Experimental and psycholinguistic approaches to studying derivation

R. Harald Baayen
Eberhard Karls University, Tübingen, Germany
and
University of Alberta, Edmonton, Canada

This chapter provides a critical overview of experimental and computational research on the processing and representation of derived words. It begins with an introductory section addressing methodological issues: The pros and cons of various popular experimental tasks, issues with respect to the selection of materials, as well as the relevance of experimental research for morphological theory. The main section reviews two opposing classes of theories for the organization of the mental lexicon: theories building on the dictionary metaphor, and theories seeking to understand lexical processing without a mental dictionary and without theoretical constructs such as the morpheme.

1 Methodology

1.1 Experimental methods

A wide range of experimental methods is available for probing the processing of complex words. In what follows, some of the more widely used methods are introduced, together with their advantages and disadvantages.

For research on comprehension, the LEXICAL DECISION TASK is used widely. Participants are presented with a sequence of stimuli, one at a time, which include both existing words (such as *table*) and non-existing words (such as *flurtle*). They are asked to decide, as quickly and accurately as possible, whether each stimulus is a word or a nonword, by pressing one of two response buttons. Stimuli can be presented in writing on a computer screen (visual lexical decision) or over headphones (auditory lexical decision). The time it takes to execute a response (the response latency or reaction time) as well as the accuracy of the lexical decision have been found to be highly informative about the processing costs of different kinds of complex words.

The lexical decision task offers several advantages. First of all, it is easy to administer, especially for the study of reading. In recent years, large-scale lexical decision studies have been carried out, collecting reaction times for tens of thousands of words (see, e.g., Balota et al., 1999, 2004, 2007; Keuleers et al., 2010, 2012; Ferrand et al., 2010). At the time of writing, several labs are running experiments using crowd sourcing, with volunteers running lexical decision experiments on remote laptops and smartphones. However, the lexical decision task also has many disadvantages. First, the task requires participants to make a metalinguistic judgement, which is far removed from normal comprehension. Second, words are presented in isolation, whereas in experience words tend to be part of sentences or utterances. As a consequence, lexical decision latencies tend to show only weak correlations with processing measures from the eye-movement record Kuperman et al. (2013a) . Third, how the nonwords are constructed (see, e.g., Keuleers and Brysbaert, 2010) as well as what kind of words are included as fillers in the list of stimuli (see, e.g., Feldman et al., 2009) may substantially affect the results obtained. Finally, reaction times and accuracy scores are uninformative about the time course of lexical processing: Typically, the early stages of lexical processing as revealed by eye-tracking studies can be very different from the later processing stages evaluating lexicality decisions (Miwa et al., 2013), and may even be misleading.

The lexical decision task is often combined with a so-called PRIMING experimental treatment, in which a given target word is preceded by carefully selected other words, the so-called primes. Primes can be words presented at a certain distance earlier in the experimental list (long-distance priming). In the masked priming task (using visual lexical decision), primes are presented for a very short duration (e.g., 60 ms), often preceded by a mask of random letters or hash marks, before the target word is presented, in which case most subjects do not become aware that a prime word was presented. When masked priming is used to study morphological processing, typical priming treatment conditions are an identity condition (*good* priming *good*), a related condition (*goodness* priming *good*), a form condition (*food* priming *good*) and an unrelated condition (*hand* priming *good*). The results one tends to obtain are that responses are fastest in the identity condition, intermediate in the related condition, and slowest in the unrelated (control) condition.

The effect of a masked related prime has been attributed to the prime word partially preparing the way for lexical access for the target word, either by 'opening' the lexical entry of the target, or by partially pre-activating the target (Forster, 1999). Theorists accepting this interpretation compare the priming effect against the unrelated baseline, in which case a related prime will elicit shorter latences than the unrelated condition. The reason that a related prime also elicits longer latencies than the identity condition is attributed to a channel capacity problem, with two words having to be processed nearly simultaneously instead of just one word. However, Norris and Kinoshita (2008) argue that in masked priming, the perceptual system cannot properly distinguish between the prime and the target as different perceptual events. As a consequence, the orthographic information of prime and target would blurr into one perceptual whole, and the more the prime differs from the target, the more noise it contributes to the perception of the target, and the longer response latencies become. Norris and Kinoshita (2008) also show that priming effects can be task-specific: present in visual lexical decision, but absent (for the same stimuli) for a same-different task. This implies that the effects of priming need not be an automatic consequence of the structure of the mental lexicon, but arise 'on-line' depending on the demands of the task.

