the empirical evidence, it will become increasingly clear that there is a remarkable convergence between the psycholinguistic evidence and WPM.

Some of the key findings of the general approach to the (mental) lexicon outlined in this chapter can be summarized as follows:

1. Lexemes and their inflected variants are organized hierarchically. One can envision this organization as a higher layer of lexemes grouped into

morphological families, and a lower level of inflected variants, which enter into paradigmatic relations within a given lexeme.

2. Inflected variants of any given lexeme are organized into paradigms, and all lexemes that form their paradigms in the same way define an inflectional class. Empirical evidence suggests that the degree to which the inflectional paradigm of a given lexeme diverges from its inflectional class affects cognitive processing over and above other relevant factors: the greater the divergence, the more costly the processing.

3. Results which will be presented here for the first time show that the processing of English derivatives can be seen as analogical. During lexical processing, a given derivative is compared with its base word, and pitted against the generalized knowledge about the relationship between all derivatives of the same type and their corresponding base words.

4. The *family size effect*, which is known to be a semantic effect, probably represents the joint effect of both semantic similarity and morphological paradigmatic structure.

Central concepts from information theory

A fundamental insight of information theory is that the amount of information I carried by (linguistic) unit u can be defined as the negative binary logarithm of its probability:

$$I_u = -\log_2 \Pr(u). \tag{6.1}$$

Consider someone in the tip-of-the tongue state saying *the eh eh eh eh eh eh key*. The word *eh* has the greatest probability, $6/8$, and is least informative. Its amount of information is $-\log_2(6/8) = 0.415$ bits. The words *the* and *key* have a probability of $1/8$ and the amount of information they carry is 3 bits. In what follows, we assume that lexical units that have a higher information load are more costly to access in long-term memory. Hence, we expect processing costs to be proportional to the amount of information. This is exactly what the word frequency effect tells us: higher frequency words, which have lower information loads, are processed faster than low-frequency – high-information words.

We estimate probabilities from relative frequencies. By way of illustration, consider the inflected variants of the Serbian feminine noun “planina” (*mountain*).

Serbian nouns have six cases and two numbers. Due to syncretism, the twelve combinations of case and number are represented by only 6 distinct inflected variants. These inflected variants are listed in column 1 of the upper part of Table 6.1. The second column lists the frequencies of these inflected variants in a two-million word corpus of written Serbian.

We consider two complementary ways of estimating probabilities from frequencies. The probabilities listed in the third column of Table 6.1 are obtained by normalizing the frequency counts with respect to a lexeme's inflectional paradigm (column three). More specifically, the probability $\mathrm{Pr}_\pi(w_e)$[2] of an inflected variant w_e of lexeme w is estimated in this table as its form-specific frequency F (henceforth *word frequency*) of occurrence, normalized for the sum of the frequencies of all the distinct inflected variants of its lexeme, henceforth *stem frequency*:

$$\mathrm{Pr}_\pi(w_e) = \frac{F(w_e)}{\sum_e F(w_e)}. \tag{6.2}$$

The corresponding amounts of information, obtained by applying (6.1), are listed in column four. Table 6.1 also lists the frequencies of the six exponents (column 5), calculated by summing the word frequencies of all forms in the corpus with these exponents. The probabilities listed for these exponents (column six) are obtained by normalizing with respect to the summed frequencies of these exponents:

$$\mathrm{Pr}_\pi(e) = \frac{F(e)}{\sum_e F(w_e)}. \tag{6.3}$$

The corresponding amount of information is listed in column seven.

The second way in which we can estimate probabilities is by normalizing with respect to the number of tokens N in the corpus. The probability of a lexeme w is then estimated as the sum of the frequencies of its inflected variants, divided by N:

$$\mathrm{Pr}_N(w) = \frac{F(w)}{N} = \frac{\sum_e F(w_e)}{N}. \tag{6.4}$$

In this approach, the probability of an inflected variant can be construed as the joint probability of its lexeme w and its exponent:

$$\begin{aligned} \mathrm{Pr}_N(w_e) &= \mathrm{Pr}(w,e) \\ &= \mathrm{Pr}(e,w) \\ &= \frac{F(w_e)}{N}. \end{aligned} \tag{6.5}$$

[2]Here and in what follows we use Pr_π to denote probabilities defined with respect to paradigmatic sets.

feminine nouns						
Inflected variant	Inflected variant frequency	Inflected variant relative frequency	Information of inflected variant	Exponent frequency	Exponent relative frequency	Information of exponent
	$F(w_e)$	$\Pr_\pi(w_e)$	I_{w_e}	$F(e)$	$\Pr_\pi(e)$	I_e
planin-*a*	169	0.31	1.69	18715	0.26	1.94
planin-*u*	48	0.09	3.47	9918	0.14	2.84
planin-*e*	191	0.35	1.51	27803	0.39	1.36
planin-*i*	88	0.16	2.64	7072	0.1	3.32
planin-*om*	30	0.05	4.32	4265	0.06	4.06
planin-*ama*	26	0.05	4.32	4409	0.06	4.06
masculine nouns						
Inflected variant	Inflected variant frequency	Inflected variant relative frequency	Information of inflected variant	Exponent frequency	Exponent relative frequency	Information of exponent
	$F(w_e)$	$\Pr_\pi(w_e)$	I_{w_e}	$F(e)$	$\Pr_\pi(e)$	I_e
prostor-*ø*	153	0.38	1.40	25399	0.35	1.51
prostor-*a*	69	0.17	2.56	18523	0.26	1.94
prostor-*u*	67	0.17	2.56	8409	0.12	3.06
prostor-*om*	15	0.04	4.64	3688	0.05	4.32
prostor-*e*	48	0.12	3.06	5634	0.08	3.64
prostor-*i*	23	0.06	4.06	6772	0.09	3.47
prostor-*ima*	23	0.06	4.06	3169	0.04	4.64

Table 6.1: Inflected nouns in Serbian. The upper part of the table shows inflected variants for the feminine noun “planina” (*mountain*), the lower part shows the inflected variants of the masculine noun “prostor” (*space*). Columns present frequencies and relative frequencies of the respective inflectional paradigm and the class to which it belongs.

Likewise, the probability $Pr(e)$ of an exponent (e.g., *-a* for nominative singular and genitive plural in Serbian feminine nouns) can be quantified as the relative frequency of occurrence of e in the corpus:

$$\mathrm{Pr}_N(e) = \frac{F(e)}{N}. \tag{6.6}$$

The probabilities considered thus far are unconditional, a priori, decontextualized probabilities. As exponents appear in the context of stems, we need to consider the conditional probability of an exponent given its lexeme, $Pr(e|w)$. Using Bayes' theorem, we rewrite this probability as:

$$\begin{aligned}
\mathrm{Pr}_N(e|w) &= \frac{\mathrm{Pr}_N(e,w)}{\mathrm{Pr}_N(w)} \\
&= \frac{F(w_e)}{N}\frac{N}{F(w)} \\
&= \frac{F(w_e)}{F(w)} \\
&= \mathrm{Pr}_\pi(w_e).
\end{aligned} \tag{6.7}$$

Likewise, the conditional probability of the lemma given the exponent is defined as:

$$\begin{aligned}
\mathrm{Pr}_N(w|e) &= \frac{\mathrm{Pr}_N(w,e)}{\mathrm{Pr}_N(e)} \\
&= \frac{F(w_e)}{N}\frac{N}{F(e)} \\
&= \frac{F(w_e)}{F(e)}.
\end{aligned} \tag{6.8}$$

For each lexical probability we can compute the corresponding amount of information. We allow for the possibility that each source of information may have its own distinct effect on lexical processing by means of positive weights ω_{1-5}:

$$\begin{aligned}
\mathsf{I}_{w_e} &= -\omega_1 \log_2 F(w_e) + \omega_1 \log_2 N \\
\mathsf{I}_{w} &= -\omega_2 \log_2 F(w) + \omega_2 \log_2 N \\
\mathsf{I}_{e} &= -\omega_3 \log_2 F(e) + \omega_3 \log_2 N \\
\mathsf{I}_{e|w} &= -\omega_4 \log_2 F(w_e) + \omega_4 \log_2 F(w) \\
\mathsf{I}_{w|e} &= -\omega_5 \log_2 F(w_e) + \omega_5 \log_2 F(e).
\end{aligned} \tag{6.9}$$

We assume that the cost of retrieving lexical information from long-term memory is proportional to the amount of information retrieved. Hence the cost of processing an inflected word w_e is proportional to at least the amounts of information in (6.9).

More formally, we can express this processing cost (measured experimentally as a reaction time RT) as a linear function:

$$\begin{aligned} RT &\propto I_{w_e} + I_w + I_e + I_{e|w} + I_{w|e} \\ &= (\omega_1 + \omega_2 + \omega_3)\log_2 N - (\omega_1 + \omega_4 + \omega_5)\log_2 F(w_e) \\ &- (\omega_2 - \omega_4)\log_2 F(w) - (\omega_3 - \omega_5)\log_2 F(e). \end{aligned} \tag{6.10}$$

There are several predictions for the effects of lexical probabilities on lexical processing that follow directly from (6.10). First, word frequency $F(w_e)$ will always elicit a facilitatory effect, as all its coefficients have a negative sign in (6.10). Second, stem frequency $F(w)$ may either facilitate or inhibit processing, depending on the relative strengths of the coefficients ω_2 and ω_4. These two coefficients balance the importance of a word's probability as such (see the second equation in (6.9)), and its importance as the domain on which the probabilities of its inflectional variants are conditioned (see the fourth equation in (6.9)). Third, the frequency of the exponent can also either speed up or hinder processing depending on the values of ω_3 and ω_5. These two weights balance the importance of an exponent's probability as such (see the first equation in (6.9)) and the exponent as the domain on which the probability of inflected forms with that exponent are conditioned (see the fifth equation in (6.9)). The first two predictions are supported by the large-scale regression study reported by Baayen, Wurm and Aycock (2008) and also the study reported in Chapter 4 of this dissertation (Kuperman, Bertram & Baayen, 2008) .

We now proceed from basic lexical probabilities that operate at the level of individual inflected words to the quantification of the information carried by inflectional paradigms and inflectional classes. The paradigm of a given lexeme can be associated with a distribution of probabilities $\{\mathrm{Pr}_\pi(w_e)\}$. For "planina" in Table 6.1, this probability distribution is given in column three. The amount of information carried by its paradigm as a whole is given by the *entropy* of the paradigm's probability distribution:

$$H = -\sum_e \mathrm{Pr}_\pi(w_e)\log_2(\mathrm{Pr}_\pi(w_e)). \tag{6.11}$$

Formally, H is the expected (weighted average) amount of information in a paradigm. The entropy increases with the number of members of the paradigm. It also increases when the probabilities of the members are more similar. For a given number of members, the entropy is maximal when all probabilities are the same. H also represents the average number of binary decisions required to identify a member of the paradigm, i.e., to reduce all uncertainty about which member of the paradigm is at issue, provided that the paradigm is represented by an optimal binary coding. We illustrate the concept of optimal coding in Figure 6.1 using as an example the inflectional class of regular feminine nouns in Serbian.

The upper panel of Figure 6.1 shows an optimal binary coding scheme, in which the most probable exponent (*-e*, $\mathrm{Pr}_\pi = 0.39$) occupies the highest leaf node in the tree. The lower the probability of the other exponents, the lower in the tree they are located. Thus, the exponents with the lowest probabilities in the inflectional class, *-om* ($\mathrm{Pr}_\pi = 0.06$) and *-ama* ($\mathrm{Pr}_\pi = 0.06$) are found at the lowest leaf nodes. The second panel of Figure 6.1 represents another possible coding, which is suboptimal in that some exponents with relatively high probabilities are located below lower-probability exponents in the tree. Finally, the third panel shows the least optimal coding, in which the less probable the exponent is, the *higher* it is positioned in the tree. The average number of binary decisions (the number of bits) required to identify a given paradigm member, i.e., to reach the paradigm member's leaf node when starting at the root node of the tree, is the sum of the products of the number of steps and the members' probabilities. This average is never greater than the entropy of the paradigm $H+1$ (Ross, 1988). For the upper panel of Figure 6.1, the average number of binary decisions is 2.33 bits, for the coding in the second panel, it is 2.83, and for the worst coding in the third panel, it is 4.29. In section 6 we will review experimental studies showing that paradigmatic entropies co-determine lexical processing.

Thus far, we have considered probabilities and the corresponding entropy at the level of the inflectional class of regular feminine nouns in Serbian. However, the probability distribution of the inflected variants of a given lexeme may differ substantially from the probability distribution of the exponents at the level of the inflectional class. As a consequence, the corresponding entropies may differ substantially from each other as well. The extent to which these probability distributions differ is quantified by the relative entropy, also known as Kullback-Leibler divergence. Consider again the Serbian feminine noun "planina" (*mountain*) and its inflectional class as shown in Table 6.1. The third column lists the estimated probabilities for the paradigm, and the sixth column lists the probability distribution of the class. Let P denote the probability distribution of the paradigm, and Q the probability distribution of the inflectional class. The relative entropy can now be introduced as:

$$D(P||Q) = \sum_e \mathrm{Pr}_\pi(w_e) \log_2 \frac{\mathrm{Pr}_\pi(w_e)}{\mathrm{Pr}_\pi(e)}. \tag{6.12}$$

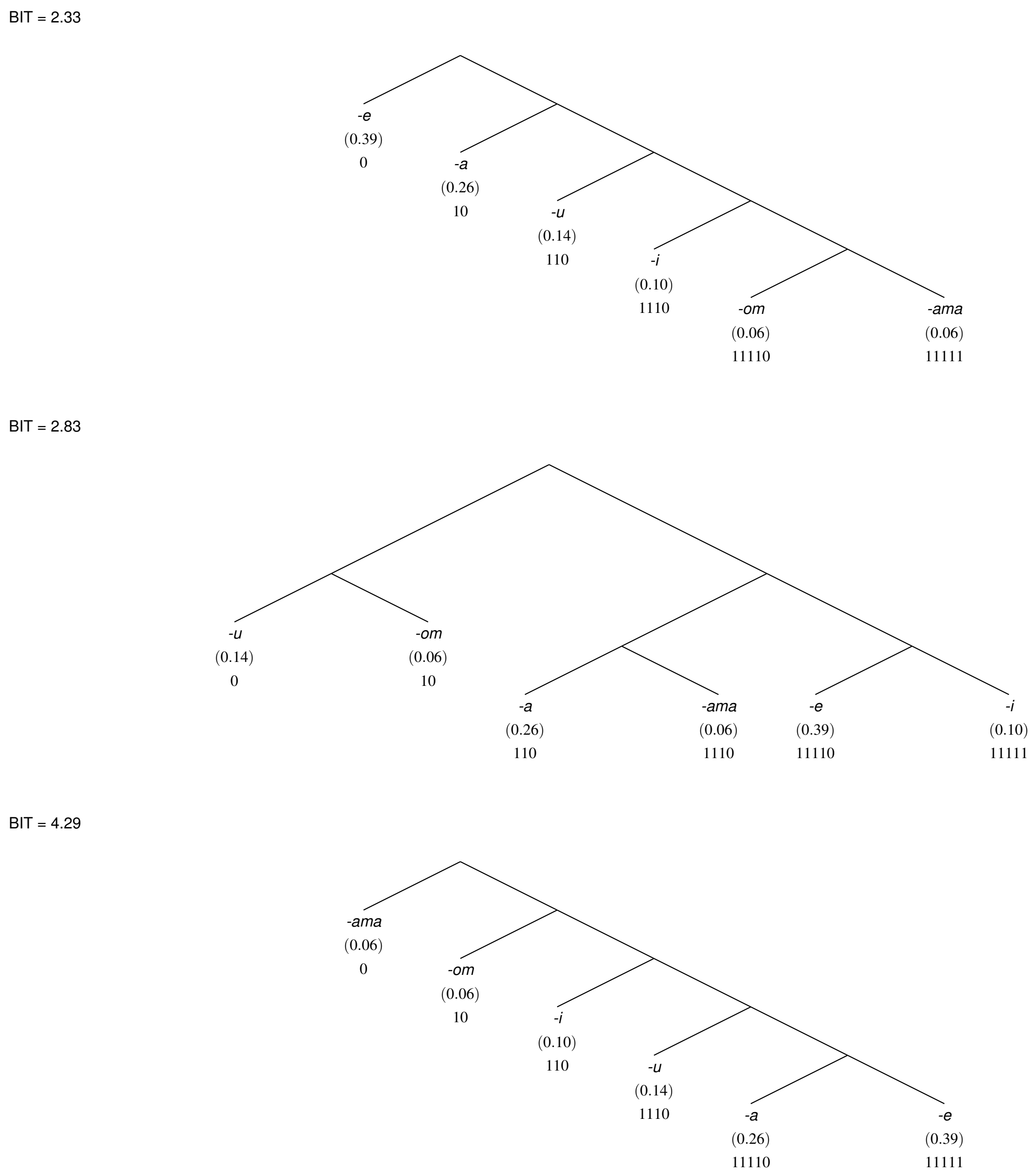


Figure 6.1: Optimal and non-optimal binary coding schemes for the inflectional class of regular feminine nouns in Serbian.

Relative entropy is also known as *information gain*,

$$\begin{aligned} D(P||Q) &= IG(\mathrm{Pr}_\pi(e|w)||\mathrm{Pr}_\pi(e|c)) \\ &= \sum_e \mathrm{Pr}_\pi(e|w)\log_2\frac{\mathrm{Pr}_\pi(e|w)}{\mathrm{Pr}_\pi(e|c)} \\ &= \sum_e \mathrm{Pr}_\pi(w_e)\log_2\frac{\mathrm{Pr}_\pi(w_e)}{\mathrm{Pr}_\pi(e)}, \end{aligned} \tag{6.13}$$

as it measures the reduction in our uncertainty about the exponent when going from the situation in which we only know its inflectional class to the situation in which we also know the lexeme. For “planina”, $H = 2.22$, and $D(P||Q) = 0.05$. For the masculine noun “prostor” listed in the lower half of Table 6.1, $H = 2.42$ and $D(P||Q) = 0.07$. In both cases, the two distributions are fairly similar, so the relative entropies (RE) are small. There is little that the knowledge of “planina” adds to what we already new about regular feminine nouns. If we approximate the probability distribution of “planina” with the probability distribution of its class, we are doing quite well. In section 6 we review a recent study demonstrating that RE is yet another information theoretic predictor of lexical processing costs.

We will now review a series of studies that illustrate how these information theoretic concepts help us to understand paradigmatic organization in the mental lexicon. Section 6 addresses the question of how the probability of an exponent given its inflectional class is reflected in measures of lexical processing costs. Section 6 reviews studies that make use of entropy and relative entropy to gauge lexical processing and paradigmatic organization. Finally, in section 6 we present new experimental results showing how concepts from information theory that proved useful for understanding inflection can help understanding derivation.

The Structure of Inflectional Classes

The consequence of the amount of information carried by an exponent for lexical processing has been explored in a series of experimental studies on Serbian (Kostić, 1991, 1995; Kostić, Marković & Baucal, 2003) . A starting point for this line of research is the amount of information carried by an exponent,

$$I_e = -\log_2 \mathrm{Pr}_\pi(e),$$

where Pr_π is estimated over all exponents within a class π. Kostić and colleagues noted that exponents are not equal with respect to their functional load. Some exponents (given their inflectional class) express only a few functions and meanings, others express many. Table 6.2 lists the functions and meanings for the exponents of the masculine and regular feminine inflectional class of Serbian. The count of numbers of functions and meanings for a given exponent were taken from an independent comprehensive lexicological survey of Serbian (see also the appendix of Kostić et al., 2003, for a shortlist of functions and meanings). Instead of using just the flat corpus-based relative frequencies, Kostić and colleagues propose to weight these probabilities for their functions and meanings. Let R_e denote the number of functions and meanings carried by exponent e. Then the weighted amount of information I'_e can be expressed as follows:

$$I'_e = -\log_2 \left(\frac{\Pr_\pi(e)/R_e}{\sum_e \Pr_\pi(e)/R_e} \right) \tag{6.14}$$

The ratio $(\Pr_\pi(e)/R_e)$ gives us the average probability per syntactic function/meaning for a given exponent. In order to take the other exponents within the inflectional class into account, this ratio is weighted by the sum of the ratios for each of the exponents (see, e.g., Luce, 1959). The resulting proportion is log-transformed to obtain the corresponding amount of information in bits. The partial effects of probability on the one hand, and the number of functions and meanings on the other, are shown in Figure 6.2. The weighted information is predicted to decrease with probability, and to increase with the number of functions and meanings. Table 6.2 lists I'_e for each of the exponents of the masculine and regular feminine inflectional classes.

To assess the predictivity of I'_e, Kostić, Marković and Baucal (2003) and Kostić (2008) calculated the mean lexical decision latency for each exponent in a given inflectional class, and investigated whether these mean latencies can be predicted from the weighted amounts of information such as those listed in Table 6.2. The Pearson correlation between the mean latencies and the weighted information scores was highly significant for both masculine and feminine nouns ($R^2 = 0.88$ for masculine nouns, $R^2 = 0.98$ for regular feminine nouns and $R^2 = 0.99$ for irregular feminine nouns). Furthermore, when mean reaction time is regressed on the weighted information load, the slopes of the regression lines are positive. Exponents carrying a greater average amount of information are more difficult to process. In other words, these data show that the average processing cost of an exponent in its inflectional class is very well predicted from its frequency and its functional load as given by (6.14) and illustrated above in Figure 6.2.