For the study of morphological processing in reading, modern eye-tracking systems offer the possibility of tracing, in considerable detail, and with great accuracy, where the eye lands in a complex word, how often it fixates within that word, and whether the eye will return to the word after having fixated elsewhere (see, e.g., Rayner, 1998; Kuperman et al., 2009, 2010). The advantages of using eye-tracking are, in addition to providing detailed insight into the time course of lexical processing, that words can be presented in sentential and/or discourse context, providing experiments with enhanced ecological validity compared to tasks involving lexical decisions. The major disadvantage of eye-tracking is that it is currently impossible to gather data with crowd sourcing. However, this may change in the near future, with the development of user interfaces for smartphones that track where the eye is fixating on the screen.

Experimental research on speech production is much more difficult than research on comprehension. Whereas in comprehension experiments, materials with desired controlled properties can be presented to participants, the challenge in production studies is to get participants to produce the words with the critical properties of interest. In principle, one could present words in writing and ask participants to read them out loud, but this has the serious drawback that results conflate an initial comprehension process with a subsequent production process. Three tasks have been widely used in research on morphological processing in speech production: picture naming, implicit priming, and the picture-word interference task.

For studying speech production from initial conceptualization to final articulation, the picture naming task is a good choice. In this task, participants are presented with line drawings or photographs, and are asked to say out loud as quickly and accurately as possible what the picture denotes. In this task, the input is non-linguistic, and hence the response variables — naming latency and accuracy — gauge the costs of preparing for speech without contamination from linguistically mediated comprehension. The task has two disadvantages, however: Only picturable nouns, verbs, and adjectives can be presented, and the temporal information obtained is restricted to the onset of articulation.

The picture-word interference task seeks to obtain information about the time-course of lexical processing by combining picture naming with the presentation of distractors, typically words presented visually or auditorily with the picture. The critical manipulation here is the amount of time between the presentation of the distractor (e.g., *lace* and the presentation of the picture (of, e.g., a shoelace), the "stimulus-onset asynchrony" (SOA). Distractors can be phonologically, semantically, or morphologically similar to the target, and different kinds of distractors typically cause maximal interference at slightly different SOAs.

A third task, implicit priming, builds on participants' ability to learn pairs of word associations (e.g., *hand/foot, beach/sea, dog/cat*), where the idea is to use the associate (e.g., *hand*) to elicit the target (e.g., *foot*). For training, pairs of words are selected such that the target words either share some critical property (e.g., they might all begin with the same phoneme, the homogeneous condition) or do not share any property (the heterogeneous condition). During testing, only the associates are presented, and participants are requested to say the corresponding targets. Response measures are reaction time and accuracy. Implicit priming has as advantage that targets are no longer restricted to being picturable, but it has as its disadvantage that participants have to perform a rather strange associative memory task with little ecological validity.

Electroencephalography (EEG, the recording of electrical activity on the scalp) and magnetoencephalography (MEG, the recording of magnetic fields produced by electrical currents in the brain), and functional magnetic resonance imaging (fMRI, the mapping of brain areas with increased blood flow in response to experimental events) have made it possible to investigate the details of the time course of lexical processing, as well as the regions in the brain that subserve these processes. These experimental techniques can be combined with behavioral tasks (lexical decision with or without priming, see, e.g., Morris et al. (2007) as well as with eye-tracking (Dambacher and Kliegl, 2007) and picture naming Jescheniak and Levelt (1994). Electroencephalography and magnetoencephalography come with a high temporal resolution, whereas functional magnetic resonance imaging comes with high-quality information on localization. While these techniques have the obvious potential of providing detailed information about the temporal and spatial reflexes of linguistic processing in the brain, they also come with disadvantages.