The probabilities that we considered in these analyses were estimated by summing across all words with a given exponent in a given inflectional class.

masculine nouns				
Exponent	Case and Number	Frequency	Functions and Meanings	Information
ø	nom sg	12.83	3	0.434
a	gen sg/acc sg /gen pl	18.01	109	5.128
u	dat sg /loc sg	4.64	43	5.744
om	ins sg	1.90	32	6.608
e	acc pl	2.21	58	7.243
i	nom pl	3.33	3	2.381
ima	dat pl/loc pl/ins pl	1.49	75	8.186

feminine nouns				
Exponent	Case and Number	Frequency	Functions and Meanings	Information
a	nom sg/gen pl	12.06	54	1.464
u	acc sg	5.48	58	2.705
e	gen sg /nom pl/acc pl	14.20	112	2.280
i	dat sg /loc sg	3.80	43	2.803
om	ins sg	1.94	32	3.346
ama	dat pl/loc pl/ins pl	1.69	75	4.773

Table 6.2: Exponents, case and number, frequency of the exponent, number of functions and meanings of the exponents, and amount of information carried by the exponents, for masculine nouns (upper table) and regular feminine nouns (lower table).

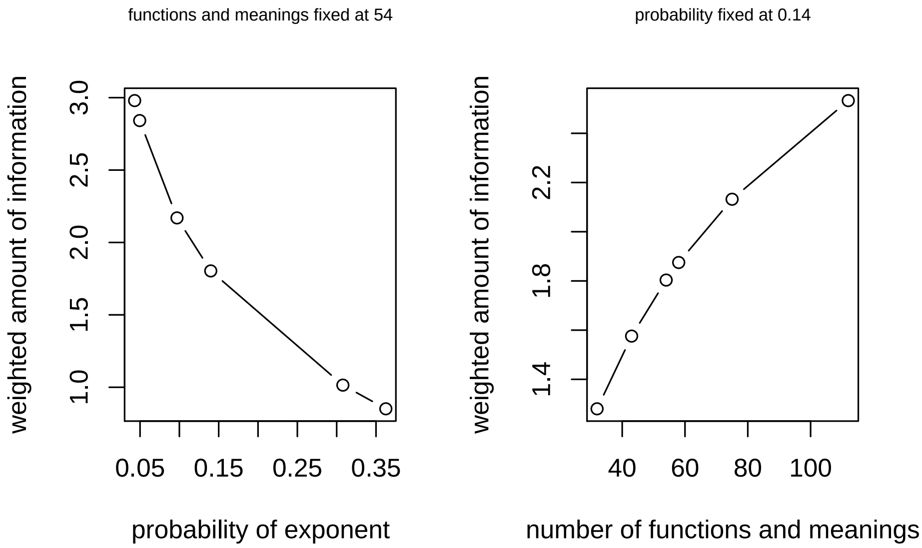


Figure 6.2: Partial effects of the probability of an exponent and its number of syntactic functions and meanings on the weighted amount of information I'_e.

In this way, the information about the probabilities of the different exponents in the inflectional paradigms of specific words is lost. In order to address the possibility that word-specific probabilities of exponents also co-determine lexical processing, Kostić, Marković and Baucal (2003) first applied the same weighting scheme underlying (6.14) at the level of individual lexemes, giving a lexeme-specific weighted information I'_{w_e}:

$$I'_{w_e} = -\log_2\left(\frac{\Pr_\pi(w_e)/R_e}{\sum_e \Pr_\pi(w_e)/R_e}\right). \quad (6.15)$$

Kostić, Marković and Baucal (2003) then constructed two sets of lexemes (henceforth Inflectional Groups) which contrasted maximally with respect to I'_{w_e}. For each of the two inflectional groups, the average value of I'_{w_e} for each of the exponents was calculated. Regression analysis showed that these group-averaged amounts of information contributed independently to the model, over and above the general class-based information values I'_e. As before, larger values for the group-averaged amounts of information I'_{w_e} corresponded to longer mean lexical decision latencies.

It is useful to probe the lexeme-specific weighted information (6.15) with respect to how it relates to the frequency properties of the lexeme and its inflected variants, as well as to the functional ambiguities existing in inflectional paradigms and classes. First consider a simple lower bound for (6.15):

$$\begin{aligned}
I'_{w_e} &= -\log_2\left(\frac{\Pr_\pi(w_e)/R_e}{\sum_e \Pr_\pi(w_e)/R_{w_e}}\right) \\
&= -\log_2\frac{\Pr_\pi(w_e)}{R_e}+\log_2\sum_e\frac{\Pr_\pi(w_e)}{R_e} \\
&\geq -\log_2\Pr_\pi(w_e)+\log_2 R_e+\log_2\prod_e\frac{\Pr_\pi(w_e)}{R_e} \\
&\geq -\log_2\Pr_\pi(w_e)+\log_2 R_e+\sum_e\log_2\frac{\Pr_\pi(w_e)}{R_e} \\
&\geq \log_2 R_e-\sum_e\log_2 R_e-\log_2\Pr_\pi(w_e)+\sum_e\log_2\Pr_\pi w_e. \qquad (6.16)
\end{aligned}$$

The third term is the amount of information carried by the inflected variant, I_{w_e}, see (6.2), and $\sum_j \log_2 \Pr_\pi w_j$ is a measure of the lexeme's stem frequency, evaluated by summing the log frequencies of its inflected variants rather than by summing the bare frequencies of its inflected variants. Consequently, at the level of the inflected variant, the amount of information (6.16) incorporates two well-known frequency effects that have been studied extensively in the processing literature. The word frequency effect ($-\log_2 \Pr_\pi(w_e)$) is facilitatory, as expected. Surprisingly, the stem frequency effect ($\sum_e \log_2 \Pr_\pi w_e$) is predicted to be inhibitory. However, both frequency effects are complemented by measures gauging ambiguity. Ambiguity of the given exponent is harmful, whereas ambiguity in the rest of the paradigm is facilitatory. Thus, the stem frequency effect emerges from this model as a composite effect with both an inhibitory and a facilitatory component. This may help explain why stem frequency effects are often much less robustly attested in experimental data (see, e.g., Baayen, Wurm & Aycock, 2008) compared to word frequency effects.

In order to evaluate how well the lower bound given in (6.16) approximates the original measure given in (6.15), we examined the exponent frequency, the group averages of the functions and meanings, the information values, and the mean reaction times for the two inflectional groups for regular feminine nouns, as listed in Table 6.3 (data from Kostić, Marković & Baucal, 2003). Note that the terms in (6.16) represent the ambiguity of the exponent, the joint ambiguity of all exponents, the word frequency effect of the inflected variant, and the stem frequency effect of its lexeme.

For the data in Table 6.3, we first carried out a linear regression analysis with RT as dependent variable and I' and Inflectional Group as predictors. The R^2 for this model was 0.863. We then carried out a linear regression analysis, but now with the two measures that figure in the lower bound of the amount of information (6.16) as predictors: exponent frequency and the number of functions and meanings of the exponent R. The R^2 of this model was 0.830. Furthermore, the effect of the number of functions and meanings was inhibitory ($\hat{\beta} = 27.5, t(8) = 2.512, p = 0.0362$) and the effect of exponent frequency was facilitatory ($\hat{\beta} = -5.2, t(8) = -5.813, p = 0.0004$) as expected given (6.16). In other words, the two variables that according to (6.16) should capture a substantial proportion of the variance explained by the amount of information I', indeed succeed in doing so: 0.830 is 96% of 0.863.

Exponent	Exponent frequency	R	I'	Inflectional Group	RT
a	12.06	3.99	1.46	high	674
e	14.20	4.72	2.28	high	687
i	3.80	3.76	2.80	high	685
u	5.48	4.06	2.71	high	693
om	1.94	3.47	3.35	high	718
ama	1.69	4.32	4.77	high	744
a	12.06	3.99	1.46	low	687
e	14.20	4.72	2.28	low	685
i	3.80	3.76	2.80	low	730
u	5.48	4.06	2.71	low	712
om	1.94	3.47	3.35	low	722
ama	1.69	4.32	4.77	low	746

Table 6.3: Mean reaction times in visual lexical decision (RT), exponent frequency, number of functions and meanings of the exponent (R), amount of information (I), and Inflectional Group (high versus low by-word amount of information) for the Exponents of the regular feminine declension class.

The lower bound estimate in (6.16) is a simplification of the full model I'_{w_e} defined by (6.15). Because the simplification allows us to separate the word and stem frequency effects, it clarifies that these two frequency effects are given the same overall weight. There is evidence, however, that stem frequency has a much more modest weight than word frequency (Baayen, Wurm & Aycock, 2008), and may even have a different functional form. This suggests that it may be preferable to rewrite (6.15) as:

$$I'_{w_e} = -\log_2 \left(\frac{\omega_1 \, \mathrm{Pr}_\pi(w_e)/R_e}{\omega_2 \, \sum_e \mathrm{Pr}_\pi(w_e)/R_e} \right), \qquad (6.17)$$

with separate weights ω for numerator and denominator. On the other hand, at the level of a given class the lower bound estimate in (6.17) reduces to the exponent frequency and the overall class frequency. Some preliminary experimental evidence for the relevance of exponent frequency (in the simplified form of inflectional formative frequency) for English is available in Baayen, Wurm and Aycock (2008), along with evidence for frequency effects for derivational affixes. However, it is presently unclear how class frequency could be generalized and gauged with derivations. Inflectional classes are well contained and it is easy to count-out their overall frequencies. Contrariwise, within and between derivational classes there are no clear partitions of the lexical space. While inflected words, in general, belong to only one inflectional class, any given base word may participate in several derivations. We shall address the issue of relations between base words and their derivatives in co-determining lexical processing in further detail in section 6.

It is also useful to rewrite (6.14) along similar lines as we did for (6.15). In this case, the lower bound for the amount of information can be written as the sum of two conditional probabilities. First consider the probability of exponent e given its inflectional class c:

$$\begin{aligned}\Pr(e|c) &= \frac{\Pr(e,c)}{\Pr(c)} \\ &= \frac{\Pr(e)}{\Pr(c)}.\end{aligned}$$

(Note that the probability of an exponent is defined strictly with respect to its inflectional class. We never sum frequencies of exponents across inflectional classes.) The information corresponding to this conditional probability is

$$\begin{aligned}I_{e|c} &= -\log_2 \frac{\Pr(e)}{\Pr(c)} \\ &= -\log_2 \Pr(e) + \log_2 \Pr(c) \\ &= -\log_2 \Pr(e) + \log_2 \sum_j \Pr(e_j) \\ &\geq -\log_2 \Pr(e) + \log_2 \prod_j \Pr(e_j) \\ &\geq -\log_2 \Pr(e) + \sum_j \log_2 \Pr(e_j) \\ &= I'_{e|c} \end{aligned} \qquad (6.18)$$

Note that $I'_{e|c}$ is a lower bound of $I_{e|c}$.

Next, let R_e denote the number of functions and meanings of exponent e in class c, and let R_c denote the total count of functions and meanings within the class. The conditional probability of the functions and meanings of exponent e given its class c is

$$\begin{aligned}
\Pr(R_e|R_c) &= \frac{\Pr(R_e, R_c)}{\Pr(R_c)} \\
&= \frac{\Pr(R_e)}{\Pr(R_c)} \\
&= \frac{R_e}{R_c}
\end{aligned}$$

and the corresponding information is therefore

$$\begin{aligned}
I_{R_e|R_c} &= -\log_2 \frac{R_e}{R_c} \\
&= -\log_2 R_e + \log_2 R_c \\
&= -\log_2 R_e + \log_2 \sum_j R_j \\
&\leq -\log_2 R_e + \log_2 \prod_j R_j \\
&\leq -\log_2 R_e + \sum_j \log_2 R_j \\
&= I'_{R_e|R_c} \qquad (6.19)
\end{aligned}$$

Here, $I'_{R_e|R_c}$ is an upper bound of $I_{R_e|R_c}$.

Taking into account that $I'_{e|c}$ is a lower bound of $I_{e|c}$, and that $I'_{R_i|R_c}$ is an upper bound of $I_{R_i|R_c}$, we can now approximate (6.14) as follows:

$$\begin{aligned}
I_{w_e} &\approx \log_2 R_e - \sum_j \log_2 R_j - \log_2 \Pr_\pi w_e + \sum_j \log_2 \Pr_\pi w_j \\
&\approx -I'_{R_e|R_c} + I'_{e|c}. \qquad (6.20)
\end{aligned}$$

In other words, the amount of information as defined in (6.14) is related to the sum of two conditional probabilities: (i) the probability of the exponent given its class, and (ii) the probability of the ambiguity of the exponent given the ambiguity in its class. The partial effects of these two conditional probabilities are shown in Figure 6.3. As expected, the partial effects are very similar to those shown in Figure 6.2.

At this point, the question arises why $I'_{R_e|R_c}$ appears with a negative sign in (6.20). To answer this question, we need to consider exponents within their classes, and differentiate between the functions and meanings that an inflected form can have in the discourse. Consider the case in which $R_e \to R_c$. The more the functions expressed by exponent e become similar to the universe of functions and meanings carried by the inflectional class, the less distinctive the exponent becomes. In other words, an exponent is more successful as a distinctive functional unit of

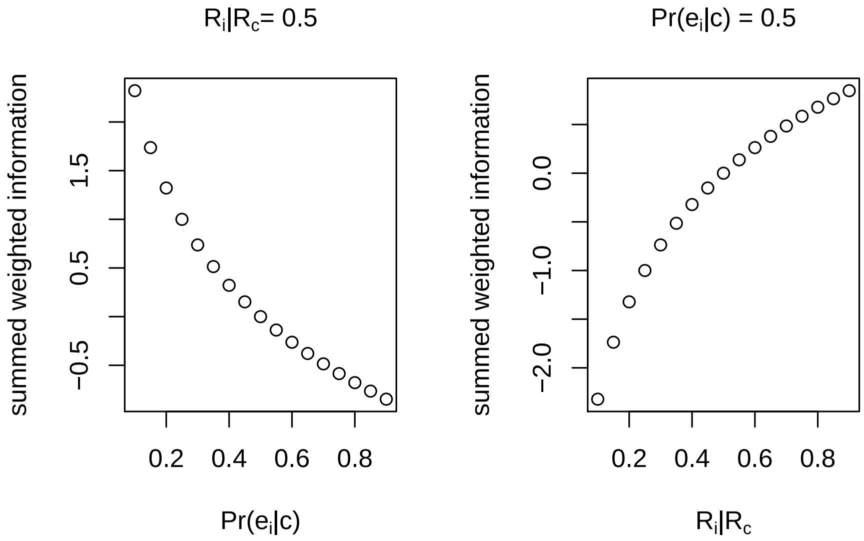


Figure 6.3: The left panel shows the partial effect of the information carried by the probability of the exponent given its class $I'_{e|c}$. The right panel shows the partial effect of the information carried by the proportion of the number of functions and meanings conditioned on the total number of functions and meanings for the class $I'_{R_e|R_c}$. Both partial effects are calibrated for the other effect evaluated at 0.5, and are calculated straightforwardly from (6.20).

the language when $R_c - R_e$ is large. If so, the amount of information $I'_{R_e|R_c}$ is large, and hence I_{w_e} in (6.20) is small, and as a consequence processing latencies are reduced. By contrast, an exponent for which $I_{R_e|R_c}$ is small is dysfunctional, and therefore harder to process, leading to longer processing latencies.

The information structure of paradigms

Entropy

Thus far, we have considered the processing load of an inflected form given its paradigm, or an exponent, given its inflectional class. Moscoso del Prado Martín, Kostić and Baayen (2004) added a new dimension to the experimental study of morphological conectivity by considering the cost of the complexity of a paradigm as such, gauged by means of the entropy measure H. Figure 6.1 is helpful for discussing the difference between Kostić's approach and the one developed by Moscoso del Prado and his colleagues. Ignoring the weighting for numbers

morphological family		inflectional paradigms		merged paradigms	
word	F	word	F	word	F
neighbour	901	*neighbour*	343	*neighbour*	343
neighbourhood	407	*neighbours*	558	*neighbours*	558
neighbouring	203			*neighbourhood*	386
neighbourliness	3	*neighbourhood*	386	*neighbourhoods*	21
neighbourly	14	*neighbourhoods*	21	*neighbouring*	203
				neighbourliness	3
				neighbourly	14

Table 6.4: Morphological family and inflectional paradigms for *neighbor*.

of functions and meanings, Kostić's measure simplifies to $-\log_2(\Pr_\pi(e))$, which reflects the number of steps from the root node to the leaf node of the exponent e in an optimal binary coding scheme (see the upper panel; for numbers of nodes that are integer powers of two, the $-\log_2(\Pr_\pi(e))$ is exactly equal to the number of steps). However, this measure is insensitive to the size and configuration of the tree. To capture these aspects of the tree, we can make use of the entropy measure. The entropy, which is the same for each and every member of the paradigm, quantifies the expected number of steps from the root to a leaf node.

Moscoso del Prado Martín, Kostić and Baayen (2004) applied the entropy measure to paradigms in Dutch, but used a much broader definition of paradigms that extended the concept of the morphological family. Table 6.4 shows the words listed in CELEX that contain *neighbour* as a constituent. The left two columns list the morphological family as defined by Schreuder and Baayen (1997) , the middle columns list the inflected variants that were found for two of the members of the family, and the rightmost columns list the set that merges the family members with the inflected variants. Moscoso del Prado and colleagues calculated the entropy over this merged set, and proposed this entropy as an enhanced measure for capturing the morphological family size effect. They pointed out that when all family members are equiprobable, the entropy of the family reduces to the log of the number of family members. Since it is exactly this log-transformed count that emerged as predictive for processing latencies, the entropy of the family can be viewed as a principled way of weighting family members for their token frequency.

Moscoso del Prado and colleagues combined this generalized entropy measure with the amount of information carried by a word (inflected or uninflected) as estimated from its relative frequency to obtain what they called the information residual:

$$I_R = I_w - H = \log N - \log_2 F_w - H. \tag{6.21}$$

This information residual performed well in a series of post-hoc analyses of processing of Dutch complex words.

By bringing several measures together in a single predictor, I_R, stem frequency and entropy receive exactly the same regression weight:

$$\begin{aligned} RT &\propto \beta_0 + \beta_1 I_R \\ &= \beta_0 + \beta_1 (I_w - H) \\ &\quad \beta_0 - \beta_1 \log_2 F_w - \beta_1 H. \end{aligned} \tag{6.22}$$

However, subsequent work (Baayen, Feldman & Schreuder, 2006) suggests that frequency, the entropy calculated over the morphological family while excluding inflected variants, and the entropy of the paradigms of individual lexemes should be allowed to have different importance (i.e, different β weights). Their study examined a wide range of lexical predictors for simple English nouns and verbs, and observed independent effects of inflectional entropy (henceforth H_i) across both the visual lexical decision and word naming tasks. An effect of derivational entropy (henceforth H_d) was present only in the visual lexical decision task. Here, it emerged with a U-shaped curve, indicating the presence of some inhibition for words with very information-rich families. In their study of the lexical processing of 8486 complex words in English, Baayen, Wurm and Aycock (2008) also observed an independent facilitatory effect of inflectional entropy, side by side with a facilitatory effect of the family size of the lexeme.

These results suggest that, when considered in terms of optimal binary coding schemes, inflected words and lexemes should not be brought together in one encompassing binary tree. Instead, lexemes form one tree, and each lexeme then comes with its own separate disjoint tree for its inflected variants.

Inflectional paradigms in languages such as Dutch and English are trivially simple compared to the paradigms one finds in morphologically rich languages. This raises the question to what extent entropy measures inform us about the processing complexity of more substantive paradigmatic structure. We address this issue for nominal paradigms in Serbian.

Relative entropy

When the inflectional entropy is computed for a given lexeme, it provides an estimate for the complexity of this lexeme's inflectional paradigm. This measure, however, does not take into account the complexity of the inflectional class, and the extent to which the probability distribution of a lexeme's paradigm diverges from the probability distribution of its inflectional class. We could consider bringing the entropy of the inflectional class into our model, but this class entropy would be the same for all lexemes in the class. Hence, it would not be much more informative than a plain name for that class (for example, Latin declension I, or Serbian declension III). Therefore, Milin, Filipović Đurđević and Moscoso del Prado Martín (2008) considered the simultaneous influence of paradigms and classes on the processing of inflected nouns in Serbian by means of relative entropy, RE.

Milin, Filipović Đurđević and Moscoso del Prado Martín (2008) investigated whether relative entropy is predictive for lexical processing in visual lexical decision using masculine and feminine nouns with the case endings *-om, -u* and *-e*. A mixed-effects analysis with word frequency and stem frequency, bigram frequency, number of orthographic neighbors and entropy as covariates revealed an independent inhibitory effect of RE, as shown in the lower right panel of Figure 6.4. Comparison with the other significant partial effects in the model shows that the magnitude of the effect of RE is comparable to that of stem frequency and orthographic neighborhood size. However, the effect of the entropy did not reach significance ($p > 0.15$).

What this experiment shows is that it is neither the probability distribution of the inflected variants in a word's paradigm, nor the probability distribution in its inflectional class considered separately that are at issue, but rather the divergence between the two distributions. The greater this divergence, the longer the response latencies. A similar pattern was observed for the accuracy measure as well: the greater the divergence of the probability distribution of the paradigm from the probability distribution of the class, the more errors were made.

From the perspective of cognitive psychology, these results are interesting in that they provide further evidence for the importance of structured lexical connectivity. From the perspective of linguistic morphology, they support the theoretical concepts of paradigms and inflectional classes. Combined with the presence of a strong effect of the word frequency, an effect that is much stronger than the effect of the word's stem (compare the upper panels in Figure 6.4), these results provide strong support for Word and Paradigm morphology (Blevins, 2003, 2006; Matthews, 1974)

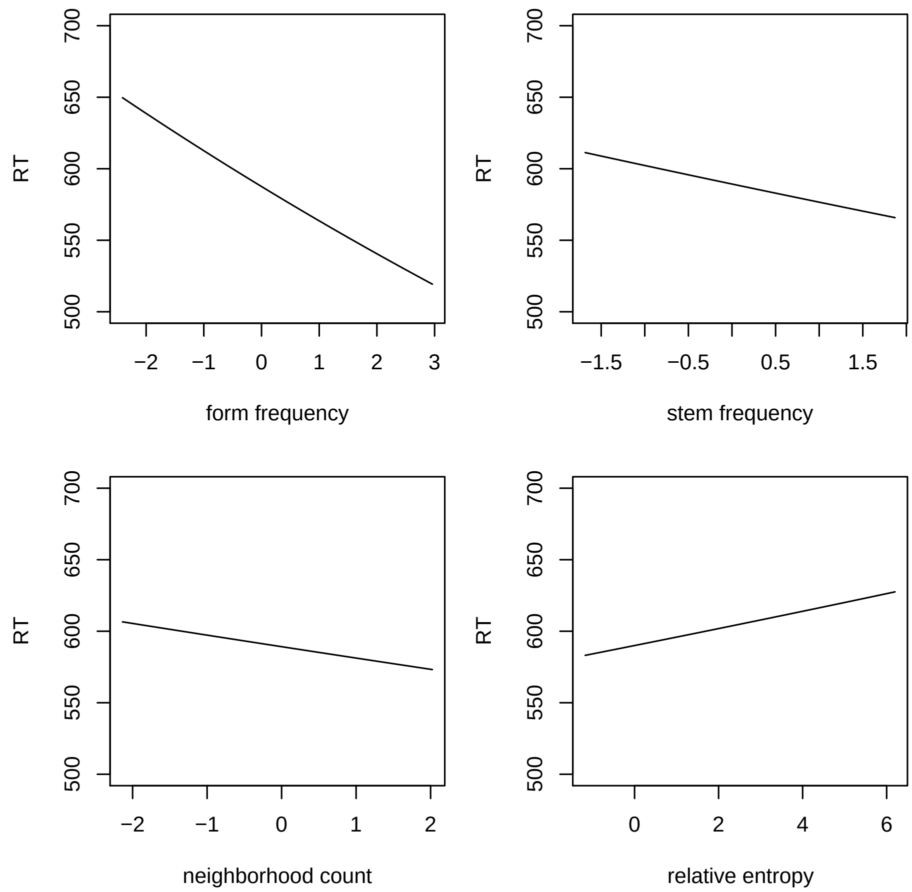


Figure 6.4: Partial effects of distributional predictors for the response latencies in visual lexical decision to Serbian nouns (Milin *et al.*, 2008).

and for exemplar-based approaches to lexical processing in general (see, e.g., Baayen, 2003).

Paradigmatic structure in derivation

In languages such as Dutch or English, morphological families consist predominantly of compounds. As a consequence, the family size effect (cf., Schreuder & Baayen, 1997) is driven almost exclusively by lexical connectivity between compounds. Little is known about the role of derived words. The problem here is that a given base word combines with only a handful of derivational affixes at best. Counts of the number of different prefixes and suffixes that English monomorphemic base words combine with, based on the English section of the CELEX lexical database (Baayen, 1995), illustrate that 60% English monomorphemic base words combine with only one affix. Table 6.5 shows a steep decrease (a Zipfian distribution) in the number of derivational affixes that are attested for a given base word. The verbs *act* and *play* are exceptional in combining with 11 different affixes. The maximum family size in English, 187, observed for *man*, is an order of magnitude larger. With such small numbers of derived family members, it becomes very difficult to gauge the role of a strictly derivational family size count in lexical processing.

Derived words, however, enter into more systematic relations than most compounds, even when we take into account that the meaning of a compound is predictable from its constituents to a much greater extent than has traditionally been assumed (Gagne, 2001; Gagne & Shoben, 1997). For instance, derived adjectives with the prefix *un-* systematically express negation. Taking this fact into account, we asked ourselves whether such systematic relations between base words and their derivatives co-determine lexical processing. As a first step towards an answer, we introduce two simple concepts: the mini-paradigm and the mini-class. Here, the term mini-paradigm refers to pairs of base words and their derivatives. Thus, *kind* and *unkind* form a mini-paradigm, and so do *clear* and *clearly*. In the same line, the term mini-class refers to the set of mini-paradigms sharing the same derivational affix. All pairs of base words and the corresponding *un-* derivatives constitute the mini-class of: *kind - unkind, true - untrue, pleasant - unpleasant*, etc. Mini-paradigms and mini-classes approximate inflectional paradigms and inflectional classes in the sense that the semantic relations within the pairs tend to be more consistent and transparent than in general morphological families or in

Number of affixes	Count of base words
1	3449
2	1391
3	516
4	202
5	105
6	31
7	13
8	11
9	2
10	3
11	2

Table 6.5: The number of monomorphemic base words that can attach the given number of affixes (prefixes or suffixes) when forming bi-morphemic derived words.

families of derived words with different prefixes and suffixes.

In what follows, we therefore investigate whether the measures of entropy and relative entropy are significant predictors for lexical processing when applied to mini-paradigms and mini-classes.

Materials

We selected six suffixes and one prefix, for which we extracted all formations listed in the CELEX lexical database and for which latencies were also available in the English Lexicon Project (Balota, Yap, Cortese *et al.*, 2007) for both the derived word and its base. The resulting counts of formations are available in Table 6.6, cross-classified by whether the base word is simple or complex. For all words, we extracted from CELEX their frequency of occurrence, their length in letters, the number of synsets for the base as listed in WordNet (Beckwith, Fellbaum, Gross & Miller, 1991; Miller, 1990), and studied by Baayen, Feldman and Schreuder (2006), the family size of the base (calculated from the morphological parses in CELEX), and their frequency in the demographic subcorpus of conversational English in the British National Corpus (Burnard, 1995). We included these variables in order to make sure that potential paradigmatic effects are not confounded with other lexical distributional properties. From the English Lexicon Project, we added the by-item

	simple base	complex base
-able	70	0
-er (comparative)	98	0
-er (deverbal)	240	24
-ly (adverbial)	21	355
-ness (complex base)	0	65
-ness (simple base)	152	0
-est (superlative)	95	0
un-	18	111

Table 6.6: Affixes in the study based on latencies extracted from the English Lexicon Project, cross-classified by the complexity of their base words.

mean naming latencies and the by-item mean lexical decision latencies.

For each pair of base and derivative, we calculated its entropy and its relative entropy. For the derived words, the entropy of the mini-paradigm was calculated on the basis of the relative frequencies of the derivative and its base word (e.g., for *kind* and *unkind*, the relative frequencies are $72/(72+390)$ and $390/(72+390)$). For the base words, we distinguished between base words with only one derivative, and base words with two or more derivatives. For base words with a single derivative, the procedure for estimating the entropy was the same as for derived words. For base words with more than one derivative, the problem arises how to calculate entropies. Selection of a single derivative seems arbitrary. Taking all derivations linked with a given base word into account is possible, but then the mini-class distribution would contain the maximum number of 11 relative frequencies (see Table 6.5), most of which would be zero for almost all words. We therefore opted for taking only two relative frequencies into account when calculating the entropy: the frequency of the base itself, and the summed frequency of all its derivatives.

The probability distribution for a given mini-class was obtained by summing the frequencies of all base words in the class on the one hand, and all derivatives in the class on the other hand. The resulting frequencies were then transformed into relative frequencies. These relative frequencies then served as the Q distribution (also known as the reference distribution) for the calculation of the relative entropy.

In the following analyses, frequency measures, family size, number of synsets, and response latencies were log-transformed to eliminate the adverse effect of outliers on the model fit.

Derived words

We investigated the predictivity of the entropy and relative entropy measures for word naming and lexical decision latencies to the derived words. For that, we applied linear mixed-effects modeling (Baayen, 2008; Baayen, Davidson & Bates, 2008; Bates, 2005, 2006), with Task (lexical decision versus naming) as a fixed-effect factor, and with the set of relevant covariates including length, (written) base frequency, (written) word frequency, spoken word frequency, number of synsets in WordNet, morphological family size, entropy and relative entropy. Word and affix were considered as random effects.

For the covariates, we investigated whether nonlinearity was present. This turned out to be the case only for word length. We also observed interactions of Task with word frequency and spoken word frequency, with length (only the quadratic term), and with entropy and relative entropy. Finally, we considered whether by-word or by-affix random slopes were required. It turned out that by-affix random slopes were necessary only for the two entropy measures.

Inspection of the coefficients for the entropy measures in the resulting model revealed that entropy and relative entropy had positive coefficients of similar magnitude ($H: 0.034, \hat{\sigma} = 0.025; RE: 0.058, \hat{\sigma} = 0.016$), with small differences across the two tasks. In word naming, the effect of entropy was slightly larger, while the effect of relative entropy was fractionally smaller (H in naming: $0.034 + 0.041$; RE in naming: $0.058 - 0.014$).

These observations invite a simplification of the regression model. Let β_0 denote the coefficient for the intercept, and let β_1 and β_2 denote the coefficients for entropy and relative entropy respectively. Given that β_1 and β_2 are very similar, we can proceed as follows:

$$\begin{aligned}\beta_0 + \beta_1 H + \beta_2 RE &\approx \beta_0 + \beta_1 H + \beta_1 RE \\ &= \beta_0 + \beta_1 (H + RE). \end{aligned} \tag{6.23}$$

Interestingly, the sum of entropy and relative entropy is equal to another information theoretical measure, the *cross entropy* (CE) (Cover & Thomas, 1991; Manning & Schutze, 1999). Applied to the present data, we have

$$\begin{aligned} CE &= H + RE = \\ &= -\sum_L \Pr_\pi(w_L)\log_2(\Pr_\pi(w_L)) + RE \\ &= -\sum_L \Pr_\pi(w_L)\log_2(\Pr_\pi(w_L)) + \sum_L \Pr_\pi(w_L)\log_2 \frac{\Pr_\pi(w_L)}{\Pr_\pi(c_L)} \end{aligned}$$

$$= -\sum_L \Pr_\pi(w_L)\log_2(\Pr_\pi(c_L)). \tag{6.24}$$

In (6.24), L indexes the base and derived lexemes for mini-paradigms, and the sets of base words and derived words for the mini-class. Thus, $\Pr_\pi(w_L)$ denotes the probability of a base or derived lexeme in its mini-paradigm, and $\Pr_\pi(c_L)$ denotes the corresponding probability in the mini-class. Technically, the cross entropy between the probability distribution of the mini-paradigm and the probability distribution of the mini-class measures the average number of bits needed to identify a form from the set of possible forms in the mini-paradigm, if a coding scheme is used based on the reference probability distribution $\Pr_\pi c_e$ of the mini-class, rather than the "true" distribution $\Pr_\pi w_e$ of the mini-paradigm. More informally, we can interpret the cross entropy as gauging the average amount of information in the mini-paradigm, corrected for the departure from the prior reference distribution of the corresponding mini-class.

We therefore replaced entropy H and relative entropy RE as predictors in our regression model by a single predictor, the cross entropy CE, and refitted the model to the data. After removal of outliers and refitting, we obtained the model summarized in Table 6.7 and visualized in Figure 6.5. The standard deviation of the by-word random intercepts was 0.0637, the standard deviation for the by-affix random intercepts was 0.0399, the standard deviation for the by-affix random slopes for cross entropy was 0.0277, and the standard deviation for the residual error was 0.0663. All random slopes and random intercepts were supported by likelihood ratio tests (all p-values < 0.0001).

With respect to the control variables, we note that word length was a strongly nonlinear (positively accelerated) predictor for especially lexical decision, with longer lengths eliciting elongated response latencies. The word frequency effect was similar for both tasks, albeit slightly stronger for lexical decision. Similarly, the spoken word frequency added facilitation specifically for lexical decision. The effect of number of synonyms, as gauged with the help of the synset count, was facilitatory and the same across the two tasks. The effect of cross entropy was inhibitory, and also did not differ across tasks. Its effect size (roughly 100 ms) exceeds that of the spoken frequency effect and that of the number of meanings. Interestingly, the model with cross entropy as predictor provides an equally tight fit to the data as the model with entropy and relative entropy as predictors, even though the latter model had two additional parameters (a beta coefficient for a second entropy measure, and a random-effects standard deviation for by-item slopes for the second entropy measure): the log likelihood of the simpler model with

	Estimate	Lower	Upper	P
Intercept	6.6679	6.5830	6.7607	0.0001
Task=naming	-0.1419	-0.2158	-0.0688	0.0001
length (linear)	0.0056	-0.0109	0.0228	0.5162
length (quadratic)	0.0012	0.0004	0.0020	0.0034
written frequency	-0.0382	-0.0428	-0.0333	0.0001
spoken frequency	-0.0183	-0.0245	-0.0117	0.0001
synset count	-0.0277	-0.0339	-0.0212	0.0001
cross entropy	0.0565	0.0164	0.0937	0.0076
Task=naming: written frequency	0.0067	0.0022	0.0112	0.0036
Task=naming:length (linear)	0.0132	-0.0025	0.0283	0.0914
Task=naming:length (quadratic)	-0.0011	-0.0019	-0.0003	0.0026
Task=naming:spoken frequency	0.0124	0.0062	0.0186	0.0001

Table 6.7: Partial effects of the predictors for the visual lexical decision and naming latencies to derived words. The reference level for Task is lexical decision. Lower, Upper: 95% highest posterior density interval; P: Markov chain Monte Carlo p-value.

cross entropy was 2364, while for the more complex model with entropy and relative entropy it was 2362.[3] From this, we conclude that the relevant entropy measure for understanding the role of paradigmatic complexity during lexical processing of derived words is the cross entropy measure.

The synset measure in our data estimates the number of meanings that a base word has (e.g., *bank* as a part of the river and a financial institution). Generally, the meaning of a derivative builds on only one of the meanings of its base word (e.g., *embank*). The lower the number of synsets, the tighter we may expect the relationship between the base and its derivatives to be. The synset measure does not interact with cross entropy, nor does it substantially affect the estimate of its slope. To further rule out potential semantic confounds, we also considered a semantic measure that specifically gauges the semantic similarity between a given derived word and its base. The measure that we used is the LSA score for the distance between the derived word and its base in co-occurrence space (Landauer & Dumais, 1997), using the software available at `http://lsa.colorado.edu`. For the subset of our mini-paradigms, the LSA scores elicited a significant facilitatory effect on lexical decision latencies ($\hat{\beta} = -0.1196,\ p = 0.0001$). As for the synset

[3]A greater log likelihood implies a better fit (for technical details consult Crawley, 2002).

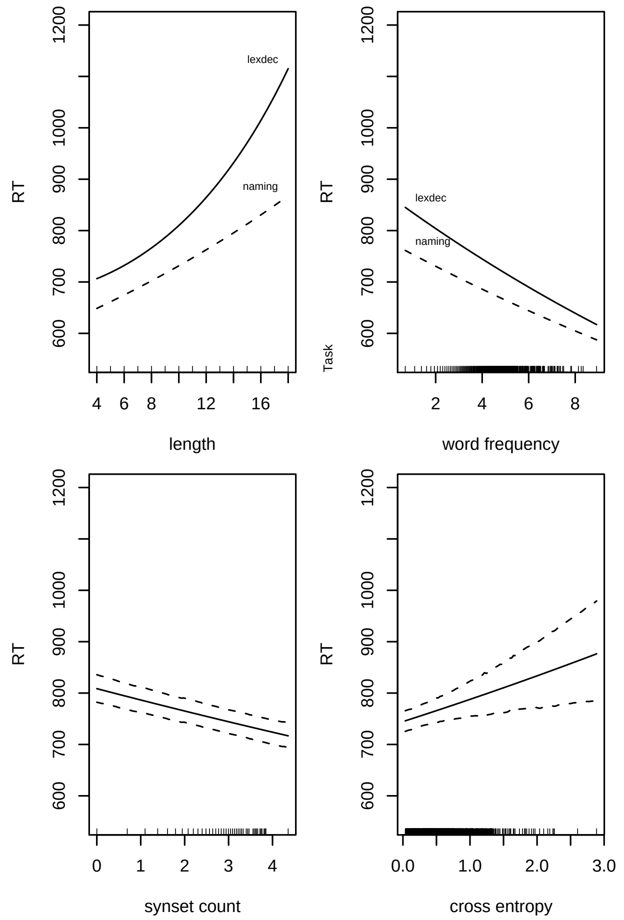


Figure 6.5: Partial effects of the predictors for word naming and visual lexical decision latencies for derived words. The lower panels are calibrated for visual lexical decision, and come with 95% highest posterior density confidence intervals.

	slope
-est (superlative)	0.097
-ly (adverbial)	0.090
-ness (complex base)	0.086
-able	0.068
-er (comparative)	0.054
-er (deverbal)	0.031
un-	0.021
-ness (simple base)	0.004

Table 6.8: Estimated slopes for derived words for the different mini-classes, positioned in decreasing order.

measure, there was no significant effect for word naming. Crucially, the measure of cross entropy retained significance also when the pairwise semantic similarity between base and derived word in mini-paradigms was taken into account.

The presence of random slopes for cross entropy in this model indicates that the effect of cross entropy varied with mini-class. Table 6.8 lists the individual slopes for the different mini-classes that we considered. Slopes range from 0.097 for superlative *-est* to 0.004 for *-ness* formations derived from simple base words.

Base words

Because complex base words (e.g., *surprising*) come with predictors such as the frequency of the stem (*surprise*) that do not apply to the simple base words, we analyzed the simple and complex base words separately. We proceeded in the same way as for the derived words. We fitted a mixed-effects model to the data, observed that again the coefficients for entropy and relative entropy were very similar and statistically indistinguishable in magnitude and had the same sign, replaced the two measures by the cross entropy measure, refitted the model and removed overly influential outliers.

The coefficients of a mixed-effects model fitted to the lexical decision and naming latencies to the complex base words are listed in Table 6.9. The corresponding partial effects are graphed in Figure 6.6.

As for the preceding data sets, we find effects of word length (longer words elicit longer latencies, upper left panel) and word frequency (more frequent words elicit shorter latencies, upper center panel). Adding frequency of use in spoken

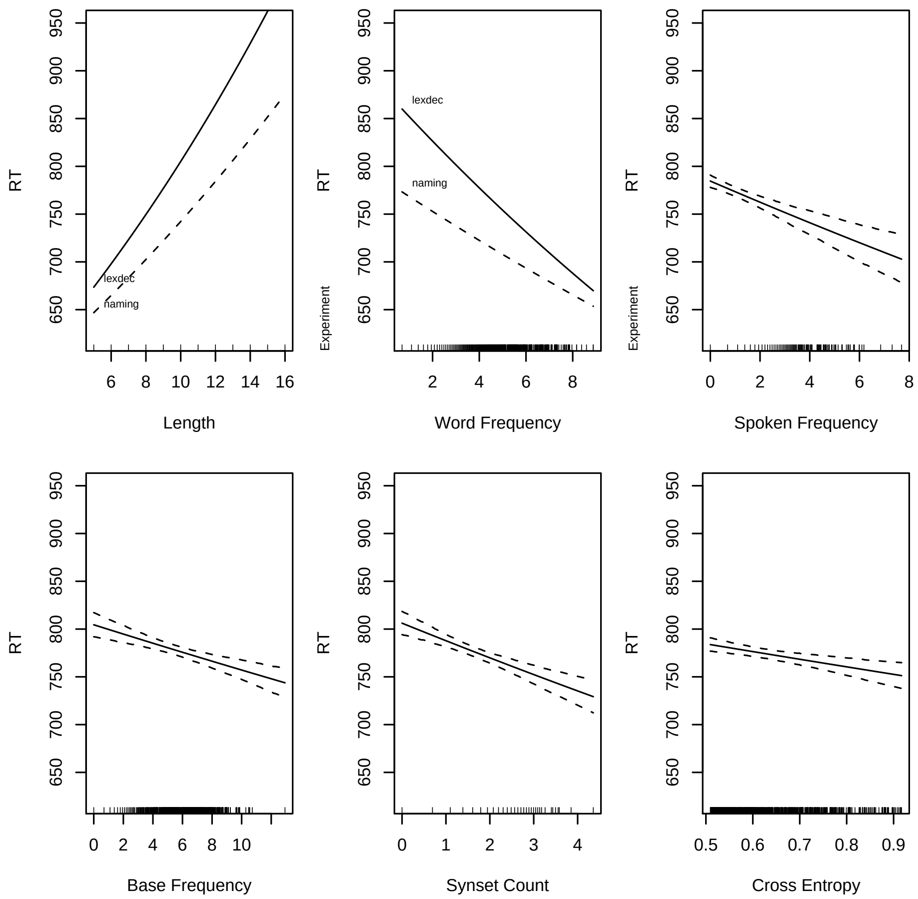


Figure 6.6: Partial effects of the predictors for word naming and visual lexical decision latencies for complex base words. Markov chain Monte Carlo based 95% confidence intervals are shown for those predictors that do not enter into interactions.