One disadvantage of the neuroimaging approaches is methodological in nature. Especially in the case of fMRI, there are so many choice points in the course of data analysis that for any given study it can be entirely unclear whether results published as "significant" are actually obtained thanks to a fishing expedition in analytical parameter space (Haller and Bartsch, 2009; Carp, 2012; Eklund et al., 2012; Vul et al., 2009). For the analysis of EEG data, a serious disadvantage in the past has been that analytical methods were limited to repeated measures analysis of variance applied to selected time intervals in which researchers observed that the waveform for a violation condition diverged from the waveform for the corresponding control condition. As a consequence, the ecological validity of EEG studies using the violation paradigm for studying language processing is questionable: In natural language, ungrammatical or nonsensical words and sentences are extremely rare, whereas in many EEG studies, violations are highly frequent, and little is known about the strategies that subjects adopt to deal with the challenge of dealing with distorted language. Fortunately, recent advances in statistical analysis make it possible to study lexical processing under more natural circumstances (see, e.g., Hauk et al., 2006; Kryuchkova et al., 2012).

Moreover, linguistic research using EEG has focused primarily on a negative inflection in the averaged waveform around 400 ms post stimulus onset (the so-called N400) and a positive inflection around 600 ms post stimulus onset (the so-called P600). The N400 has been linked to semantic violations, and the P600 to syntactic violations. Some studies have taken this to indicate that words would be understood only after at least 400 ms post stimulus onset. However, words can be read at a rate of 5/second (Rayner, 1998), which makes it very unlikely that a word's meaning would become available only during the reading of following words (see, e.g., Kliegl et al., 2012; Segalowitz et al., 2009; Rubin and Turano, 1992). Another disadvantage coming especially with MEG and fMRI are the high costs associated with these techniques. For studies with no immediate medical benefit, and hence without the generous financial support typical for medical and clinical research, these high costs increase the pressure to publish, which in turn increase the risk of fishing expeditions and post-hoc explanations.

1.2 Selection of materials

As pointed out by Forster (2000), the materials going into many experimental studies are not selected randomly when researchers seek to match stimuli for lexical properties across experimental conditions. Often, researchers use their own knowledge of the language and experimental experience to accept certain items, and reject others. Researchers may have intuitions about what items might work, and which might not. The consequences of non-random stimulus selection is, from a statistical perspective, disastrous. First, results do not generalize beyond the items in the experiment. Second, the risk of replication failure is unnecessarily large. For this reason, the 'mega-studies' using thousands and even tens of thousands of words (Spieler and Balota, 1998; Balota et al., 1999, 2001, 2007; Lemhoefer et al., 2008; Ferrand et al., 2010; Keuleers et al., 2012, 2010) are extremely important: The risk of adverse effects of undocumented and undocumentable selection criteria is much reduced.

A further problem in experimental studies concerns the widespread practice of dichotomizing numeric variables. For instance, high-frequency words might be contrasted with low-frequency words. The problem here is that almost all lexical distributional variables are intercorrelated. Higher frequency words tend to be shorter, they tend to have more lexical neighbors that themselves tend to be more frequent, they tend to be composed of higher-frequency letters and letter pairs, they tend to have more meanings, and to occur in higher-frequency word sequences. Traditional studies depended heavily on analysis of variance, and hence sought to build binary contrasts in frequency (high versus low frequency) while matching on a subset of other lexical variables. It turns out that from a statistical perspective, this procedure has several severe disadvantages. First, statistical power is reduced (Baayen, 2010c): It becomes more difficult to ascertain that an effect is truly there. Second, the materials in an experiment are not a random sample, but a sample with very specific properties that run the risk of being atypical for the population. Third, matching constraints tend to severely reduce the number of items, sometimes to such an extent that matching criteria have to be relaxed in order to be able to run an experiment at all.

This brings us to the linguistic quality of the materials. In some studies, the necessity of having sufficient items in each experimental condition has led to the inclusion of words that from a linguistic perspective should not have been included. As a case in point, consider the influential study of Rastle et al. (2004). This study contrasts three sets of words: suffixed words such as *worker*, words containing a potential suffix but which are not morphologically complex such as *corner*, and a control group. In this study, *fruitful* is included in the pseudo-suffixed group along with *corner*. The rationale of these authors must have been that *fruitful* does not mean 'full of fruit'. However, the authors ignore that *fruitful* in the sense of 'successful' contrasts with *fruitless* ('unsuccesful') and that we can speak of *the fruits of one's labors* (see Baayen et al., 2011, for detailed discussion).