	Estimate	Lower	Upper	P
Intercept	6.6006	6.5428	6.6596	0.0001
Experiment=naming	-0.0397	-0.0750	-0.0031	0.0326
Length	0.0357	0.0325	0.0387	0.0001
Word Frequency	-0.0305	-0.0363	-0.0250	0.0001
Spoken Frequency	-0.0143	-0.0195	-0.0090	0.0001
Base Frequency	-0.0061	-0.0086	-0.0035	0.0001
Synset Count	-0.0230	-0.0311	-0.0147	0.0001
cross entropy	-0.1038	-0.1605	-0.0483	0.0002
Experiment=naming:Length	-0.0082	-0.0115	-0.0052	0.0001
Experiment=naming:Word Frequency	0.0100	0.0057	0.0141	0.0001

Table 6.9: Partial effects of the predictors for word naming and visual lexical decision latencies for complex base words. Lower, Upper: 95% highest posterior density interval; P: Markov chain Monte Carlo p-value.

English as a predictor again contributes significantly to the model over and above the written frequency measures (upper right panel). The frequency of the base word (lower left panel of Figure 6.6) also emerged as a significant predictor, but with a slope that is substantially shallower than that of the word frequency effect. The Synset Count of the embedded base word is predictive as well. It is facilitatory, just as observed for the derived words (lower center panel). Finally, the lower right panel shows that there is a small effect of cross entropy. But while for the derived words the effect of cross entropy was inhibitory, it is facilitatory for the base words.

Before discussing this unexpected change in sign, we first inquire whether facilitation for cross entropy also characterizes the set of simple base words. Table 6.10 lists the partial effects of the predictors that were retained after stepwise variable elimination. Figure 6.7 visualizes these partial effects. The upper left panel shows the effect of orthographic length, which shows a clear minimum near the median length (5 letters) for visual lexical decision but not for word naming. For the latter task, the shorter the word, the easier it is to articulate. For the former task, 5-letter words emerge as most easily read. The upper right panel shows that, as for the derived words, spoken frequency allows greater facilitation for visual lexical decision than for word naming.

The lower left panel presents the expected facilitatory effect of the Synset Count, and illustrates that words with more meanings elicit shorter latencies, for both word naming and lexical decision. Surprisingly, the lower central panel shows that the

	Estimate	Lower	Upper	P
Intercept	6.8433	6.7756	6.9097	0.0001
Experiment=naming	-0.2520	-0.3213	-0.1885	0.0001
Length (linear)	-0.0613	-0.0797	-0.0430	0.0001
Length (quadratic)	0.0067	0.0052	0.0080	0.0001
Spoken Frequency	-0.0251	-0.0286	-0.0216	0.0001
Family Size	0.0107	0.0021	0.0193	0.0158
Word Frequency	-0.0090	-0.0125	-0.0054	0.0001
cross entropy	-0.1316	-0.1823	-0.0869	0.0001
Synset Count	-0.0235	-0.0321	-0.0154	0.0001
Experiment=naming:Length (linear)	0.0507	0.0305	0.0722	0.0001
Experiment=naming:Length (quadratic)	-0.0034	-0.0050	-0.0018	0.0002
Experiment=naming:Spoken Frequency	0.0173	0.0141	0.0202	0.0001

Table 6.10: Partial effects of the predictors for word naming and visual lexical decision latencies for simple base words. Lower, Upper: 95% highest posterior density interval; P: Markov chain Monte Carlo p-value.

partial effect of Family Size is inhibitory, instead of facilitatory, as reported for previous experiments. We return to this finding below. The partial effect of cross entropy is presented in the lower right panel of Figure 6.7. As for the complex base words, the effect of cross entropy for simple base words is again facilitatory.

The analyses of the two sets of base words leave us with two questions. First, how should we understand the change in sign of the cross entropy effect between derived words and base words? Second, why do we have inhibition from the morphological family size for simple base words, and no effect of family size for complex base words?

With respect to the first question, we note that there is bottom-up support for only the base word, and no such support for their derivatives. By contrast, in the case of the derived words, there is bottom-up support for the derived word itself, its base word, and its affix. In sum, for derived words, three of the four elements in a proportional analogy such as

$$\underbrace{\textit{great} : \textit{greatest}}_{\text{mini paradigm}} = \underbrace{\text{A} : \textit{-est}}_{\text{mini class}} \tag{6.25}$$

are actually present in the signal. For derived words, we can therefore understand the effect of cross entropy as reflecting the cost of resolving the proportional analogy between mini-paradigm and mini-class. More specifically, the cross entropy

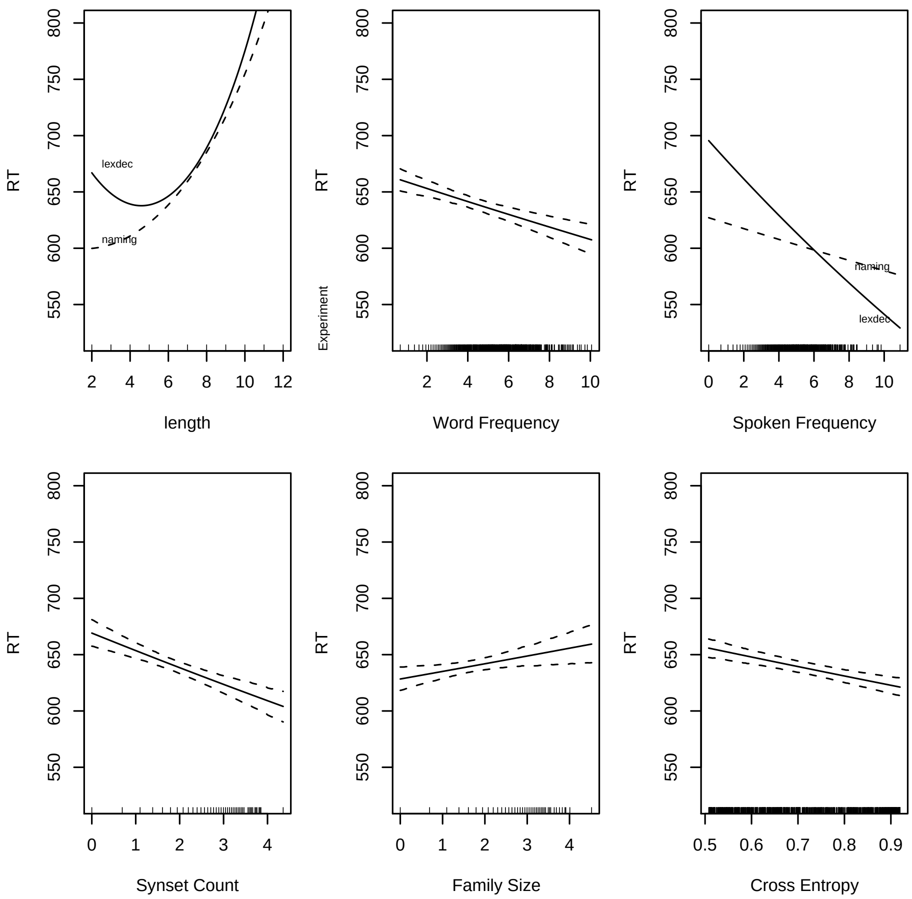


Figure 6.7: Partial effects of the predictors for word naming and visual lexical decision latencies for simple base words. Markov chain Monte Carlo based 95% confidence intervals are shown for those predictors that do not enter into interactions.

reflects the average complexity of identifying the derived word in its mini-paradigm on the basis of the generalized probability distribution of the mini-class. Thus, the cross entropy can be understood as reflecting the cost of resolving the ambiguity in the visual input with the help of generalized knowledge in long-term memory about the corresponding mini-class. From this perspective, the inhibitory effect of cross entropy for derived words makes perfect sense: The higher the cross entropy, the more information has to be retrieved from memory to resolve the proportional analogy.

Let us now consider the facilitatory effect of cross entropy for simple base words. For simple base words, the visual input is unambiguous, with bottom-up support only for the word itself. There is no cost of a call on proportional analogy to resolve morphological ambiguity. In the absence of a morphological parsing problem, the cross entropy effect apparently reverses and emerges as a measure of the amount of support the base receives from related derived words co-activated by the base. Crucially, it is not simply the count of related derived words (we checked that this count is not predictive for the present data) but rather the analogical support for the base given its derivative (defined in the mini-paradigm) and the general likelihood of a base word having derivatives (defined in the mini-class).

The second question to be considered is why we observe inhibition from the morphological family size for simple base words, and no effect of family size for complex base words. The unexpected inhibitory effect of family size is probably due to what is known in the statistical literature as suppression (see, e.g., Friedman & Wall, 2005): When predictor variables are correlated, and both are correlated with the dependent variable, then, depending on the strength of the former correlation, the beta coefficient of one of the predictors can become non-significant or even change sign. Table 6.11 presents the correlation matrix for key predictors, and reveals a large positive coefficient for the correlation of Family Size and the Synset Count, and the expected negative correlations for Family Size and response latencies in lexical decision and naming. This by itself is a warning that suppression might be at issue here.

We therefore inspected whether Family Size was significant in a model for the simple base words, excluding the Synset Count as predictor. It was not ($p > 0.8$). When cross entropy was also removed as predictor, the Family Size measure emerged as significant ($p < 0.01$), now with a negative slope, as expected given previous studies. For the complex base words, excluding only the Synset measure was sufficient to allow a facilitatory effect of Family Size to emerge. What this

	Frequency	Family Size	Synset Count	cross entropy	RT lexdec	RT naming
Frequency	1.000	0.320	0.345	-0.527	-0.379	-0.266
Family Size	0.320	1.000	0.643	0.245	-0.473	-0.392
Synset Count	0.345	0.643	1.000	0.092	-0.552	-0.434
cross entropy	-0.527	0.245	0.092	1.000	-0.085	-0.101
RT lexical decision	-0.379	-0.473	-0.552	-0.085	1.000	0.648
RT naming	-0.266	-0.392	-0.434	-0.101	0.648	1.000

Table 6.11: Pairwise correlations between key predictors and lexical decision (lexdec) and naming latencies for the set of simple base words.

suggests is that the Family Size effect, which has always been understood as a semantic effect (see, e.g., Moscoso del Prado Martín *et al.*, 2004; Schreuder & Baayen, 1997), is a composite effect that bundles effects of semantic similarity and effects of paradigmatic structure. Effects of similarity would then be better captured by means of the Synset Count, and effects of derivational paradigmatic structure would then be better captured by means of the cross entropy measure.

The question that arises at this point is whether the semantic aspect of the Family Size effect has any specific morphological component. To answer this question, we first partioned the Synset Count into two disjunct counts, a count for morphologically related synsets, and a count for morphologically unrelated synsets. A morphologically related synset is a synset in which at least one of the synset members is morphologically related to the target word (not counting the target word itself). A morphologically related synset, therefore, is a family size count that only includes semantically highly related family members.

In the model for the simple base words, we then replaced the Family Size measure and the Synset Count by the counts of morphologically related and unrelated synset counts. A mixed-effects analysis revealed that, for visual lexical decision, both counts were significant predictors with very similar coefficients (-0.018 and -0.015 respectively). For the naming latencies, however, only the synset count of morphologically unrelated synsets was significant. This interaction ($p = 0.0049$) shows that in a task such as word naming, which does not require deep semantic processing, semantic ambiguity that arises through morphological connectivity does not play a role. By contrast, the lexical decision task, which invites deeper semantic processing, allows the effect of morphologically related words that are also very similar in meaning to become visible. We therefore conclude that morphologically related words that are also semantically very similar have a special status compared to semantically similar but morphologically unrelated words (see also Moscoso del Prado Martín *et al.*, 2004).

Concluding remarks

In the preceding sections we reviewed and presented a range of studies addressing specific aspects of the complexities of paradigmatic structure in lexical processing. In order to obtain a model for the full complexity for an inflected variant w_e, we combine equations (6.10), (6.14), and (6.15) and add the effects of the entropy and relative entropy measures, leading to the following equation:

$$\begin{aligned} I \propto \beta_0 \quad &+ \quad \beta_1 \log_2 \Pr_N(w_e) + \beta_2 \log_2 \Pr_N(w) + \\ &+ \quad \beta_3 \log_2 \left(\frac{\Pr_\pi(e)/R_e}{\sum_e \Pr_\pi(e)/R_e} \right) + \\ &+ \quad \beta_4 \log_2 \left(\frac{\Pr_\pi(w_e)/R_e}{\sum_e \Pr_\pi(w_e)/R_e} \right) + \\ &+ \quad \beta_5 H_d + \\ &+ \quad \beta_6 H_i + \beta_7 RE. \end{aligned} \tag{6.26}$$

Large regression studies are called for to bring all these variables into play simultaneously. However, even though (6.26) is far from simple, it is only a first step towards quantifying the complexities of inflectional processing. We mention here only a few of the issues that should be considered for a more comprehensive model.

First, Kostić, Marković and Baucal (2003) calculated the number of functions and meanings R_e of exponent e conditionally on a lexeme's inflectional class. For instance, the number of functions and meanings listed for the exponent *a* for masculine nouns in Table 6.2, 109, is the sum of the numbers of functions and meanings for masculine genitive and the masculine accusative singular. This provides a lower bound for the actual ambiguity of the exponent, as the same exponent is found for nominative singulars and genitive plurals for regular feminine nouns. The justification for conditioning on inflectional class is that the stem to which an exponent attaches arguably provides information about its inflectional class. This reduces the uncertainty about the functions and meanings of an exponent to the uncertainty in its own class. Nevertheless, it seems likely that an exponent that is unique to one inflectional class (e.g., Serbian *ama* for regular feminine nouns) is easier to process than an exponent that occurs across all inflectional classes (e.g., *a, u*), especially when experimental items are not blocked by inflectional class. (Further complications that should be considered are the consequences of, for instance, masculine nouns (e.g., “sudija” (*judge*), “sluga” (*servant*)) taking the same inflectional exponents as regular feminine nouns do, and of animate masculine nouns being associated with a pattern of exponents that differs from that associated with inanimate masculine nouns.)

Second, the standard organization of exponents by number and case has not played a role in the studies that we discussed. Thus far, preliminary analyses of the experimental data available to us have not revealed an independent predictive role for case, over and above the attested role of ambiguity with respect to numbers of functions and meanings. This is certainly an issue that requires further empirical investigation, as organization by case provides insight into the way that functions and meanings are bundled across inflectional classes.

Third, we have not considered generalizations across, for instance, irregular and regular feminine nouns in Serbian, along the lines of Clahsen, Hadler, Eisenbeiss and Sonnenstuhl-Henning (2001). The extent to which inflected forms inherit higher-order generalizations about their phonological form provides further constraints on lexical processing.

Fourth, the size of inflectional paradigms has not been investigated systematically. Although the nominal inflectional classes of Serbian are an enormous step forward compared to the nominal paradigms of English or Dutch, the complexities of verbal paradigms can be much larger. From an information-theoretic perspective, the entropy of the complex verbal paradigms of Serbian must be much larger than the entropy of nominal paradigms, and one would expect this difference to be reflected in elongated processing latencies for inflected verbs. The study by Traficante and Burani (2003) provides evidence supporting this prediction. They observed that inflected verbs in Italian elicited longer processing latencies than inflected adjectives.

Nevertheless, it should be noted that the question of what constitutes a verbal paradigm is still open. In one, traditional, sense each verb may have not one, but several paradigms defined over various tenses and aspects. In the other sense, verbs have one exhaustive paradigm that encompasses all verbal inflected variants. Baayen, Wurm and Aycock (2008) have addressed a similar question for the paradigms of English nouns and they concluded that lexemes and their inflected variants should not be considered together as a single paradigm. In a similar way, we can tackle the question of verbal paradigmatic organization in the mental lexicon

– using information theory and large-scale regression modelling. Two alternatives can be tested empirically and the result should be straightforwardly in favor of either (a) a single entropy measure calculated over all verbal inflected variants or (b) entropies within each tense and aspect, and one computed over all tenses and aspects.

Fifth, all results reported here are based on visual comprehension tasks (lexical decision, word naming). Some of the present results are bound to change as this line of research is extended to other tasks and across modalities. For instance, the effect of inflectional entropy reported by Baayen, Feldman and Schreuder (2006) for visual lexical decision and word naming was facilitatory in nature. However, in a production study by Bien (2007), inflectional entropy was inhibitory (see also Baayen, Levelt, Schreuder & Ernestus, 2008). In lexical decision, a complex paradigm is an index of higher lexicality, and may therefore elicit shorter response latencies. In production, however, the paradigm has to be accessed, and a specific word form has to be extracted from the paradigm. This may explain why, in production, a greater paradigm complexity appears to go hand in hand with increasing processing costs. Generally, it will be important to establish paradigmatic effects for lexical processing in natural discourse using tasks that do not, or only minimally, impose their own constraints on processing.

Sixth, it will be equally important to obtain distributional lexical measures that are more sensitive to contextual variation than the abstract frequency counts and theoretical concepts of functions and meanings that have been used thus far. Interestingly, Moscoso del Prado Martín, Kostić and Filipović Đurđević (2008) and Filipović Đurđević (2007) report excellent predictivity for lexical processing of more complex information theoretic measures of morphological and semantic connectivity derived bottom-up from a corpus of Serbian.

It is clear that the information theoretic measures that we have proposed and illustrated in this chapter capture only part of the multidimensional complexity of lexical processing. Hence, each measure can be undersood as a plane cross-cutting this multidimensional space. In spite of these limitations, the extent to which the present information-theoretic approach converges with WPM is striking. Across our experimental data sets we find evidence for exemplars, irrespective of whether the language under investigation is Dutch, English, or Serbian. At the same time, we observe the predictivity of entropy measures, which generalize across probability distributions tied to subsets of these exemplars, and evaluate the complexity of paradigms and the divergence between different levels of morphological organization. However, all the results discussed here pertain to the processing of familiar words. In order to properly gauge the processing complexity of new inflected and derived words, it will be necessary to combine WPM and the present information theoretic approach with computational models of language processing.

Such an integration is especially challenging because across computational models of linguistic generalization, whether abstractionist and implementing greedy learning (Albright & Hayes, 2003), or memory-based and implementing lazy learning (Daelemans & Van den Bosch, 2005; Keuleers, 2008; Keuleers, Sandra, Daelemans *et al.*, 2007), a common finding is that it is type frequencies and not token frequencies on which generalization is based. In fact, type-based generalization has been found to be reflected in processing measures as well (see, e.g., Ernestus, Baayen & Ling, 2004; Krott, Hagoort & Baayen, 2004). Typically, current computational models (cf., Albright, 2008) make use of much more sophisticated analogies than the traditional four-part analogy that we have referred to as a possible explanation for the effect of cross-entropy.

To resolve this paradox, we note, first of all, that our hypothesis is not a hypothesis about the choice of a linguistic form, but rather a measure of the cost of selecting a given complex word from its mini-paradigm given its mini-class. Furthermore, note that for most of the derivational suffixes we have considered, there are no rival suffixes comparable to the rivalling options that characterize the past tense in English (Albright & Hayes, 2003), or plural selection in Dutch (Keuleers, Sandra, Daelemans *et al.*, 2007). There is only one way in English to express the comparative, the superlative, or adverbs through suffixation. Hence, the probability of the selection of *-er, -est* or *-ly* is equal to one. For this 'degenerate' case, four-part analogy provides a reasonable model. In fact, we think it is precisely this uniformity in the analogical support for a given suffix that allows us to see the effect of cross-entropy. Because there are no competing sets of exemplars supporting different outcomes, there are no overriding type frequency effects. As a consequence, the more subtle relevance of the token counts becomes visible only for the basic, type-uniform four-part analogy. The real challenge for future research, therefore, is to clarify whether subtle effects of token frequencies also codetermine the fine details of lexical processing when more complex, type-frequency driven analogies come into play.

References

Albright, A. (2008). Modeling analogy as probabilistic grammar. In Blevins, J. and Blevins, J., editors, *Analogy in Grammar*, this volume. OUP, Oxford.

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

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

Aronoff, M. (1994). *Morphology by itself: stems and inflectional classes*. The MIT Press, Cambridge, Mass.

Baayen, R. (2003). Probabilistic approaches to morphology. In Bod, R., Hay, J., and Jannedy, S., editors, *Probability theory in linguistics*, pages 229–287. The MIT Press.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge University Press, Cambridge.

Baayen, R. H., Davidson, D. J., and Bates, D. (2008a). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, in press.

Baayen, R. H., Feldman, L., and Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 53:496–512.

Baayen, R. H., Levelt, W., Schreuder, R., and Ernestus, M. (2008b). Paradigmatic structure in speech production. *Chicago Linguistics Society*, 43, in press.

Baayen, R. H., McQueen, J., Dijkstra, T., and Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In Baayen, R. H. and Schreuder, R., editors, *Morphological structure in language processing*, pages 355–390. Mouton de Gruyter, Berlin.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX lexical database (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Baayen, R. H., Wurm, L. H., and Aycock, J. (2008c). Lexical dynamics for low-frequency complex words. a regression study across tasks and modalities. *The Mental Lexicon*, 2:419–463.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39(3):445–459.

Bates, D. (2006). Linear mixed model implementation in lme4. Department of Statistics, University of Wisconsin-Madison.

Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, 5:27–30.

Beard, R. (1995). *Lexeme-morpheme base morphology: a general theory of inflection and word formation.* State University of New York Press, Albany, NY.

Beckwith, R., Fellbaum, C., Gross, D., and Miller, G. (1991). WordNet: A lexical database organized on psycholinguistic principles. In Zernik, U., editor, *Lexical Acquisition. Exploiting On-Line Resources to Build a Lexicon*, pages 211–232. Lawrence Erlbaum Associates, Hillsdale, NJ.

Bien, H. (2007). *On the production of morphologically complex words with special attention to effects of frequency*. Max Planck Institute for Psycholinguistics, Nijmegen.

Blevins, J. P. (2003). Stems and paradigms. *Language*, 79:737–767.

Blevins, J. P. (2006). English inflection and derivation. In Aarts, B. and McMahon, A. M., editors, *Handbook of English Linguistics*, pages 507–536. Blackwell, London.

Burnard, L. (1995). *Users guide for the British National Corpus*. British National Corpus consortium, Oxford university computing service.

Clahsen, H., Hadler, M., Eisenbeiss, S., and Sonnenstuhl-Henning, I. (2001). Morphological paradigms in language processing and language disorders. *Transactions of the Philological Society*, 99(2):247–277.

Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. John Wiley & Sons, New York.

Crawley, M. J. (2002). *Statistical computing. An introduction to data analysis using S-plus*. Wiley, Chichester.

Daelemans, W. and Van den Bosch, A. (2005). *Memory-based language processing*. Cambridge University Press, Cambridge.

Ernestus, M. and Baayen, R. H. (2004). Analogical effects in regular past tense production in Dutch. *Linguistics*, 42:873–903.

Filipović Đurđević, D. (2007). *The polysemy effect in processing of Serbian nouns*. PhD thesis, University of Belgrade, Serbia.

Friedman, L. and Wall, M. (2005). Graphical views of suppression and multicollinearity in multiple regression. *The American Statistician*, 59:127–136.

Gagné, C. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27:236–254.

Gagné, C. and Shoben, E. J. (1997). The influence of thematic relations on the comprehension of modifier-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23:71–87.

Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In Hale, K. and Keyser, S. J., editors, *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, volume 24 of *Current Studies in Linguistics*, pages 111–176. MIT Press, Cambridge, Mass.

Hockett, C. (1954). Two models of grammatical description. *Word*, 10:210–231.

Keuleers, E. (2008). *Memory-based learning of inflectional morphology*. University of Antwerp, Antwerp.

Keuleers, E., Sandra, D., Daelemans, W., Gillis, S., Durieux, G., and Martens, E. (2007). Dutch plural inflection: The exception that proves the analogy. *Cognitive Psychology*, 54:283–318.

Kostić, A. (1991). Informational approach to processing inflected morphology: Standard data reconsidered. *Psychological Research*, 53(1):62–70.

Kostić, A. (1995). Informational load constraints on processing inflected morphology. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*. Lawrence Erlbaum Inc. Publishers, New Jersey.

Kostić, A. (2008). The effect of the amount of information on language processing. *Manuscript submitted for publication*.

Kostić, A., Marković, T., and Baucal, A. (2003). Inflectional morphology and word meaning: Orthogonal or co-implicative domains? In Baayen, R. H. and Schreuder, R., editors, *Morphological Structure in Language Processing*, pages 1–44. Mouton de Gruyter, Berlin.

Krott, A., Hagoort, P., and Baayen, R. H. (2004). Sublexical units and supralexical combinatorics in the processing of interfixed Dutch compounds. *Language and Cognitive Processes*, 19:453–471.

Kuperman, V., Bertram, R., and Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, in press.

Landauer, T. and Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22:1–38.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177.

Luce, R. D. (1959). *Individual choice behavior*. Wiley, New York.

Manning, C. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge.

Matthews, P. H. (1974). *Morphology. An Introduction to the Theory of Word Structure*. Cambridge University Press, London.

Milin, P., Filipović Đurđević, D., and Moscoso del Prado Martín, F. (2008). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Manuscript submitted for publication*.

Miller, G. A. (1990). Wordnet: An on-line lexical database. *International Journal of Lexicography*, 3:235–312.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., and Baayen, R. H. (2004a). Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30:1271–1278.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (2004b). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94:1–18.

Moscoso del Prado Martín, F., Kostić, A., and Filipović Đurđević, D. (2008). The missing link between morphemic assemblies and behavioral responses: a Bayesian Information-Theoretical model of lexical processing. *Manuscript submitted for publication*.

New, B., Brysbaert, M., Segui, F. L., and Rastle, K. (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, 51:568–585.

Pinker, S. (1991). Rules of language. *Science*, 153:530–535.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Weidenfeld and Nicolson, London.

Ross, S. M. (1988). *A First Course in Probability*. Macmillan Publishing Company, New York.

Schreuder, R. and Baayen, R. H. (1995). Modeling morphological processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 131–154. Lawrence Erlbaum, Hillsdale, New Jersey.

Schreuder, R. and Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, 37:118–139.

Stemberger, J. P. and MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition*, 14:17–26.

Traficante, D. and Burani, C. (2003). Visual processing of Italian verbs and adjectives: the role of the inflectional family size. In Baayen, R. H. and Schreuder, R., editors, *Morphological structure in language processing*, pages 45–64. Mouton de Gruyter, Berlin.

Summary and Conclusions

Chapter 7

This dissertation explored the role of morphological structure in the comprehension and production of polymorphemic words (i.e., words with two or more morphemes, e.g., *dish-wash-er*) in Dutch, English, Finnish and Serbian. The primary research question was how people use the probabilistic information carried by morphemes, morphological paradigms (i.e., sets of words that share a morpheme), and complex words as wholes. We tackled this question by investigating (i) the time-course of activation of morphemes and whole words in silent reading and speech production, (ii) the effect of morphemes hierarchically and orthographically embedded in larger structural blocks (e.g., *wash-* and *-er* in *washer*) on visual comprehension and acoustic production of polymorphemic words, (iii) the role of morphological paradigms (e.g., *washbasin*, *washroom*, *washcloth*) in the lexical processing of inflected, derived and compound words, and (iv) the interactions between morphological and other linguistic predictors as co-determinants of the costs of lexical processing for complex words.

In a series of experiments on silent reading of polymorphemic words, as well as in the acoustic analyses of such words, we found evidence that the lexical processing of complex words involves a larger pool of information sources than previously thought. Our experimental findings on reading of compounds and derived words enabled us to formulate a probabilistic model of visual morphological processing based on concepts and tools of information theory. In this chapter, we review the results of our experiments, as well as the main aspects of our probabilistic model, and we outline topics that call for further research.

The time-course of morphological processing

The eye-tracking experiments reported in Chapters 3 and 4 investigated the time-course of morphological effects on the speed of reading of Dutch compounds presented as isolated words (Chapter 3), and of Finnish compounds presented in sentential contexts (Chapter 4). In both experiments we used relatively long compounds (in the range of 8-12 characters in Chapter 3 and 10-18 characters in Chapter 4). Due to visual acuity constraints, our eyes generally cannot process words of this length range without fixating two or more times on the target words. Typically, at the first fixation readers obtain a sharp, foveal view of the compound's left constituent (e.g., *dish*), while subsequent fixations allow for the foveal inspection of the right constituent (e.g., *washer*). This gradual visual uptake for long words makes it easier to determine the relative order of activation of morphemes versus whole words in the eye-movement record, as compared to the cases where morphological effects emerge jointly in a single fixation or in a lexical decision latency. The eye-tracking study in Chapter 5 further explored morphological processing in sentential reading of Dutch derived words (*succes-vol* "successful"): Unlike compounds, these derived words were mostly read in a single fixation, and yet they shed light on the temporal unfolding of activation of morphological structure.

The three experiments in Chapters 3-5 showed a robust, similar, temporal pattern of morphological effects, which is remarkable given the cross-experimental differences in languages (Dutch vs. Finnish), length ranges of target words, experimental tasks (lexical decision on isolated words in Chapter 3 vs. reading of sentences in Chapters 4 and 5), and the type of morphological word-formation (compounding vs. derivation). We found that:

- The order of activation of compounds' morphological constituents closely follows the typical progress of the visual uptake, from left to right (Chapters 3 and 4): Morphological properties (frequency and family size) of the compound's left constituents showed earlier, stronger and longer lasting effects on eye-movement measures than the properties of the right constituents. The head of the compound (i.e., its right constituent) is generally semantically closer to the compound as a whole (cf., *washer* in *dishwasher*), and it has been argued that it may therefore function as the key for lexical access to the compound (e.g., Juhasz *et al.*, 2003). However, the dominant role of the compound's left constituent that we observed does not support this suggested crucial role for the right constituent (cf., Juhasz, 2007). As a result of its later availability for the visual system, identification of a compound's right constituent may proceed against the backdrop of existing knowledge gleaned from the left constituent. These results make a strong case for explicitly taking into consideration the order of visual uptake in models of morphological processing, which so far have treated complex words as if they are immediately and entirely available to the visual system (e.g., Baayen & Schreuder, 2000; Giraudo & Grainger, 2001; Taft & Forster, 1975, 1976; for an exception, see Pollatsek *et al.*, 2003).

We found further evidence for the left-to-right order of activation of morphemes in the silent reading of Dutch derived words (Chapter 5). The vast majority of those words (*succes-vol* "successful") are read in just one fixation, so that both the left morphological constituent (base word, *succes* "success") and the right constituent (suffix, *-vol* "-ful") are simultaneously available to the visual system. However, one of the morphological predictors of reading times, i.e., the relative entropy of the morphological families of the word's constituents, was sensitive to whether we assume the left-to-right order of morphemic activation or the right-to-left order. The assumption of the left-to-right order, even given the simultaneous visual availability of both the left and the right word's constituents, led to a statistical model that fit the observed data significantly better than a model with as a predictor the alternative measure of relative entropy based on the putative right-to-left activation order.

- The effect of whole word frequency is a hallmark of full-form lexical access. In the two experiments on long compounds (Chapters 3 and 4), we observed an effect of compound (whole word) frequency on reading times as early as during the first fixation, simultaneously with the effect of the left constituent frequency and family size. Given the lengths of our compounds, it is likely that the early effect of compound frequency precedes the complete identification of all characters and of the right constituents of our long compounds. This effect suggests that readers make inferences about the compound's identity as soon as they have available any (potentially incomplete) information about the word, such as the word's length, the identity of the left constituent or the preview of the right constituent. The early compound frequency effect is problematic for those sublexical models of word processing that require both the left and the right constituent to be activated prior to activation of the whole compound at the lemma level (e.g., Pinker, 1999; Taft & Ardasinski, 2006). The fact that the compound frequency effect is simultaneous with the effect of the left constituent frequency and family size challenges supralexical models, since they predict activation of morphological constituents to occur after, and not in parallel with, activation of the full-form. This finding strongly supports the parallel processing of full-forms and morphemes, advocated by dual- and multiple-route models of morphological processing.

- The effect of compound frequency lingers on throughout the entire course of reading a compound, overlapping first with the effects associated with the compound's left constituent and then with those associated with its right constituent. This continuous involvement of the full-form is noticeable even for most compounds that are likely candidates for morphological decomposition (due to their high constituent frequencies or large constituent families). Since both the full-form and separate morphemes contribute to word recognition, our findings are not easily reconcileable with the "winner-takes-it-all" principle implemented in the dual-route model of Frauenfelder and Schreuder (1991), Schreuder and Baayen (1995).

- Morphological constituent families of compounds and derived words (defined as sets of words sharing a given constituent, *ice pick*, *ice cube*, *ice cream*) are activated as soon as those constituents become available to the visual system (Chapters 3-5). This finding contrasts with the received view that the effect of morphological families develops late in the identification process of the complex word and is due to activation spreading through morphological paradigms and to semantic resonance of family members, which are related in form and (often) meaning with the word being recognized (cf., De Jong, Schreuder & Baayen, 2000). In Chapters 3, 4 and 6 we refine the current knowledge on the role of families by making the case that, even though a semantic component is intrinsic in the effects of the family, it is substantially complemented by the morphological characteristics of the paradigm members.

While Chapters 3-5 addressed the temporal unfolding of the effects that morphological structure shows in eye-movements during reading, the study in Chapter 2 shed light on the time-course of morphological effects in speech production. We explored the acoustic duration of interfixes *-s-* and *-e(n)-* in Dutch polymorphemic compounds (e.g., *dier-en-arts* "veterinary") as a function of the probabilistic bias towards those interfixes stemming from the morphological families of compounds' constituents. One of our findings was that the amount of information in the morphological family of the right constituent, i.e., the right positional entropy, codetermines the duration of the interfix *before* the actual articulation of that right constituent takes place: The higher the entropy in the right constituent family, the longer the acoustic realization of the interfix is. We interpret this effect as evidence that (i) morphological families of upcoming morphemes (e.g., compounds' right constituents) are activated in the process of planning the articulation of those morphemes, and (ii) the amount of information in the paradigm of the morpheme under planning interferes with the articulation of preceding morphemes (i.e., interfixes), see also Pluymaekers *et al.* (2005).

Taken together, the findings outlined in this section indicate that current models of morphological processing in speech production (e.g., Levelt, Roelofs & Meyer, 1999) and visual perception (e.g., Baayen & Schreuder, 1999; Giraudo & Grainger, 2001; Taft, 1991) are too restrictive in their architectures to account for the empirically established time-course of morphological effects.

Morphological paradigms

The experiments described in this dissertation do not only contribute to our knowledge of the time-course of morphological processing, but also reveal previously unknown aspects of the paradigmatic organization of the mental lexicon and its relevance for the recognition and production of compound words (Chapters 2-4), derived words (Chapters 5 and 6), and inflected words (Chapter 6). The summary of our findings is as follows:

- In all visual recognition studies in this dissertation (Chapters 3-6), characteristics pertaining to families of morphological constituents (family size or family frequency) invariably showed stronger effects on the processing speed of complex words than the frequencies of occurrence for constituents as isolated words. This suggests that recognition of morphemes in complex words proceeds predominantly against their morphological context and does not hinge on retrieving the relevant morphemes as independent, context-free units from the lexical long-term memory (mental lexicon).

- Morphological families of constituents are generally considered as a source of paradigmatic support for the recognition of those constituents (cf., De Jong, Schreuder & Baayen, 2003). Hence morphological families that have more members, or more frequent members, or a larger amount of information (measured as the entropy of the family) are argued to facilitate decomposition of complex words into morphemes during comprehension (cf., e.g., De Jong, Schreuder & Baayen, 2000; Moscoso del Prado Martín, Bertram, Häikiö *et al.*, 2004; Moscoso del Prado Martín, Kostić & Baayen, 2004). In Chapter 5, however, we found that the impact of families on parsing was more intricate. Specifically, we observed that the processing of Dutch derived words (e.g., *succesvol*, "successful") was optimal when the morphological family of the base (e.g., *succes*, "success") and that of the suffix (e.g., *-vol* "-ful") were equally small or equally large. The more the two families diverged in size, the more inflated the processing costs were, as estimated by eye-movement measures of reading times. Apparently, this effect (which we modelled using the information-theoretical measure of relative entropy) bears witness to some form of competition between morphological families. Imbalance in the amount of lexical support for the morphemes may delay the integration of the two morphemes into a coherent representation of the derived word as a whole.

- Morphological families as sets of complex words with a shared constituent only reflect one possible type of paradigmatic organization in the mental lexicon. In our lexical decision study of English derived words (Chapter 6) we tested the relevance of a different type of organization based on earlier information-theoretical studies of inflected words (extensively reviewed in Chapter 6), namely, mini-paradigms (pairs of base words and their derivatives, e.g., *kind* and *unkind*) and mini-classes (the set of mini-paradigms sharing the same derivational affix: e.g., *kind* - *unkind*, *true* - *untrue*, *pleasant* - *unpleasant*, etc.). We found that the more the mini-paradigm diverges in its probability distribution from the probability distribution defined over the entire mini-class (as quantified by the information-theoretical measure of cross entropy), the longer the lexical decision and the naming latencies for the derived word in the mini-paradigm, and the shorter those latencies for the base word. This study offers new evidence that lexical processing of derived words is not only sensitive to the amount of information in its complete morphological paradigm. It is also reflects the probabilistic relationships between different levels of paradigmatic organization, from the micro-level of mini-paradigms to the macro-level of mini-classes.

While Chapters 3-6 showed abundant and novel evidence for the role of morphological paradigms in codetermining processing costs of visual word recognition, our production study (Chapter 2) examined whether distributional characteristics of those paradigms can affect the way morphemes are realized in speech. Recent studies showed that the amount of probabilistic bias towards an interfix in a Dutch compound is codetermined in a given compound by the choice of the interfix in the words that share the left constituent of that compound (e.g., *kandidaat-s-examen* "bachelor's examination", *kandidat-en-lijst* "list of candidates", and *kandidaat-stelling* "nomination", cf., Krott, Baayen & Schreuder, 2001). We observed that the probabilistic bias towards an interfix correlated with the acoustic duration of that interfix, such that the more biased (predictable) the interfix was, the longer its duration was. This finding presented an apparent paradox for an influential class of speech production theories that postulates a *negative* correlation between the probability of a speech unit and the amount of articulatory effort (e.g., the acoustic duration) realized in the production of that unit (cf., Jurafsky, Bell, Gregory & Raymond, 2001; Aylett & Turk, 2004; 2006; Van Son & Van Santen, 2005). We explain this intriguing result by distinguishing between predictability from a syntagmatic perspective, which is negatively correlated with acoustic salience, and the amount of paradigmatic support for one of a small number of alternatives, which appears to be positively correlated with acoustic salience. This account, which we labeled the Paradigmatic Signal Enhancement Hypothesis, makes testable predictions about the acoustic realizations of other lingusitic units that have paradigmatic alternatives.

Considered together, our results suggest that paradigmatic organization of the mental lexicon plays a stronger and a more complex role in the lexical processing of complex words that previously assumed.

Morphemes at lower hierarchical levels

The current psycholinguistic literature does not offer much data on the lexical processing of complex words with three or more morphemes (see for exceptions, e.g., De Almeida & Libben, 2005; Inhoff, Radach & Heller, 2000; Krott, Baayen & Schreuder, 2001; Krott, Hagoort & Baayen, 2004). Yet such words are interesting in that they show a multi-level morphological structure, from the whole word (e.g., *dishwasher*) to its immediate morphological constituents (e.g., *dish* and *washer*) to morphemes (deeply) embedded in those morphological constituents (e.g., *wash-* and *-er*). In Chapters 2, 3 and 4 we explored polymorphemic compounds in Dutch and Finnish to establish the role of morphemes deeply embedded in morphological structure in visual word recognition and speech production. Results reported in Chapters 3 and 4 clearly show that also morphemes at low levels of the morphological hierarchy can be recognized and used in compound identification as independent units of meaning, rather than being treated as an unanalyzable part of the character string (e.g., *washer*). We found that:

- more productive affixes embedded in compounds' constituents (e.g., *-ing* in *plaatsingsbeleid* "placing policy") elicited shorter reading times, just like productive affixes in bimorphemic derivations;

- relatively salient (e.g., longer, more frequent and structurally invariant) derivational suffixes embedded in trimorphemic compounds (e.g., the suffix *-sto* in the Finnish compound *kirjastokortti* "library card") serve as better parsing cues for segmentation of the compounds into their immediate constituents (*kirjasto* "library" and *kortti* "card") and thus facilitate lexical processing.

- compounds embedded in trimorphemic compounds (e.g., *zaal+voet-bal* "indoor football") come with longer reading times than derivations embedded in compounds (e.g., *plaatsingsbeleid* "placing policy"), which we interpreted as an indication of increased costs of semantic integration for words with three, rather than two, free-standing lexemes;

Moreover, the results of our production study (Chapter 2) outlined above further demonstrate that the processing of morphemes deeply embedded in the morphological structure (*-s-* in *oorlog-s-verklaring* "declaration of war") is sensitive to the amount of information carried by that morpheme, and – crucially – also by the informativeness of the morphemes that are realized prior to (e.g., *oorlog*) or after (e.g., *verklaring*) the deeply embedded morpheme.

These findings add granularity to our knowledge of morphological processing. We interpret them as compelling and novel evidence that morphemes structurally and orthographically embedded in larger morphological structures participate in the process of complex word recognition, with the likelihood of their activation being codetermined by the lexical-distributional and orthographic salience of those morphemes. These findings challenge many current models of morphological processing with the task of accounting for embedded morphemes as information sources in their own right.

Interdependent contributions of morphological structure

In all studies reported in this dissertation we observed that the magnitude and sometimes even the presence of the effects of some morphological units depends on the magnitude of effects elicited by other morphological and orthographical predictors. This interactive use of morphological sources of information contrasts with current single route and most parallel dual route models, which tend to simplify morphological processing to activation of autonomous lexical representations that are blind to each other's activation (cf., Laudanna & Burani, 1985; Frauenfelder & Schreuder, 1991, and Schreuder & Baayen, 1995).

The most common type of interactions we found showed that in words with salient (e.g., frequent or paradigmatically supported) morphemes, the properties of the full-form are used to a lesser extent than the properties of the embedded morphemes, while the situation is reverse in complex words where full-form processing is favored over decomposition by virtue of a frequent or short full-form, or the absence of clear segmentation cues for morphological parsing. In the production study of Chapter 2, we found an interaction between the measure of the probabilistic bias for the compound's interfix and the average amount of information in the left constituent family of that compound. In Chapter 3, we found that compound frequency interacts with left constituent frequency in codetermining reading times for Dutch (Figure 3.1), so that compound frequency has the weakest effect on the compounds with higher-frequency left constituents (i.e., those for which morphological decomposition is a preferred processing route). Similarly, in Chapter 4, compound frequency was found to elicit the weakest effect in the Finnish compounds with larger left or right constituent families (Figures 4.1 and 4.2), which again make the compounds likely candidates for decompositional processing. In our study of Dutch derived words (Chapter 5), suffix length emerged as a parameter that regulated the use readers made of available processing routes, and hence it modulated the magnitude of several morphological effects. Words with extremely short and hence non-salient suffixes favored full-form lexical access and showed the strongest facilitating effect of derived word frequency on reading times, with only weak effects associated with the word's morphemes. As affixal salience increased with suffix length, the effect of derived word frequency was attenuated and virtually vanished for words with longer suffixes, while the effects associated with parsing of derived word's morphemes increased in size.