A final issue in psychology in general, and unfortunately psycholinguistics is no exception, is a strong publication bias. Experimental studies that failed to find an effect, as well as studies reporting a replication failure, tend not to be published. Even worse, studies replicating earlier work, without adding a newsworthy new finding of their own, are almost impossible to publish. As a consequence, far more significant results have been published than warranted by the alpha levels of the field (see, e.g., Ioannidis and A., 2007; Ioannidis, 2008; Francis, 2013). In other words, unfortunately, a fair proportion of studies report false positives.

1.3 Relevance for morphological theory

The research goal of theoretical morphology is often conceived of as providing a complete and insightful description of the internal structure of words within and across languages. Such descriptions typically aim for a balance between enumeration and analysis. Furthermore, such descriptions are either neutral with respect to the MODALITY of use (speaking, writing, reading, listening, signing), or they implicitly take a production perspective (especially in generative frameworks). Although experimental research on lexical processing is fraught with methodological difficulties, as outlined above, it nevertheless has the potential of enriching our understanding of how language in general, and morphology in particular, work.

First of all, it is worth noting that the processes of speaking, writing, reading and listening are very different. For instance, an experienced reader can process 5 words per second, whereas in auditory comprehension, 200 ms typically captures only part of a syllable. In production, one proceeds from the message to a carefully orchestrated sequence of articulatory gestures, whereas in auditory comprehension, the direction reverses, the task now being to map a highly variable speech signal onto meaning. Not only do speech production and auditory comprehension have very different time courses, they are also subject to different constraints. In auditory comprehension, the number of words compatible with the speech input (the competitors in the lexical cohort) reduces as the acoustic signal unfolds, whereas in speech production, the initial processing stages have to deal with semantically-driven competition (e.g., the selection between near-synonyms or between a hypernym and one of its hyponyms).

Furthermore, differences between individual language users may lead to remarkably different use of the possibilities offered by the grammar of 'the language'. It is well known that women tend to have slightly superior verbal skills compared to men (Kimura, 2000), and this difference extends to morphological processing. Ullman et al. (2002) and Hartshorne and Ullman (2006) observed a frequency effect for regular inflected words in English for women, but not for men. They interpret this finding within the declarative-procedural model of language (Ullman, 2004), which basically takes Bloomfield (1933)'s conception of the lexicon and maps it onto neural structures taken to subserve declarative memory (containing the unpredictable) and procedural memory (rule-based processing). Women, but not men, would then have a declarative memory containing even some higher-frequency regular complex word forms. The female/male split, however, is not this absolute. Various studies have replicated stronger frequency effects for regular complex words for females, but these studies also documented weaker, but still significant, frequency effects for males (Tabak et al., 2005, 2010; Balling and Baayen, 2008; Lemhoefer et al., 2008).

Differences between speakers may also arise as a consequence of differences in experience with language. Older speakers tend to know more words then younger speakers. The entropy of their vocabularies is greater than that of younger speakers. As a consequence, retrieving words from their mental lexicons requires more time (see, e.g., Baayen, 2008, , page 181). This in turn leads older speakers to rely more heavily on the use of pronouns (see for extensive discussion, Ramscar et al., 2013b).

The consequences of experience have recently been well documented for reading. For reasons of experimental convenience, most research on lexical processing is carried out using reading. However, the participants in experimental studies in psycholinguistics tend to be convenience samples from undergraduate students in psychology who are required to participate in experiments for course credit. As a consequence, the results in the published literature are strongly biased in that they describes the performance of predominantly highly-educated students of which a large majority is female (Francis et al., 2001; Sander and Sanders, 2006). This has not restrained researchers from drawing far-reaching conclusions about lexical processing in general and the architecture of the language faculty. However, when the population of readers is broadened to include students from vocational tracts, qualitatively very different patterns of reading are observed (see, e.g., Kuperman and Van Dyke, 2011, 2013, below, we will return to their work when discussing the balance of storage and computation in the processing of derived words.). Although individual differences are well-studied in (educational) psychology, for many years, many psycholinguists implicitly adopted the model of the ideal native speaker from generative linguistics, and had no interest whatsoever in individual variation. Fortunately, this is now changing.