While the possibility of interdependencies in morphological processing has been considered in earlier experimental and modeling studies (e.g., Bertram & Hyönä, 2003; Baayen & Schreuder, 2000), our data confirm that such interdependencies are so common that they virtually regulate the lexical processing of complex words. We believe that they reflect trade-offs between available routes of morphological processing, including storage in long-term memory and online computation. The empirical data allow us to conclude that any model of morphological processing should explicitly account for the fact that the contribution of one information source to lexical processing modulates the contributions of other available sources.

Modeling morphological processing in visual word recognition

The empirical data of Chapters 3-6 laid the ground for the formulation of a new PRObabilistic Model of Information SourcEs (PROMISE). This model proposes a mathematical apparatus, based on the tools of information theory, to describe multiple-route parallel morphological processing in visual recognition of polymorphemic words. PROMISE builds on research by Kostić (1991; 1995) and Moscoso del Prado Martín, Kostić & Baayen (2004), who were the first to apply information-theoretical insights to explain experimental evidence on morphological processing in inflected, derived and compound words. Our model also extends the modeling framework outlined in Baayen, Wurm and Aycock (2007) by including more sources of morphological information, and implementing the interactions between those sources.

In Chapter 3, we outlined specifications for any model aiming to capture the complex pattern of morphological effects in compound processing. The specifications include:

- explicit consideration of the temporal order of information uptake, including the left-to-right order of activation of morphemes in complex words read in multiple fixations;
- absence of strict sequentiality in the processing of information, i.e., simultaneous processing of available information at different levels in the representational hierarchies, such as the full-forms of complex words (e.g., *dishwasher*), their immediate morphological constituents (*dish* and *washer*), morphemes at deeply embedded structural levels (*wash* and *-er*), and morphological paradigms of these constituents (*dishcloth*, *dish soap*, *dish rack*, etc.);
- the possibility for one processing cue (e.g., perceptual salience of a morphological constituent) to modulate the contribution of other cues (e.g., full-form properties, such as whole word frequency).

The study in Chapter 4 implemented these specifications in the PROMISE model. The conceptual framework of the model considers the mental lexicon as a long-term memory store for lexical information. An incoming visual stimulus is a key for accessing this lexical information. The accumulated knowledge of words and their paradigmatic and syntagmatic properties defines a word's information load, and hence the speed with which information about that word can be retrieved from lexical memory. PROMISE formalizes the information load of morphological structure using the perhaps most basic statement of information theory, that information (I) of a (linguistic) unit can be quantified as minus log probability (P) of that unit. As P decreases, I increases: less probable events are more informative. A fundamental assumption of our model is that the time spent by the eye on a constituent or word is proportional to the total amount of lexical information available in long-term memory for identification of that constituent or word at that timepoint (cf., Moscoso del Prado Martín, Kostić & Baayen, 2004). Events with small probability and hence a large information load require more processing resources and more processing time.

In Chapters 3-6 we identified a broad range of potential sources of morphological information, including unconditional probabilites for whole words and their constituent morphemes (i.e., likelihoods of guessing those words without further contextual information); conditional probabilities of morphemes given their position in the complex word or given that other morphemes are already available to the lexical processing system; probabilities of morphemes in their paradigms; and measures of distance between probability distributions associated with morphemes and their classes (Chapters 5, 6), operationalized by means of relative entropy and cross-entropy measures. The processing time is proportional to the sum of those amounts of information, each taken with its own weight (see General Discussion of Chapter 4 and especially equation (4.15) for the detailed elaboration of the mathematical framework of PROMISE).

The weights modulate the impact of individual information sources on the processing times, and they serve as the parameters of the model. Since PROMISE represents the information load carried by the variety of morphological sources in terms of distributional characteristics of morphological structure (e.g., word frequencies, family sizes and frequencies, etc.), parameters of PROMISE can be directly estimated from the data using multiple regression models.

PROMISE meets the requirements formulated in Chapter 3 (and itemized above) in the following way.

- The study in Chapter 4 specifies how the model can handle the temporal order of visual availability of morphological information (typically, progressing from left to right). Information sources that are available early in the time-course of the visual uptake are demonstrably more important in compound recognition (cf. the weaker role of right constituent measures as compared to properties of the left constituent). In the model equation, weights *w* for "early" information sources can be multiplied by a time-step coefficient α_1, such that $\alpha_1 > 1$. For "late" information sources, the value of α_2 is equal to or smaller than 1. As with weights w, the value of α can be directly estimated from comparing regression coefficients of a predictor in models for early measures of the visual uptake versus models for later measures.

- PROMISE takes as a basic point of departure the notion that many sources of information are evaluated jointly, e.g., whole word frequency alongside frequencies of morphemes, family-based estimates of morpheme's lexical connectivity, etc. Thus, PROMISE formalizes the abovementioned desideratum of simultaneous processing of available information at different levels.

- The statistical interactions between morphological predictors observed in Chapters 3-6 gave rise to the idea that sources of information are used interactively in complex word recognition, with the contributions of some such sources modulating the contributions of other sources to lexical processing. The PROMISE model formalizes this intuition in several ways. First, estimates of most probabilities bring in several sources with opposing weights. For instance, the probability of a word is defined as the frequency of that word divided by the corpus size. Since the informational load of a word is the minus (binary) log probability of that word, it follows that the stronger the facilitatory impact of word frequency on processing costs, the more inhibition comes from the corpus size (the greater the corpus, the greater the vocabulary richness and the greater the problem of identifying the target becomes), see General Discussion of Chapter 4 for details. This example reflects but one of many trade-offs involved in the processing of complex words. Second, in Chapter 4 we explicitly implemented the interactivity of information sources by making the coefficient associated with one information source (left constituent family frequency) propotional to the value of another cue (compound frequency), see equations (4.16) - (4.19). Finally, in Chapters 5 and 6 we used the information-theoretical tool of relative entropy to estimate how the contribution of one constituent's morphological family is modulated by the size of the other constituent's family, and we used cross-entropy to estimate how the divergence of probability distributions in a derivational mini-paradigm and in its mini-class translates into higher processing costs.

To summarize, the PROMISE model is a formalization of the idea that readers and listeners use multiple sources of information for the recognition of complex words (see Libben, 2006). Parameters of PROMISE can be pitted against the regression coefficients of statistical models that emerge from fitting the empirical data. Estimated values of parameters do not only shed light on which sources of information are preferred over others, but also specify at what timesteps of the visual uptake and at what cost to the processing system. Importantly, PROMISE allows dealing with different word formation types (e.g., derived or compound words) in a unified way.

Topics for further research

Research presented in this dissertation has shown that the study of morphological complexity yields new insights about the organization of the mental lexicon and cognitive processes involved in word production and recognition. In what follows we identify several areas of morphological inquiry that call for further research.

First, many models of morphological processing, including PROMISE, make no commitment as to how morphological activation is implemented in the brain. Yet it is crucial to know how the complex pattern of morphological effects obtained via behavioral studies (e.g,. lexical decision or eye-tracking experiments in this dissertation) reflects in the electrical, magnetic or biochemical activity of the brain. Such a link would provide a physiological basis for the hypotheses that we made with respect to the time-course of lexical processing and paradigmatic organization of the lexical long-term memory. One promising way of building such a link is in combining eye-tracking studies with experiments using event-related potentials. Further studies on acquisition of morphology in children may also shed light on how the constraints of lexical knowledge, memory and representation modulate the role of morphological structure in learning, comprehension and production of polymorphemic words.

Second, the PROMISE model can be extended and refined in several ways. For instance, PROMISE would benefit from multiple regression studies of different types of morphological complexity across languages and experimental tasks. This would allow specification of the boundary conditions and expected value ranges for the parameter space of the model. Once the parameter values have been validated against experimental replications, the model can serve as a predictive, and not only descriptive, tool for probing psycholinguistic processes involved in complex word recognition.

Third, PROMISE can be easily incorporated into general models of eye-movement control in reading, such as E-Z Reader (e.g., Reichle, Rayner & Pollatsek, 2003) or SWIFT (e.g., Engbert, Nuthmann, Richter & Kliegl, 2005). Consideration of parameters of PROMISE along with other motor-visual, orthographic and lexical parameters may improve predictions of such models for the processing of morphologically complex words.

Also, while PROMISE provided insightful results for compounding and derivation as types of word-formation, its application to inflection (based on a large-scale regression study of inflected words) is a task for the future. Finally, the challenge for the PROMISE model, developed for visual comprehension, is whether it has predictive power for auditory comprehension and also for speech production of morphologically complex words.

Concluding remarks

Most psycholinguistic research on morphological processing has been dominated by the symbolic theoretical perspective, which proposes deterministic rules for combining discrete morphemes into regular complex structures, and memory storage for irregular complex forms (cf., Pinker, 1999). An alternative, subsymbolical view holds that morphological structure is fundamentally probabilistic and emerges from the statistical regularities that characterize the mappings between forms and meanings of words (cf., Hay & Baayen, 2005). The latter view also inspires the Word and Paradigm morphology (e..g, Blevins, 2003), which assumes that words (both simple and complex) are the basic units in the lexicon and that words are organized into paradigms, which are further organized into higher-level classes. The ultimate aim of this dissertation was to improve current theories of morphological processing in visual recognition and speech production, and throughout the studies reported here I found evidence in favor of the probabilistic view of morphology. I hope that this dissertation will contribute to the empirical foundations and the theoretical advancement of this inspiring approach to research in the processing of morphologically complex words.

References

Aylett, M. and Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47:31–56.

Aylett, M. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllabic nuclei. *Journal of the Acoustic Society of America*, 119:3048–3058.

Baayen, R. H. and Schreuder, R. (1999). War and peace: morphemes and full forms in a non-interactive activation parallel dual route model. *Brain and Language*, 68:27–32.

Baayen, R. H. and Schreuder, R. (2000). Towards a psycholinguistic computational model for morphological parsing. *Philosophical Transactions of the Royal Society (Series A: Mathematical, Physical and Engineering Sciences)*, 358:1–13.

Baayen, R. H., Wurm, L. H., and Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *Mental Lexicon*, 2:419–463.

Bertram, R. and 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.

Blevins, J. P. (2003). Stems and paradigms. *Language*, 79:737–767.

de Almeida, R. and Libben, G. (2005). Changing morphological structures: The effect of sentence context on the interpretation of structurally ambiguous English trimorphemic words. *Language and Cognitive Processes*, 20:373–394.

De Jong, N. H., Schreuder, R., and Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, 15:329–365.

De Jong, N. H., Schreuder, R., and Baayen, R. H. (2003). Morphological resonance in the mental lexicon. In Baayen, R. H. and Schreuder, R., editors, *Morphological structure in language processing*, pages 65–88. Mouton de Gruyter, Berlin.

Engbert, R., Nuthmann, A., Richter, E., and Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112:777–813.

Frauenfelder, U. H. and Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In Booij, G. E. and Marle, J. v., editors, *Yearbook of Morphology 1991*, pages 165–183. Kluwer Academic Publishers, Dordrecht.

Giraudo, H. and Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review*, 8:127–131.

Inhoff, A. W., Radach, R., and Heller, D. (2000). Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory and Language*, 42:23–50.

Juhasz, B. (2007). The influence of semantic transparency on eye movements during english compound word recognition. In Van Gompel, R., Fischer, M., Murray, W., and Hill, R., editors, *Eye Movements: A Window on Mind and Brain*, pages 373–389. Elsevier, Amsterdam.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., and Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology*, 94:223–244.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P., editors, *Frequency and the emergence of linguistic structure*, pages 229–254. John Benjamins, Amsterdam.

Kostić, A. (1991). Informational approach to processing inflected morphology: Standard data reconsidered. *Psychological Research*, 53(1):62–70.

Kostić, A. (1995). Informational load constraints on processing inflected morphology. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*. Lawrence Erlbaum Inc. Publishers, New Jersey.

Krott, A., Baayen, R. H., and Schreuder, R. (2001). Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics*, 39(1):51–93.

Krott, A., Hagoort, P., and Baayen, R. H. (2004). Sublexical units and supralexical combinatorics in the processing of interfixed Dutch compounds. *Language and Cognitive Processes*, 19:453–471.

Krott, A., Schreuder, R., and Baayen, H. (2002). Linking elements in Dutch noun-noun compounds: Constituent families as predictors for response latencies. *Brain and Language*, 81(1-3):708–722.

Laudanna, A. and Burani, C. (1995). Distributional properties of derivational affixes: Implications for processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 345–364. Lawrence Erlbaum Associates, Hillsdale, N. J.

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22:1–38.

Libben, G. (2006). Why study compound processing? In Libben, G. and Jarema, G., editors, *The representation and processing of compound words*, pages 1–23. Oxford University Press, Oxford.

Moscoso del Prado Martín, F., Kostić, A., and Baayen, R. H. (2004). Putting the bits together: An information theoretical perspective on morphological processing. *Cognition*, 94:1–18.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., and Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30:1271–1278.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Weidenfeld and Nicolson, London.

Pluymaekers, M., Ernestus, M., and Baayen, R. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62:146–159.

Pollatsek, A., Reichle, E., and Rayner, K. (2003). Modeling eye movements in reading: Extensions of the E-Z Reader model. In Hyönä, Y., Radach, R., and Deubel, H., editors, *The mind's eye: Cognitive and applied aspects of eye movement research*, pages 361–390. Elsevier, Amsterdam.

Reichle, E., Reyner, K., and Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26:445–526.

Schreuder, R. and Baayen, R. H. (1995). Modeling morphological processing. In Feldman, L. B., editor, *Morphological Aspects of Language Processing*, pages 131–154. Lawrence Erlbaum, Hillsdale, New Jersey.

Taft, M. (1991). *Reading and the mental lexicon*. Lawrence Erlbaum, Hove, U.K.

Taft, M. and Ardasinski, S. (2006). Obligatory decomposition in reading prefixed words. *Mental Lexicon*, 1:183–189.

Van Son, R. and Van Santen, J. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, 47:100–123.

Samenvatting en conclusies

Chapter 8

Het onderwerp van dit proefschrift is de rol van morfologische structuur bij het begrijpen en produceren van polymorfemische woorden (d.w.z. woorden met twee of meer morfemen, bv. *vaat-was-er*) in het Nederlands, Engels, Fins en Servisch. De voornaamste onderzoeksvraag was hoe mensen gebruik maken van de probabilistische informatie die vervat is in morfemen, morfologische paradigma's (d.w.z. verzamelingen van woorden met een gemeenschappelijk morfeem) en volledige woorden. Deze vraag werd aangepakt door het onderzoeken van (i) het tijdsverloop van de activatie van morfemen en volledige woorden bij stillezen en spraakproductie, (ii) het effect van morfemen die hiërarchisch en orthografisch deel uitmaken van een grotere structuur (bv. *was-* en *-er* in *wasser*) op de visuele herkenning en akoestische productie van polymorfemische woorden, (iii) de rol van morfologische paradigma's (bv. *wasbak*, *waskamer*, *wasbeurt*) bij de lexicale verwerking van geflecteerde woorden, derivaties en samenstellingen en (iv) de interacties tussen morfologische en andere linguïstische prediktoren als codeterminanten van de kosten van lexicale verwerking van complexe woorden.

Een reeks experimenten over het stillezen van polymorfemische woorden en de akoestische analyse van deze woorden gaf ondersteuning voor het gebruik van een groter aantal informatiebronnen bij de lexicale verwerking van complexe woorden dan voorheen werd aangenomen. Op basis van onze experimentele bevindingen over het lezen van samenstellingen en derivaties, formuleerden we een probabilistisch model van visuele morfologische verwerking dat gebruik maakt van concepten en instrumenten uit de informatietheorie. In dit hoofdstuk bespreken we de experimentele resultaten, zetten we de belangrijkste aspecten van ons probabilistisch model uiteen en schetsen we een aantal onderwerpen die om verder onderzoek vragen.

Het tijdsverloop van morfologische verwerking

Met de oogbewegingsexperimenten uit Hoofdstukken 3 en 4 onderzochten we het het tijdsverloop van morfologische effecten op het lezen van in isolatie aangeboden Nederlandse samenstellingen (Hoofdstuk 3), en in zinscontext aangeboden Finse samenstellingen (Hoofdstuk 4). In beide experimenten werd gebruik gemaakt van relatief lange samenstellingen (in de orde van 8-12 letters in Hoofdstuk 3 en van 10-18 letters in Hoofdstuk 4). Door beperkingen in visuele scherpte kunnen onze ogen in het algemeen geen woorden van deze lengte verwerken zonder twee of meer maal te fixeren. Gedurende de eerste fixatie verkrijgen lezers in het algemeen een scherp foveaal zicht van de linkerconstituent van de samenstelling (bv. *vaat*), terwijl daaropvolgende fixaties foveale inspectie van de rechterconstituent (bv. *wasser*) mogelijk maken. Door de graduele visuele opname van lange woorden kan het de weergave van het patroon van oogbewegingen gebruikt worden om op een eenvoudigere manier de relatieve orde vast te stellen waarin morfemen vs. volledige woorden worden geactiveerd, dan in die gevallen waar morfologische effecten samen optreden in één enkele fixatie of, bij lexicale beslissing, in één reactietijd. In de oogbewegingsstudie in Hoofdstuk 5 werd de morfologische verwerking van Nederlandse derivaties (*success-vol*) tijdens het lezen in een zinscontext verder onderzocht. In tegenstelling tot samenstellingen werden deze woorden meestal in een enkele fixatie gelezen, maar toch geven ze inzicht in de temporele activatie van morfologische structuur.

De drie experimenten in Hoofdstukken 3 tot en met 5 toonden een vergelijkbaar en robuust temporeel patroon van morfologische effecten. Gegeven de verschillen in taal (Nederlands vs. Fins), lengte van de doelwoorden, taak (lexicale beslissing op geïsoleerde woorden in Hoofdstuk 3 vs. lezen in een zinscontext in Hoofdstukken 4 en 5) en het soort woordvorming (samenstelling vs. derivatie), is dit opmerkelijk. Onze bevindingen kunnen als volgt samengevat worden.

- De volgorde waarin de morfologische constituenten van een samenstelling geactiveerd worden, is nauw verbonden met het typische verloop van visuele verwerking, van links naar rechts (Hoofdstukken 3 en 4). De morfologische eigenschappen (frequentie en familiegrootte) van de linkerconstituent van de samenstelling toonden eerdere, sterkere en langdurigere effecten op oogbewegingsmaten dan de eigenschappen van de rechterconstituent. Wat betekenis betreft, ligt het hoofd van een samenstelling (de rechterconstituent) in het algemeen dichter bij de betekenis van de samenstelling als geheel (cf. *wasser* in *vaatwasser*). Men heeft daarom wel voorgesteld dat het hoofd van een samenstelling kan dienen als sleutel voor lexicale toegang tot de samenstelling als een geheel (bv. Juhasz et al., 2003). De door ons geobserveerde dominante rol van de linkerconstituent geeft echter geen ondersteuning voor een cruciale rol van de rechterconstituent (cf. Juhasz et al., 2007). Omdat de rechterconstituent later beschikbaar is voor het visuele systeem zou de identificatie ervan plaats kunnen vinden tegen de achtergrond van reeds bestaande kennis over de linkerconstituent. Deze resultaten pleiten er sterk voor dat modellen van morfologische verwerking de volgorde van visuele verwerking expliciet in beschouwing zouden moeten nemen. Tot dusver veronderstelden deze modellen dat complexe woorden onmiddellijk en volledig beschikbaar zijn voor het visuele systeem (bv. Baayen & Schreuder, 2000; Giraudo & Grainger, 2001; Taft & Forster, 1975, 1976; maar zie Pollatsek et al., 2003 voor een uitzondering).

Verder bewijs voor de links-rechts activatie van morfemen vonden we bij het stillezen van Nederlandse derivaties (Hoofdstuk 5). In de meeste gevallen worden deze woorden (*succes-vol*) in één enkele fixatie gelezen. Zowel de linkerconstituent (de basisvorm *succes*) en de rechterconstituent (het suffix *-vol*) zijn dan simultaan beschikbaar voor het visuele systeem. Desondanks vonden we dat het effect van één van de morfologische prediktoren van leestijd, namelijk de relatieve entropie van de morfologische families van de constituenten, verschilde naargelang er een links-rechts of rechts-links volgorde van morfologische activatie werd aangenomen. Hoewel de linker- en rechterconstituent visueel simultaan beschikbaar waren, leidde de aanname van een links-rechts volgorde tot een statistisch model met een significant betere fit op de geobserveerde gegevens dan wanneer een rechts-links volgorde werd aangenomen.

- Het frequentie-effect bij volledige woorden wordt karakteristiek gezien als ondersteuning voor lexicale toegang via de volledige vorm. In de twee experimenten met lange samenstellingen (Hoofdstukken 3 en 4) vonden we vanaf de eerste fixatie, en simultaan met het effect van de frequentie en familiegrootte van de linkerconstituent, een effect van samenstellingsfrequentie (d.w.z. van de volledige vorm) op leestijd. Gezien de lengte van de gebruikte samenstellingen kunnen we aannemen dat het vroege effect van samenstellingsfrequentie voorafgaat aan de volledige identificatie van alle letters en van de rechterconstituent van de lange samenstellingen. Dit effect suggereert dat lezers zich een idee beginnen te vormen over de identiteit van een samenstelling zodra er (misschien onvolledige) informatie beschikbaar is over woordlengte, identiteit van de linkerconstituent, of een vooruitblik op de rechterconstituent. Het vroege effect van samenstellingsfrequentie is problematisch voor sublexicale modellen van woordverwerking die vereisen dat zowel de linker- als rechterconstituent geactiveerd worden voordat het volledige woord op het lemmaniveau geactiveerd wordt (bv. Pinker, 1999; Taft & Ardasinski, 2006). Het simultane effect van samenstellingsfrequentie en de frequentie en familiegrootte van de linkerconstituent stelt vraagtekens bij supralexicale modellen, omdat zij voorspellen dat morfologische constituenten ná de volledige vorm geactiveerd worden en niet tegelijkertijd met de activatie van de volledige vormen. Anderzijds biedt dit effect sterke ondersteuning voor de parallelle verwerking van volledige vormen en morfemen, zoals twee- en meer-route modellen van morfologische verwerking verdedigen.