Beyond task and individual differences, experimental studies are also of interest to morphological theory because they may provide evidence that supports or challenges the adequacy of the cognitive architectures posited by linguistic theories. For example, as mentioned above, it has been argued that the traditional distinction between rules and lexicon can be mapped onto procedural and declarative memory respectively (Ullman, 2004). However, instance, Ramscar and Gitcho (2007) offer a very different neural theory, contrasting implicit striatal learning for word forms with top-down control processes involving the pre-frontal cortex and the anterior cingulate cortex (see also Ramscar et al., 2013a). This alternative approach has far-reaching consequences for theories of the acquisition of morphologically complex words, as shown by Ramscar and Yarlett (2007). Or consider distributed morphology and the separation hypothesis (Halle and Marantz, 1993; Beard, 1995), according to which morphemes are no longer linguistic signs. Various studies have sought to demonstrate for comprehension that morpheme-forms are necessarily accessed before higher-order structures (Pylkkänen et al., 2004; Solomyak and Marantz, 2010). This approach in turn is challenged by studies indicating the involvement of semantics and higher-level knowledge at the earliest stages of lexical processing (Feldman et al., 2009; Kuperman et al., 2010). Discussions such as these have the potential of informing morphological theory about which of several formal architectures are more compatible with the experimental evidence.

Finally, experimental research may shed light on questions that remain unresolved within declarative theories. By way of example, consider phonaesthemes in English (e.g., *glow, glimmer, glare, glisten* where *gl* appears to refer to the emission or reflection of light. Since for any putative phonaestheme, there are many counterexamples (e.g., *glove, glue, glad*), it is hard to tell from the distributional data alone whether series such as *glow, glimmer, glare, glisten* have processing consequences similar to those of regular morphemes. Experimental studies by Bergen (2004) and Pastizzo and Feldman (2009) indicate, surprisingly, that there are indeed strong similarities with the processing of derived words.

2 The organization of the mental lexicon

Dictionaries for Indo-European languages such as English, French, and Greek are organized by entries which are ordered by a wordform as a key. In order to access the meaning of a word, this form key has to be found first, either by paging through a paper dictionary, or by entering the key into the search slot of an electronic dictionary. Once the relevant entry has been located, its contents become available.

Many theories take the organization of dictionaries as exemplary for the organization of the mental lexicon. Models of reading, for instance, typically assume that comprehension is a two-staged process. First, the word's form entry has to be identified. During this identification process, lexical competition takes place with similar word forms. Once access to the lexical form is completed, this form entry would then provide a pointer to the word's semantic and syntactic properties.

Other theories try to free themselves from the dictionary metaphor. These theories instead make use of the network metaphor. In network models, activation is claimed to spread from form units to semantic units, crucially without critical mediation by some form of 'dictionary' entry or units representing dictionary entries. In what follows, theories building on the dictionary metaphor are introduced first. Most work in psycholinguistics has been carried out within this general approach. Network theories, which seek to model lexical processing without a 'mental lexicon', are discussed next.

2.1 Theories building on the dictionary metaphor

2.1.1 Reading

A central question in research on the processing of complex words is whether morphological structure serves the purpose of facilitating lexical access, i.e., the identification of the proper form entry that provides access to semantics. Given a vocabulary of V entries, the complexity of finding an entry is $O(V)$ when a linear search is used, and $O(\log V)$ for a binary search. Whatever algorithm is used, a greater vocabulary implies an increased search problem. Now suppose that the V vocabulary items are grouped into F word families by their first constituent (e.g., all words beginning with *work*, such as *work, workable, workload, workbag, workbasket, worker, working, workings*, ..., would be in one word family), then the initial search complexity is reduced from roughly 50,000 to 15,000 (counts based on the CELEX lexical database, Baayen et al., 1995). Since word families tend to be small (in English, the median family size is 2, with a range of 1 to 187 for content words, and a range of 1 to 433 if prefixes are included), it has been proposed that finding the form entry can be speeded up by breaking down the search problem into an initial word family based search, followed by a second search in the much reduced search space of the word family itself. Early models of lexical access worked out this idea under the assumption of lexical searches being linear searches through frequency-ordered lists of entries (see, e.g., Taft and Forster, 1975, 1976; Knuth, 1973).