- Het effect van samenstellingsfrequentie blijft verder duren tijdens het lezen van een samenstelling. Eerst valt het samen met effecten die met de linkerconstituent geassocieerd zijn, later met de effecten die geassocieerd worden met de rechterconstituent. Deze continue betrokkenheid van de volledige vorm blijkt zelfs bij de meeste samenstellingen waarvoor morfologische decompositie aannemelijk is, bv. wanneer de constituenten een hoge frequentie of een grote familie hebben. Omdat zowel de volledige vorm als de afzonderlijke morfemen een bijdrage leveren aan de herkenning van een woord, zijn onze bevindingen niet eenvoudig te verzoenen met het "winner-takes-it-all" principe in het twee-route model van Frauenfelder en Schreuder (1991) en Schreuder en Baayen (1995).

- De families van de constituenten van samenstellingen en derivaties (gedefinieerd als verzamelingen van woorden die een bepaalde constituent delen: bv. *ijs-baan*, *ijs-klomp*, *ijs-schaats*) worden geactiveerd van zodra de constituenten beschikbaar worden voor het visuele systeem (Hoofdstukken 3-5). Deze bevinding staat in contrast met de algemeen aanvaarde mening dat het effect van morfologische families zich pas laat in het identificatieproces van een complex woord ontwikkelt en dat het toe te schrijven is aan de verspreiding van activatie binnen morfologische paradigma's en aan de semantische resonantie van familieleden die in vorm en ook vaak in betekenis gerelateerd zijn met het te herkennen woord (cf. De Jong, Schreuder & Baayen, 2000). In Hoofdstukken 3, 4 en 6 verfijnen we de bestaande kennis over de rol van families door aan te tonen dat de intrinsieke semantische component van het familieeffect aanzienlijk aangevuld wordt door de morfologische kenmerken van de leden van het paradigma.

Terwijl in Hoofdstukken 3-5 de temporele ontwikkeling van de effecten van morfologische structuur op oogbewegingen tijdens het lezen aan de orde werd gesteld, gaf de studie in Hoofdstuk 2 inzicht in het tijdsverloop van morfologische effecten bij spraakproductie. We onderzochten hoe de akoestische duur van de interfixen *-s-* en *-e(n)-* in Nederlandse polymorfemische samenstellingen (bv. *dier-en-arts*) varieert als functie van de probabilistische "bias" voor een interfix t.g.v. de morfologische families van elk van de constituenten. Eén van onze bevindingen was dat de hoeveelheid informatie in de morfologische familie van de rechterconstituent, d.w.z. de rechter-positionele entropie, een codeterminant is van de duur van het interfix *voordat* de rechterconstituent gearticuleerd wordt. Met andere woorden, hoe hoger de entropie van de familie van de rechterconstituent, hoe langer de akoestische realisatie van het interfix. Dit effect wordt geïnterpreteerd als ondersteuning voor: (i) activatie van morfologische families van volgende morfemen (bv. rechterconstituenten van samenstellingen) en (ii) de hoeveelheid informatie in het paradigma van het morfeem dat gepland wordt leidt tot interferentie met het articuleren van voorgaande morfemen (d.w.z. interfixen); zie hiervoor ook Pluymaekers et al. (2005).

De bevindingen in deze sectie wijzen er op dat huidige modellen van morfologische verwerking tijdens spraakproductie (bv. Levelt, Roelofs & Meyer, 1999) en visuele perceptie (bv. Baayen & Schreuder, 1999; Giraudo & Grainger, 2001; Taft, 1991) architecturaal te beperkt zijn om het empirisch vastgestelde tijdsverloop van morfologische effecten te verklaren.

Morfologische paradigma's

De in dit proefschrift beschreven experimenten dragen niet alleen bij tot onze kennis van het tijdsverloop van morfologische verwerking, maar tonen ook nog onbekende facetten van de paradigmatische organisatie van het mentale lexicon en de relevantie daarvan voor de herkenning en de productie van samenstellingen (Hoofdstukken 2-4), derivaties (Hoofdstukken 5 en 6) en geflecteerde woorden (Hoofdstuk 6). Onze bevindingen kunnen als volgt samengevat worden:

- In alle visuele herkenningsstudies in dit proefschrift (Hoofdstukken 3-6) hadden kenmerken van de families van morfologische constituenten (familiegroote of familiefrequentie) een groter effect op de snelheid waarmee complexe woorden verwerkt worden dan de frequentie van de constituenten zelf. Dit wijst erop dat deze herkenning van de morfemen van complexe woorden zich vooral afspeelt in hun morfologische context en dat deze herkenning niet afhangt van het oproepen van de relevante morfemen als onafhankelijke, contextvrije eenheden uit het lexicale langetermijngeheugen (het mentale lexicon).

- Morfologische families van constituenten worden in het algemeen beschouwd als een bron van paradigmatische ondersteuning voor de herkenning van die constituenten (cf. De Jong, Schreuder & Baayen, 2003). Daarom wordt aangevoerd dat morfologische families met meer leden, of meer frequente onderdelen, of meer informatie (gemeten door de entropie van de familie) de decompositie van complexe woorden in morfemen faciliteren (zie bv. De Jong, Schreuder & Baayen, 2000; Moscoso del Prado Martín, Bertram, Häikiö et al., 2004; Moscoso del Prado Martín, Kostić & Baayen, 2004). In Hoofdstuk 5 ontdekten we echter dat de invloed van families op het parseren complexer is. We merkten dat de verwerking van Nederlandse derivaties (bv. *succesvol*) optimaal was wanneer de morfologische families van de stam (bv. succes) en van het suffix (*-vol*) even klein of even groot waren. Hoe meer de grootte van de twee families uiteenliep, hoe groter de verwerkingskosten (geschat door meting van leestijden met behulp van oogbewegingregistratie). Dit effect (dat gemodelleerd werd door middel van de informatietheoretische maat relatieve entropie) geeft ondersteuning voor een soort competitie tussen morfologische families. Onevenwichtigheid in de hoeveelheid lexicale ondersteuning voor de morfemen kan een vertraging teweegbrengen in hun integratie tot een coherente representatie van de volledige derivatie.

- Morfologische families als verzamelingen van complexe vormen met een gedeelde constituent zijn slechts één mogelijke weerspiegeling van paradigmatische organisatie in het mentale lexicon. In onze lexicale beslissingsstudie met Engelse derivaties (Hoofdstuk 6) onderzochten we de toepasbarheid van een ander soort organisatie, gebaseerd op eerdere (in Hoofdstuk 6 uitvoerig besproken) informatietheoretische studies van geflecteerde woorden: mini-paradigma's (paren van stammen en hun derivaties, bv. *kind-unkind* "vriendelijk-onvriendelijk") en mini-klassen (de verzameling mini-paradigma's met hetzelfde derivationele affix, bv. *kind-unkind*, *true-untrue* "waar-onwaar", *pleasant-unpleasant* "aangenaam-onaangenaam", etc.). We ontdekten dat hoe meer de kansverdeling van het mini-paradigma afwijkt van de kansdistributie van de volledige mini-klasse (gemeten door de informatietheoretische maat *cross-entropy*), hoe trager de derivatie in het mini-paradigma benoemd wordt en hoe trager ze herkend wordt bij lexicale beslissing. Voor de stam geldt het omgekeerde: snellere benoeming en snellere herkenning bij lexicale beslissing. Deze studie toont aan dat de lexicale verwerking van derivaties niet enkel gevoelig is voor de hoeveelheid informatie in hun volledige morfologische paradigma, maar dat ze ook een weerspiegeling is van de probabilistische relaties tussen de verschillende niveaus van paradigmatische organisatie, van het micro-niveau van mini-paradigma's tot het macro-niveau van mini-klassen.

Terwijl in Hoofdstukken 3-6 veel nieuwe gegevens werden aangevoerd ter ondersteuning van de rol van morfologische paradigma's in het codetermineren van verwerkingskosten in visuele woordherkenning, onderzocht onze productiestudie (Hoofdstuk 2) of de gesproken realisatie van morfemen beinvloed wordt door de distributionele kenmerken van deze paradigma's. Uit recente studies blijkt dat de hoeveelheid probabilistische bias voor een interfix in een Nederlandse samenstelling gecodetermineerd wordt door de keuze van het interfix in de woorden die de linkerconstituent van de samenstelling delen (bv. *kandidaat-s-examen*, *kandidaat-en-lijst* en *kandidaat-stelling*, cf. Krott, Baayen & Schreuder, 2001). We zagen dat de probabilistische bias voor een interfix gecorreleerd is met de akoestische duur van dat interfix: Hoe voorspelbaarder het interfix, hoe langer de duur ervan. Deze bevinding vormt een paradox voor een belangrijke klasse van spraakproductiemodellen die stellen dat er een *negatieve* correlatie is tussen de waarschijnlijkheid van een eenheid van spraakproductie en de hoeveelheid articulatorische inspanning (bv. de akoestische duur) die gepaard gaat met de productie van die eenheid (cf. Jurafsky, Bell, Gregory & Raymond, 2001; Aylett & Turk, 2004; 2006; Van Son & Van Santen, 2005). Dit intrigerende resultaat wordt door ons verklaard door een onderscheid te maken tussen voorspelbaarheid vanuit syntagmatisch perspectief, wat negatief gecorreleerd is met akoestische saillantheid, en de hoeveelheid paradigmatische ondersteuning voor één van een klein aantal alternatieven, wat een positieve correlatie met akoestische saillantheid lijkt te hebben. Deze interpretatie, die we de Paradigmatische Signaalversterkingshypothese genoemd hebben, maakt toetsbare voorspellingen over de akoestische realisatie van andere linguïstische eenheden met paradigmatische alternatieven.

Samengevat wijzen onze resultaten erop dat de paradigmatische organisatie van het mentale lexicon een meer belangrijke en complexe rol speelt bij de lexicale verwerking van complexe woorden dan voorheen werd aangenomen.

Morfemen op lagere hiërarchische niveaus

De bestaande psycholinguïstische literatuur zegt weinig over de lexicale verwerking van complexe woorden met drie of meer morfemen (voor uitzonderingen, zie b.v. De Almeida & Libben, 2005; Inhoff, Radach & Heller, 2000; Krott, Baayen & Schreuder, 2001; Krott, Hagoort & Baayen, 2004). Desondanks zijn deze woorden interessant omdat ze verschillende niveaus van morfologische structuur hebben, van het hele woord (bv. *vaatwasser*), tot zijn onmiddellijke morfologische constituenten (bv. *vaat* en *wasser*) en de (diep) in deze constituenten ingebedde morfemen (bv. *vaat*, *was* en *-er*). In Hoofdstukken 2, 3 en 4 bestudeerden we polymorfemische samenstellingen in het Nederlands en het Fins om de rol van diep in de morfologische structuur ingebedde morfemen tijdens visuele woordherkenning en spraakproductie te onderzoeken. De resultaten in Hoofdstukken 3 en 4 laten duidelijk zien dat morfemen ook op diep niveau in de morfologische hiërarchie herkend en gebruikt kunnen worden als onafhankelijke betekeniseenheden bij het identificeren van samenstellingen. En de resultaten laten ook zien dat zij niet functioneren als alleen maar niet verder analyseerbare delen van letterstrings (zoals *-er* in *wasser*). Onze bevindingen hierover kunnen als volgt worden samengevat:

- Productievere affixen die ingebed zijn in de constituenten van samenstellingen (e.g,. *ing* in *plaatsingsbeleid*) lokken kortere leestijden uit, net zoals productieve affixen in bimorfemische derivaties.

- Relatief saillante (bv. langere, frequentere en structureel invariante) derivationele suffixen die ingebed zijn in trimorfemische samenstellingen (e.g, het suffix *-sto* in de Finse samenstelling *kirjastokortti* "bibliotheekkaart") geven betere parseringsaanwijzingen bij het segmenteren van samenstellingen in hun onmiddellijke constituenten (*kirjasto* "bibliotheek" en *kortti* "kaart") en faciliteren zo lexicale verwerking.

- Samenstellingen die ingebed zijn in trimorfemische samenstellingen (bv. *zaal+voet+bal*) leiden tot langere leestijden dan samenstellingen die ingebed zijn in derivaties (bv. *plaatsingsbeleid*). Dit wordt door ons geïnterpreteerd als een aanwijzing voor de hogere kosten van semantische integratie bij woorden met drie, vergelijken met twee, vrijstaande lexemen.

Verder tonen de resultaten van de eerder besproken productiestudie (Hoofdstuk 2) dat de verwerking van diep in de morfologische structuur ingebedde morfemen (*-s-* in *oorlog-s-verklaring*) gevoelig is voor de door het morfeem gedragen hoeveelheid informatie en -doorslaggevend- ook van de informativiteit van de morfemen die voor (bv. *oorlog*) of na (bv. *verklaring*) het diep ingebedde morfeem gerealiseerd worden.

Deze bevindingen voegen *granularity* toe aan onze kennis van morfologische verwerking. In onze interpretatie leveren deze nieuwe bevindingen het sterke evidentie dat morfemen die structureel en orthografisch ingebed zijn in grotere morfologische structuren een rol spelen bij het proces van complexe woordherkenning, waarbij de kans dat deze morfemen geactiveerd worden gecodetermineerd wordt door hun lexicaal-distributionele en orthografische saillantheid. Deze bevindingen stellen vele bestaande modellen van morfologische verwerking voor de uitdaging om een verklaring te vinden voor ingebedde morfemen als op zichzelf staande informatiebronnen.

Onderling afhankelijke bijdragen van morfologische structuur

In alle in dit proefschrift gerapporteerde studies zagen we dat de grootte en soms zelfs de aanwezigheid van de effecten van sommige morfologische eenheden afhankelijk waren van de grootte van de effecten die door andere morfologische en orthografische predictoren veroorzaakt werden. Dit interactief gebruik van morfologische informatiebronnen contrasteert met huidige één-route modellen en met de meeste parallelle twee-route modellen, die geneigd zijn om morfologische verwerking te vereenvoudigen tot de activatie van autonome lexicale representaties die blind zijn voor elkaars activatie (cf. Laudanna & Burani, 1985; Frauenfelder & Schreuder, 1991, en Schreuder & Baayen, 1995, een uitzondering is Baayen & Schreuder, 2000).

De meest voorkomende types interacties die we aantroffen toonden dat bij woorden met saillante (bv. frequente of paradigmatisch ondersteunde) morfemen, de kenmerken van de volledige vorm minder van invloed zijn dan de kenmerken van de ingebedde morfemen. Voor complexe woorden is het omgekeerde het geval: volledige vormverwerking wordt verkozen boven decompositie op grond van een frequente of korte volledige vorm, of op grond van de afwezigheid van duidelijke segmentatieaanwijzingen voor morfologische parsering. In de productiestudie in Hoofdstuk 2 vonden we een interactie tussen de maat voor de probabilistische bias voor het interfix van een samenstelling en de gemiddelde hoeveelheid informatie in de familie van de linkerconstituent van die samenstelling. In Hoofdstuk 3 vonden we dat de frequentie van een samenstelling interacteert met de frequentie van de linkerconstituent bij het codetermineren van leestijden in het Nederlands (Figuur 3.1), met als gevolg dat de frequentie van een samenstelling het kleinste effect heeft op samenstellingen waarvan de linkerconstituent een hogere frequentie heeft (d.w.z. die samenstellingen waarvoor morfologische decompositie de voorkeur draagt als verwerkingsroute). Vergelijkbaar hieraan vonden we in Hoofdstuk 4 dat de frequentie van een samenstelling het kleinste effect had op de Finse samenstellingen met een grotere famile voor de linker- of rechterconstituent (Figuren 4.1 en 4.2), wat verwerking door decompositie voor deze samenstellingen opnieuw aannemelijk maakt. In onze studie van Nederlandse derivaties (Hoofdstuk 5), bleek de lengte van het suffix een parameter in het reguleren van de beschikbare verwerkingsroutes, en op die manier in de modulatie van de grootte van verschillende morfologische effecten. Voor woorden met extreem korte, en daardoor niet-saillante, suffixen, werd de voorkeur gegeven aan lexicale toegang via de volledige vorm, en toonde de frequentie van de derivatie het grootste faciliterende effect op leestijden, terwijl de morfemen slechts een zwak effect vertoonden. Naarmate de saillantheid van het affix toenam met de lengte van het suffix, verminderde het effect van de frequentie van de derivatie en verdween dit effect bijna voor woorden met langere suffixen, terwijl de effecten die in verband staan met de parsering van de morfemen van een derivatie, toenamen.

Hoewel interdependenties in morfologische verwerking in overweging genomen werden in eerdere experimentele studies en modelleringsonderzoek (bv. Bertram & Hyönä, 2003; Baayen & Schreuder, 2000), bevestigen onze data dat deze interdependenties zo algemeen zijn dat ze de lexicale verwerking van complexe woorden als het ware reguleren. We denken dat dit een weerspiegeling is van de wisselwerking tussen de beschikbare routes voor morfologische verwerking, waaronder opslag in het langetermijngeheugen en berekening. Op basis van de empirische gegevens kunnen we besluiten dat elk model van morfologische verwerking een expliciete verklaring moet bieden voor het feit dat de bijdrage van één informatiebron de bijdrage van andere beschikbare bronnen bij morfologische verwerking moduleert.

Het modelleren van morfologische verwerking bij visuele woordherkenning

De empirische gegevens uit Hoofdstukken 3-6 boden de basis voor de formulering van een nieuw probabilistisch model (PROMISE). Dit model geeft ons een wiskundig, op informatietheorie gebaseerd, instrumentarium waarmee parallelle multipele route modellen van morfologische verwerking bij de visuele herkenning van polymorfemische woorden beschreven kunnen worden. PROMISE bouwt verder op onderzoek van Kostić (1991; 1995) en Moscoso del Prado Martín, Kostić & Baayen (2004), die als eersten informatietheoretische inzichten toepasten op experimentele gegevens over de morfologische verwerking van geflecteerde woorden, derivaties en samenstellingen. Ons model breidt ook het modelleringskader van Baayen, Wurm en Aycock (2007) uit. Het voegt er meer morfologische informatiebronnen aan toe en het implementeert de interacties tussen deze bronnen.

In Hoofdstuk 3 schetsten we de noodzakelijke kenmerken van modellen die het complexe patroon van morfologische effecten in de verwerking van samenstellingen willen verklaren. Deze kenmerken omvatten:

- expliciete aandacht voor de temporele orde van informatieopname, waaronder een links-rechts volgorde van activatie van de morfemen bij complexe woorden die in verschillende fixaties gelezen worden;
- afwezigheid van stricte opeenvolging in het verwerken van informatie, d.w.z. simultane verwerking van de beschikbare informatie op de verschillende niveaus van representationele hiërarchieën, zoals de volledige vorm van complexe woorden (bv. *vaatwasser*), hun onmiddellijke constituenten (*vaat* en *wasser*), morfemen op diep ingebedde niveaus van de morfologische structuur (*was* en *-er*), en morfologische paradigma's van deze constituenten (*vaatdoek*, *vaatkwast*, *vaatwater*, enz.);
- de mogelijkheid dat één verwerkingscue (bv. perceptuele saillantheid van een morfologische constituent) een modulerende invloed heeft op de bijdrage van andere verwerkingscues (bv. kenmerken van de volledige vorm, zoals frequentie).

In Hoofdstuk 4 werden deze kenmerken in het PROMISE model geïmplementeerd. Het conceptuele kader van dit model beschouwt het mentale lexicon als een langetermijngeheugen voor lexicale informatie. Een binnenkomende visuele stimulus vormt een sleutel voor de toegang tot deze lexicale informatie. De geaccumuleerde kennis van woorden en hun paradigmatische en syntagmatische kenmerken bepalen de informatielading van een woord en daardoor ook de snelheid waarmee informatie over dat woord uit het lexicale geheugen opgehaald kan worden. PROMISE formaliseert de informatielading van een morfologische structuur door gebruik te maken van de wellicht meest fundamentele uitdrukking uit de informatietheorie: De informatie (I) van een (linguïstische) eenheid kan uitgedrukt worden als min de log-waarschijnlijkheid (P) van die eenheid. Naarmate P afneemt, neemt I toe: Minder waarschijnlijke gebeurtenissen dragen meer informatie. Een fundamentele veronderstelling van ons model is dat de tijd die het oog aan een constituent of aan een woord besteedt, evenredig is met de totale hoeveelheid lexicale informatie die op dat moment in het langetermijngeheugen aanwezig is voor de identificatie van die constituent of dat woord (cf. Moscoso del Prado Martín, Kostić & Baayen, 2004). Gebeurtenissen met een lage waarschijnlijkheid, en daardoor een grote informatielading, vragen meer verwerkingstijd en -middelen.

In Hoofdstukken 3-6 identificeerden we een uitgebreide reeks bronnen van morfologische informatie, waaronder: de niet-conditionele waarschijnlijkheid van een volledige woordvorm en van zijn morfemen (d.w.z. de aannemelijkheid dat deze woorden geraden worden zonder verdere contextuele informatie); de conditionele waarschijnlijkheid van een morfeem gegeven zijn positie in een complexe woordvorm of gegeven de beschikbaarheid van andere morfemen in het lexicale verwerkingssysteem; de waarschijnlijkheid van een morfeem in zijn paradigma; en de afstand tussen de kansverdeling van een morfeem en die van zijn klasse (Hoofdstukken 5, 6), zoals gemeten door de begrippen entropie en cross-entropie. De verwerkingstijd is evenredig met de gewogen som van die hoeveelheden informatie (zie de algemene bespreking in Hoofdstuk 4 en vergelijking (4.15) voor de uitvoerige behandeling van het wiskundige kader van PROMISE).

De gewichten moduleren de impact van individuele informatiebronnen op verwerkingstijd en zijn tevens de de parameters van het model. Omdat PROMISE de informatielading van de uiteenlopende morfologische bronnen weergeeft in termen van de distributionele kenmerken van morfologische structuur (bv. woordfrequentie, familiegrootte, frequentie, enzovoort), kunnen de parameters van PROMISE onmiddellijk worden geschat op de gegevens met behulp van multipele regressiemodellen.

PROMISE voldoet op de volgende manier aan de in Hoofdstuk 3 geformuleerde (en hierboven gespecificeerde) voorwaarden.

- De studie in Hoofdstuk 4 omschrijft hoe het model omgaat met de temporele visuele beschikbaarheid van morfologische informatie (die doorgaans van links naar rechts verloopt). Informatiebronnen die vroeg in de visuele opname beschikbaar zijn, hebben een aantoonbaar groter belang bij de herkenning van samenstellingen (cf. de kleinere rol van variabelen die in verband staan met de rechterconstituent in vergelijking met eigenschappen van de linkerconstituent). In de modelvergelijking kunnen de gewichten *w* voor "vroege" informatiebronnen vermenigvuldigd worden met een tijdscoëfficient α_1 , zodat $\alpha_1 > 1$. Voor "late" informatiebronnen is de waarde van α_2 kleiner of gelijk aan 1. Net zoals voor de gewichten w kan de waarde van α meteen geschat worden door een vergelijking van de regressiecoëfficiënten van een predictor voor modellen met een vroege en latere meting voor visuele opname.

- Een fundamentele veronderstelling van PROMISE is dat een groot aantal informatiebronnen gemeenschappelijk beoordeeld worden, bv. frequentie van de volledige vorm, morfeemfrequentie, familiegebaseerde schattingen van de lexicale connectiviteit van een morfeem, enz.. De hierboven uiteengezette desiderata voor het simultaan verwerken van beschikbare informatie op verschillende niveaus, worden op deze manier door PROMISE geformaliseerd.

- De statistische interacties tussen morfologische predictoren, die we in Hoofdstukken 3-6 opmerkten, leidde tot het idee dat informatiebronnen bij woordherkenning interactief gebruikt worden, waarbij de bijdrage van sommige bronnen de bijdrage van andere bronnen op lexicale verwerking moduleert. Het PROMISE model formaliseert deze intuïtie op verschillende manieren. Om te beginnen brengen schattingen van de meeste kansen, bronnen met tegenovergestelde gewichten met zich mee. De waarschijnlijkheid van een woord wordt bijvoorbeeld gedefinieerd als de frequentie van dat woord gedeeld door de grootte van het corpus. Omdat de informatielading van een woord gelijk is aan min de (binaire) log-waarschijnlijkheid van dat woord, volgt dat hoe meer woordfrequentie verwerking faciliteert, hoe meer inhibitie er ontstaat vanuit de grootte van het corpus: hoe groter het corpus, hoe groter de rijkdom van de woordenschat en hoe moeilijker de identificatie van de doelvorm (zie Hoofdstuk 4, algemene bespreking, voor meer details). Dit voorbeeld weerspiegelt één van de vele wisselwerkingen in de verwerking van complexe woordvormen. Vervolgens implementeerden we in Hoofdstuk 4 expliciet de interactie tussen informatiebronnen door de coëfficiënt m.b.t. een bepaalde informatiebron (de familiefrequentie van de linkerconstituent) evenredig te maken met de waarde van een andere verwerkingsaanwijzing (de frequentie van de samenstelling); zie vergelijkingen (4.16)-(4.19). Tenslotte maakten we in Hoofdstukken 5 en 6 gebruik van relatieve entropie als informatietheoretische maat voor de manier waarop de bijdrage van de morfologische familie van één constituent gemoduleerd wordt door de familiegrootte van de andere constituent. Met behulp van cross-entropie maakten we een schatting van hoe de divergentie tussen de kansverdeling van het mini-paradigma van een derivatie en die van haar mini-klasse zich vertaalt in grotere verwerkingskosten.

Samengevat kan PROMISE beschouwd worden als een formalisering van de gedachte dat lezers en luisteraars uiteenlopende informatiebronnen gebruiken bij het herkennen van complexe woordvormen (zie Libben, 2006). De parameters van PROMISE kunnen afgezet worden tegen de regressiecoëfficiënten van statistische modellen die gefit werden op empirische gegevens. De geschatte waarde van de parameters vertelt ons niet enkel waarom sommige informatiebronnen boven andere verkozen worden, maar ook op welk moment in de tijd en met welke kosten voor het verwerkingssysteem. Belangrijk hierbij is dat PROMISE verschillende soorten woordvorming (bv. derivatie en samenstelling) op dezelfde manier behandelt.

Thema's voor verder onderzoek

Het onderzoek dat in dit proefschrift werd voorgesteld, toont aan dat de studie van morfologische complexiteit nieuwe inzichten biedt in de organisatie van het mentale lexicon en in de cognitieve processen m.b.t. woordproductie en -herkenning. We belichten nu verschillende gebieden van morfologisch onderzoek die om verdere studie vragen.

Ten eerste spreken modellen van morfologische verwerking, waaronder ook PROMISE, zich niet uit over de implementatie van morfologische activatie in de hersenen. Toch is het noodzakelijk om te weten hoe het complexe patroon van morfologische effecten dat uit gedragsstudies blijkt (bv. de lexicale beslissings- en oogbewegingsexperimenten in dit proefschrift) zich weerspiegelt in de elektrische, magnetische of biochemische werking van het brein. Een dergelijk verband zou een fysiologische basis bieden voor de hier ontwikkelde hypotheses over het tijdsverloop van lexicale verwerking en de paradigmatische organisatie van het lexicale langetermijngeheugen. Een veelbelovende manier om die verband te maken is de combinatie van oogbewegingsstudies met experimenten die gebruik maken van event-related potentials. Verder onderzoek naar de verwerving van morfologie door kinderen kan een inzicht bieden in hoe restricties op het gebied van lexicale kennis, geheugen en representatie, de rol van morfologische structuur moduleren in het leren, begrijpen en produceren van polymorfemische woorden.

Ten tweede kan het PROMISE model op verschillende vlakken uitgebreid en verfijnd worden. Multipele regressiestudies van verschillende types morfologische complexiteit bij verschillende talen en taken zouden bijvoorbeeld nuttig zijn om de grenscondities en de verwachte parameterwaarden van het model te specificeren. Wanneer de parameterwaarden gevalideerd zijn bij experimentele replicaties, kan het model ook dienst doen als predictief - en dus niet enkel descriptief - instrument voor het onderzoeken van de psycholinguïstische processen bij de herkenning van complexe woordvormen.

Ten derde kan PROMISE op een eenvoudige manier verenigd worden met algemene modellen van oogbewegingscontrole, zoals E-Z reader (bv. Reichle, Rayner & Pollatsek, 2003) of SWIFT (bv. Engbert, Nuthmann, Richter & Kliegl, 2005). Het samen in overweging nemen van de parameters van PROMISE en andere visueel motorische, orthografische en lexicale parameters, kan de voorspellingen van deze modellen i.v.m. verwerking van morfologisch complexe woorden verbeteren.

Hoewel PROMISE inzichtelijke resultaten bood voor samenstelling en derivatie, is de toepassing op flexie (gebaseerd op een grootschalige studie van geflecteerde woorden) nog een taak voor de toekomst. Tenslotte staat PROMISE, dat ontwikkeld werd voor visuele woordherkenning, voor de uitdaging om ook voorspellingen te maken voor auditieve woordherkenning en voor de productie van morfologische complexe woorden.

Slotopmerkingen

In psycholinguïstisch onderzoek naar morfologische verwerking is het symbolisch theoretisch perspectief vaak dominant geweest. Dit perspectief stelt dat discrete morfemen tot regelmatige complexe structuren gecombineerd worden door deterministische regels en dat onregelmatige complexe vormen in het geheugen opgeslagen worden (cf. Pinker, 1999). Een alternatieve subsymbolische opvatting stelt dat morfologische structuur fundamenteel probabilistisch is en dat ze voortvloeit uit de statistische regelmatigheden waarmee woordvorm en -betekenis met elkaar verbonden zijn (cf. Hay & Baayen, 2005). Die opvatting is ook een inspiratie geweest voor Word and Paradigm morfologie (bv. Blevins, 2003), waarin gesteld wordt dat woorden (zowel eenvoudige als complexe) de basiseenheden zijn van het lexicon en dat deze woorden georganiseerd zijn in paradigma's die op een hoger niveau dan weer georganiseerd zijn in klassen. Dit proefschrift had als ultieme doel om bestaande theorieën van morfologische verwerking in visuele herkenning en spraakproductie te bevorderen. In de studies die hier gerapporteerd werden, vond ik aanwijzingen in het voordeel van de probabilistische benadering na morfologie. Ik hoop dat dit proefschrift een bijdrage levert aan de empirische grondslagen en aan de theoretische vooruitgang van deze inspirerende benadering.

Acknowledgments

This part of the dissertation is perhaps the most gratifying one, as it gives me the chance to thank all those who filled the three years of my Ph.D. project with help, encouragement, excitement and joy. My *promotores*, Harald Baayen, Rob Schreuder and Mirjam Ernestus, duly receive the lion share of my appreciation. Harald took the risk of inviting me to become a Ph.D. student in his group without ever talking to me in person, and that changed my life in more ways than he could probably imagine. In the years to follow Harald was a true mentor to me. He spent countless hours teaching me the trade of psycholinguistics from scratch, delving into current theories of language production and perception, experimental design, statistics, data analysis, writing and many other topics. Also - whether that was his didactic goal or not - in my eyes he set an example of enthusiasm and unbending dedication to scientific pursuit. Our work was no less productive when Harald moved to Canada, thanks to his visits to Nijmegen, my trip to Edmonton and the blessings of skype, and I am grateful for that continuity as well. There is something that I cherish as much as my apprenticeship with Harald, and that is the fact that I always enjoy his company, his music, his sense of humor, and the bursts of laughter that we constantly exchange when we talk. Thank you, Harald, for leading me through this rite de passage with a smile. Tineke and Tera, thank you for being wonderful hostesses during my stay in Edmonton, and my apologies for always taking up so much of Harald's time.

In the course of these three years I set a rule for myself. Never embark on a project without discussing it with Rob Schreuder first. Rob's vast erudition in psycholinguistics and cognitive psychology helped me more than once to put my thoughts in the right perspective and to link my ideas to the literature in the field. Of equal importance was Rob's ability to remind me of mundane practicalities of academic life, such as deadlines, submissions, reports, research contacts, conferences, and many others. At the same time, I am grateful to you, Rob, for keeping me constantly aware of the fact that there's life beyond science and its

demands are not to be ignored. Thank you for helping out with arranging my life in Nijmegen, for your unfailing political advise, and for hearing out my joys and frustrations throughout these years.

It is thanks to Mirjam and her line of research that my project was not confined to the "eyes" (visual processing) but also covered the "ears" (acoustic production and auditory comprehension). I gained much knowledge from Mirjam on phonology, phonetics, and speech corpora, as well as on the collection and analysis of acoustic data. Mirjam taught me persistance and meticulousness in getting answers to my questions regardless of the effort it takes. The labor she took to re-read and re-edit my manuscripts showed to me what precision and clarity are required of scientific prose. Besides, most importantly, Mirjam was not only generous with her time when I barged in her office with a question or a vague thought in the midst of her busy day. She also found time to cheer me up by chatting about non-academical matters or sharing a joke. Thank you, Mirjam, for all your support, commitment and for making me feel self-confident even in difficult times.

I also wish to thank the members of the manuscript committee, Lou Boves, Herbert Schriefers and Reinhold Kliegl, for their valuable feedback on the dissertation. I deeply appreciate Reinhold Kliegl's kind gesture of including me as an associate fellow in the ECRP project "Eye-movement control in reading" and for sharing his experimental data and discussing research issues of mutual interest. Thanks are due to the members of my corona for reading the dissertation and giving their feedback on my research: Doug Davidson, Teun Dijkstra, and Dominiek Sandra.

I feel privileged to have been working in what was formerly known as the IWTS group and what now is the CLSM research group. Mark Pluymaekers was an absolutely essential part of my first year here. He kept patiently answering the many of my newbie questions, he taught me many tricks in LATEX, statistical regression modeling, spoken corpora of Dutch, and scientific writing, and finally – fortunately for me – he shared the effort of preparing my first publication within this project. Thank you, Mark. I am happy to have enjoyed the help and the friendly company of Vitória Piai in the last year of my project. On several occasions her creative feedback made me sharpen my ideas and come up with new interpretations of the data. She took over most of the laborious technical issues involved in preparing and running eye-tracking and speech production experiments. And thank you, Vitória, for being my paranimf. Laura, thank you for being the perfect roommate, the patient listener of my attempts at conversations in Dutch, and the only Olympic person I've ever come across so far. At CLSM, my fellow PhD students, post-docs, visitors, masters students and student assistants kept me company at lunches, courses and conferences. They made sure that my experiments are well-planned and run, my table tennis thirst is quenched, my bookshelves are assembled, and my official paperwork written in the Dutch officialese is filled in and submitted. And of course, we had lots of stimulating discussions on topics ranging from experimental stimuli to statistical modeling to baroque music to jenever tasting. Thank you: Annelie, Christophe, Edwin, Esther, Francisco, Hanke, Hilde, Karen, Kevin, Kors, Lee, Lucie, Marco, Marieke, Mybeth, Roel, and Wieke.

My list of thanks would have been incomplete without collaborators and colleagues world-wide who contributed to my research, and in many cases became my friends. Raymond Bertram taught me most of what I know about eye-tracking and led my way into the research field of eye-movement control in reading. He was my kind host during my three-week stay at University of Turku in 2007, and he also came to Nijmegen on two research visits in 2007 and 2008. This collaboration came with long hours in experimental booths, temperamental discussions, tedium of data analysis, captivating conversations about meaning of life, and also numerous happy get-togethers. Thanks, Raymond, for bearing with me over several joint projects, some of them ongoing, and even more so for the strong friendly ties between us and our families. I am also grateful to Jukka Hyönä for his consultations on complicated aspects of eye-movement research.

I am grateful to Tom Wasow who sponsored my three-month stay at Stanford University in 2008 and who I will be working with for the next two years while carrying out my NWO-funded Rubicon research programme at Stanford. Thank you, Tom, for helping me settle down at Stanford, for the experiment that we set together and for the theoretical expertise that you contribute to our collaboration. During my Stanford visit I enjoyed talking to Joan Bresnan, Dan Jurafsky, Arto Anttila, Neal Snider, Tyler Schnoebelen and many graduate students at the Department of Linguistics. Thank you all for feedback and inspiration. I am looking forward to spending more time in your collegial and enthusiastic circle.

Petar Milin and Dušica Filipović Đurđević introduced me to the advanced information-theoretical way of thinking about morphological paradigms and are responsible for keeping my Slavic past alive. Thank you for that, Petar and Dušica, and for the pleasure that I always take in talking to both of you. Thanks are due to Emmanuel Keuleers who made me revisit more than once the methodological foundations of my research and data analysis, and who embellishes my life with witty conversations, passion for guitar music and perfect cooking. Michael Dambacher is my ERP comrade-in-arms. Thank you, Michael, for subjecting my wild flights of phantasy to rigorous sanity checks. Gary Libben, Fermin Moscoso del Prado Martín and Ingo Plag provided me with friendly and valuable advice on several occassions – my appreciation for that.

Working as a Radboud University PhD student at the Max-Planck-Institute gives one access to vast intellectual and technical resources. In the course of my project I benefited from the meetings of the Language and Speech Technology group, the formal Max-Planck-Institute colloquium, the MPI lunch talks, the CLS colloquium, the Nijmegen lectures and the LOT schools. Even more so I benefitted from communication with researchers at the Max-Planck-Institute and the Center for Language and Speech. Thanks are also due to the Technical Group of the MPI, especially Tobias and Johan, for their support of the experimental and computational facilities I made extensive use of. Els den Os and Lisenka Fox (Bureau Onderzoek Letteren) and Kitty Rikken (Personeelszaken) helped me a lot in organizational and financial aspects of my stay in Nijmegen. It is also a privilege for me to publish this dissertation in the MPI Series for Psycholinguistics.

My PhD life would have been one-sided and dull if it were not for the friends to cheer it up. Thank you: Ria, Adriana (my paranimf!) and Doug, Kathrin and Andre, Golda and Shurik, for the precious hours of levity. Golda deserves extra thanks for providing her drawing for the cover page.

I am grateful to my brother Max and my mom Victoria for proudly supporting every step of my way into science, and for their love and encouragement. I would not be where I am now without having my late father Boris as a role model of a parent and a true scientist. I thank my children, Boris and Joseph, for gracefully accepting my frequent spells of absence and for being such an inexhaustible source of shared affection and comfort. They alone have (and exert!) the power to make me forget all about the lures of research life. And Katya, my love, this book is as much your achievement as it is mine. Thank you for believing in me and for making me able to carry on.

Curriculum Vitae

Victor Kuperman was born in St. Petersburg, Russia, on 29 July 1973. In 1990, he started his undergraduate studies in Computer Science at the Technical University, St. Petersburg. In 1994 he continued his education at the Hebrew University of Jerusalem and was awarded the BA degree in Linguistics and the English Language and Literature in 1997. In 2001, he graduated from the master's programme in information science at the School for Library, Archive and Information Studies, Hebrew University of Jerusalem. His MA thesis was on the application of Zipf's law to the patterns of authorship in the Internet mailing lists. In 2001-2005, he worked as a technical support analyst in the Fourth Dimension Software and Roundtrip Systems software companies based in the San Francisco Bay area. In 2005, he started a PhD project "With an eye and an ear to multiply complex structures" at the Interfaculty Unit for Language and Speech (IWTS, now CLSM) at Radboud University Nijmegen. He has been awarded the Rubicon grant by the Netherlands Organization for Scientific Research (NWO) to carry out a two-year post-doctoral research programme on sentence processing at the Department of Linguistics, Stanford University.

Publications

Kuperman, V., Bertram, R., Schreuder, R. & Baayen, R.H. (2008). Reading of Polymorphemic Dutch Compounds: Towards a Multiple Route Model of Lexical Processing. In press in *Journal of Experimental Psychology: Human Perception and Performance*.

Kuperman, V., Bertram, R. & Baayen, R.H. (2008). Morphological Dynamics in Compound Processing. In press in *Language and Cognitive Processes*.

Milin, P., Kuperman, V., Kostić, A. & Baayen, R.H. (2008). Paradigms bit by bit: an information-theoretic approach to the processing of inflection and derivation. In press in Blevins, James P. and Juliette Blevins (eds.), *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press.

Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, R.H. (2007). Morphological redundancy and the acoustic salience of interfixes in Dutch compounds. *Journal of the Acoustical Society of America* 121(4), 2261-2271.

Kuperman, V. (2006). Productivity in the Internet Mailing Lists: A Bibliometric Analysis. *Journal of the American Society for Information Science and Technology* 57, 51-59.

Kuperman, V. (2006). Narratives of Psychiatric Malingering in Works of Fiction. *Medical Humanities* 32, 67-72.

Kuperman, V. & Zislin, J. (2005). Semiotic Perspective of Psychiatric Diagnosis. *Semiotica* 155(1-4), 1-13.

Zislin, J., Kuperman, V. & Durst, R. (2002). The Generation of Psychosis: A Pragmatic Approach. *Medical Hypotheses* 58(1), 9-10.

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*

2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsaggital articulography. *Niels O. Schiller*

3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*

4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*

5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*

6. Gesture and speech production. *Jan-Peter de Ruiter*

7. Comparative intonational phonology: English and German. *Esther Grabe*

8. Finiteness in adult and child German. *Ingeborg Lasser*

9. Language input for word discovery. *Joost van de Weijer*

10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*

11. Producing past and plural inflections. *Dirk Janssen*

12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*

13. From speech to words. *Arie van der Lugt*

14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*

15. Interpreting indefinites: An experimental study of childrenâĂŹs language comprehension. *Irene Krämer*

16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*

17. Moving eyes and naming objects. *Femke van der Meulen*

18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*

19. Morphology in speech comprehension. *Kerstin Mauth*

20. Morphological families in the mental lexicon. *Nivja H. de Jong*

21. Fixed expressions and the production of idioms. *Simone A. Sprenger*

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*

24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*

25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*

26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*

27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*

29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*

30. A grammar of Jalonke argument structure. *Friederike Lüpke*

31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*

32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*

33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*

34. Phonetic and lexical processing in a second language. *Mirjam Broersma*

35. Retrieving semantic and syntactic word properties. *Oliver Müller*

36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*

37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*

38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*

39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*

40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*

41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*

42. The acquisition of auditory categories. *Martijn Goudbeek*

43. Affix reduction in spoken Dutch: Probabilistic effects in production and perception. *Mark Pluymaekers*

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*

45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*

46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*

48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*

49. Phoneme inventories and patterns of speech sound perception. *Anita E. Wagner*

50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*