In this approach, two lexical-distributional measures have played an important role as diagnostic litmus tests. If the initial search takes place on the basis of the first constituent, then this search should be completed earlier the more frequent this constituent is. In linear search models, the assumption is that the constituents are ordered by frequency, with the highest frequency forms first in the list. Hence for higher frequency words, the number of search steps required is shorter, which predicts shorter processing times. In interactive activation models (McClelland and Rumelhart, 1981) — network models in which interconnected nodes excite or inhibit each other — higher-frequency words are assigned higher resting activation levels, allowing these words to be stronger competitors, which enables them to suppress similar words more quickly. Thus, the frequency of the first constituent becomes a diagnostic for lexical access taking place through morphological decomposition: The visual input is parsed into its constituents, of which the first is used as a pointer to its word family.

The second diagnostic is the frequency of the complex word itself. The more frequent the complex word is, the faster it should be accessible. In serial search models, this second frequency effect is accounted for by ordering the entries in the word families by frequency. In interactive activation models, nodes for constituents pass on activation to nodes representing whole words. These whole-word units of higher-frequency words are assigned higher resting activation levels, which allows them to reach threshold activation level more quickly than low-frequency complex words (see, e.g., Taft, 1991, 1994). Thus, constituent frequency effects are attributed to early morphological decomposition, whereas whole-word frequency effects are attributed to subsequent recombination (Taft, 2004), to look-up within word families, or to whole-word nodes that receive their activation from constituent nodes lower down in the interactive activation hierarchy.

An important property of this general approach is that the morphological parsing process is assumed to be blind and automatic. Whenever a potential base is encountered, it is assumed to be parsed out, and to serve as a key to a word family. Taft and Forster (1975); Taft (1981) argued that when prefixed words are read, the prefix is stripped off, and access proceeds on the basis of the stem. Since prefixes tend to have prefix families that are larger than the word families of their stems, prefix stripping is supposed to provide computational efficiency (Knuth, 1973) (actual corpus-based estimates suggest otherwise, however, see Schreuder and Baayen, 1994). Under blind decomposition, prefixes are also argued to be stripped off in unprefixed words such as *precipice* and *unique*. Taft and Forster (1975); Taft (1981) provide experimental evidence suggesting that for such pseudo-prefixed words, prefix stripping comes with a processing cost, as *cipice* and *ique* are not valid access keys to word families.

In the more recent literature, the role of blind obligatory decomposition has focused on pseudo-suffixed words such as *corner*, where a parse into a stem *corn* and a suffix *-er* is possible, but misleading. A large number of studies using the implicit priming paradigm have argued that *corn* is parsed out of *corner* just as *work* is parsed out of *worker*, as the amount of facilitation (with respect to an unrelated baseline) obtained by presenting *corner* and *worker* as primes for *corn* and *work* respectively was found to be equivalent, see Rastle et al. (2004); Rastle and Davis (2008); Kazanina (2011); Lavric et al. (2007, 2012), and also Devlin et al. (2004); Lewis et al. (2011); Solomyak and Marantz (2010). Studies using overt priming have reported similar results (see, e.g., Smolka et al., 2009). The evidence for obligatory decomposition is not unequivocal, however, as other studies reported evidence for truly affixed words having a processing advantage over pseudo-affixed words (Diependaele et al., 2005; Christianson et al., 2005; Feldman et al., 2009; Morris et al., 2007; Dunabeitia et al., 2011). Furthermore, an fMRI study by Bozic et al. (2007) suggests that brain regions with a reduced Bold response (i.e., regions with reduced oxygenation compared to an unrelated control condition) for form relations (*corner/corn*) are distinct from brain regions showing a reduced Bold response for semantic relations (*notion/idea*). Interestingly, both areas show a reduced Bold response for morphologically complex words (*boldly/bold*). Although the Bold response is slow, and does not provide information about the earliest stages of lexical processing, this pattern of results suggests that if indeed there is early morpho-orthographic parsing, it does not have long-lasting effects on lexical processing.

There are various reasons for the lack of consistency in the literature on the possible role of blind obligatory decomposition. First, from a linguistic perspective, stimuli selected as pseudo-affixed are semantically heterogenous (e.g., *fruitful*, as mentioned above, see Baayen et al., 2011, for detailed discussion). Second, the nature of the filler materials can influence the strategies used by subjects to meet the task requirements (Feldman et al., 2009). Third, whether pseudo-affixed words side with unrelated controls or with truly affixed words may possibly vary depending on participants' spelling skills and vocabulary size (Andrews and Lo, 2013).

Perhaps the most important problem with obligatory decomposition is why it would take place. The central functions of morphology are semantic (Lieber, 2004) and syntactic (Kastovsky, 1986), rather than to provide some efficient hash code for lexical access in reading. Prefix-stripping might seem advantageous, but when the distributional properties of languages such as English and Dutch are considered carefully, the disadvantages of obligatory decomposition outweight the advantages. A further complication is that the majority of derived words have idiosyncratic shades of meaning. For instance, a *worker* is, according to the online Merriam Websters, “one that works especially at manual or industrial labor or with a particular material”, “a member of the working class”, or “any of the sexually underdeveloped and usually sterile members of a colony of social ants, bees, wasps, or termites that perform most of the labor and protective duties of the colony”. Obligatory decomposition of *worker* into *work* and *-er*, combined with subsequent processes of compositional semantics, will never be able to reconstruct the conventionalized meanings of *worker* (see, e.g., Pham and Baayen, 2013; Kuperman et al., 2013b). The only way in which *work* and *-er* can be constructed to work properly is to construe them as hash codes for table look-up to the abovementioned meanings. Unfortunately, hash coding is an engineering solution that fails to predict semantic effects in lexical processing. Furthermore, there are non-morphological engineering solutions that perform better (e.g., letter trees, see Sproat, 1992).

Instead of assuming obligatory morpho-orthographic decomposition, Giraudo and Grainger (2001, 2003) have argued that all words have an orthographic access representation (the dictionary key to meaning) that is activated from the visual input. Once such an access representation reaches threshold activation (suppressing its competitors), the corresponding meaning or meanings become available. Effects of morphological constituents observed across a wide range of studies using primed and unprimed lexical decision as well as eye-tracking (see, e.g., Feldman, 2000; Kuperman et al., 2010; Bertram et al., 2000b; Burani et al., 1997; Laudanna and Burani, 1995; Laudanna et al., 1994; Burani and Caramazza, 1987; Miwa et al., 2013) are explained in this theory as arising due to post-access processes. Once the meaning of *goodness* is understood as “the quality or state of being good” (ignoring for ease of exposition its use as an interjection expressing mild surprise), activation would fan out to the meaning *good* and from there to the corresponding access representation. In other words, according to this supralexical theory, constituent effects arise as a consequence of having accessed a word’s meaning, instead of reflecting mediation by the constituents of the access to meaning.

Parallel dual route models present a hybrid of the obligatory decomposition theories and the supralexical theories. These models assume that form representations exist for both whole words and constituents. Two processes run in parallel and independently, a direct route and a parsing route. The first route to provide access to a word’s meaning is hypothesized to determine behavioral measures such as response latencies, as well as fixation durations. The parsing route operates on the access representations of the constituents, and attempts a combinatorial interpretation. The direct route operates on the word’s access representation, and makes use of a pointer from this access representation to the word’s semantics (Burani and Caramazza, 1987; Caramazza et al., 1988; Frauenfelder and Schreuder, 1992; Schreuder and Baayen, 1995). There are several reasons for positing a direct route in addition to a parsing route. First, a dual route system is more robust and more efficient than a single-route system (see, e.g., Baayen et al., 1997a). Second, the presence of whole-word access representations provides some protection against the many possible competing morphological parse trees that arise in the morpheme-driven route (Baayen and Schreuder, 2000). Third, dual route models are supported by eye-tracking studies indicating that first fixation durations are co-determined not only by constituent frequencies but also by whole-word frequencies (Kuperman et al., 2009; Miwa et al., 2013; Kuperman et al., 2008; Pollatsek et al., 2000).

A serious problem for dual route models is that the two routes appear not to work independently. Across several experiments, a tug of war between constituent measures and whole-word frequency has been observed. For instance, Kuperman et al. (2008) observed the effect of compound frequency to be strongest for the modifiers and heads with smaller word families, and Kuperman et al. (2009)

observed a similar interaction of compound frequency by modifier frequency. For Dutch derived words, the effect of whole-word frequency was modulated by suffix length (Kuperman et al., 2010). Such interactions are also present in lexical decision (Baayen et al., 2007). An attempt to address these kinds of interactions using probabilities defined over morphemes, complex words, and word families can be found in Kuperman et al. (2008). However, as more refined statistical methods that have become available for addressing numerical interactions in experimental data (Wood, 2006; Baayen et al., 2010) typically show even more complex patterns, they challenge explanations invoking a morphemic probability calculus. A further complication is that the tug of war between constituent and whole-word properties has been found to vary systematically between readers as a function of education level and reading skill. Skilled readers revealed strong lexical competition between whole words (*worker*) and base words (*work*), while poor readers received a processing advantage from higher-frequency base words (Kuperman and Van Dyke, 2011).

Kuperman et al. (2010) provide an example for Dutch of the complexities that arise when reading suffixed words in sentential context. Focusing on words with a single fixation, the duration of this fixation is co-determined by how far into the word the eye lands (landing positions that are too early or too late induce longer durations), by the length of the preceding word (the longer the preceding word, the longer the duration), and by the plausibility of the word in the sentence (the more plausible, the shorter the duration). Single fixation durations are also shorter for more frequent suffixed words. However, this effect decreases with increasing length of the suffix, and is totally absent for the longest suffixes (length 5). In other words, when the suffix has substantial support from the visual input, because it is long, the effect of word frequency disappears. Furthermore, for words with longer suffixes, processing costs increase with increasing imbalance of the morphological family sizes of base and suffix, as gauged with the Kullback-Leibler divergence (Milin et al., 2009a,b). The more the family size of the base and the family size of the suffix are similar, the shorter the fixation durations are. Dual-route theory does not provide predictions of this complexity, and it is unclear how it could be modified to do so.

The discussion thus far has addressed research investigating how "dictionary entries" are accessed from the visual input, with special attention to the role of the form representations for the whole word and its constituents. However, the paradigmatic relations between complex words within word families has also been found to have consequences for lexical processing. Here, it is useful to make a distinction between morphological family size, the type count of words in a word's morphological family (defined as the set of all words sharing that word as a constituent), and morphological family frequency (the summed frequencies of all complex words in the word family). Various studies on Dutch (Schreuder and Baayen, 1997; Bertram et al., 2000a; De Jong et al., 2000) indicate that the predictor relevant for predicting visual lexical decision latencies is the family size measure and not the family frequency measure (but see Ford et al., 2010). The processing advantage of words with large morphological families has been observed also for English (Baayen et al., 1997b; Feldman and Pastizzo, 2003; Pylkkänen et al., 2004; Baayen et al., 2007), as well as for Finnish and Hebrew (Moscoso del Prado Martín et al., 2005; Moscoso del Prado Martín et al., 2004).

The morphological family size effect is usually understood as a consequence of activation spreading in a network of connected dictionary entries from the base of a complex word to its family members. Within the multiple-readout framework of Grainger and Jacobs (1996), the co-activation of many family members provides evidence for a positive lexicality decision. Alternatively, if activation is allowed to resonate within a morphological family, this resonance can significantly boost the activation of the presented word, and hence afford shorter response latencies (De Jong et al., 2003).

The family size effect is semantic in nature. This is seen clearly in the results obtained for Hebrew. Moscoso del Prado Martín et al. (2005). They studied derived words with homonymic roots such as X-SH-B, which contribute to two semantic families, one involving concepts related to thinking (e.g., *Xa-SHaB* 'to think', *maXSHaBa* 'a thought', *XaSHiBa* 'thinking', and one relating to concepts involving arithmetic (e.g., *XiSHeB* 'to calculate', *XeSHBon* 'arithmetic', *XiSHuB* 'calculation'). Response latencies turn out to be sensitive to which semantic family (within the root family) a word belongs to. When a derived word from one semantic word family is read, response latencies decrease for increasing family size of that family, whereas response latencies increase for increasing family size of the other, semantically unrelated, root family members. In other words, the effect is sensitive not just to the presence of a shared consonantal root, but to the semantic fields supported by a given root.

In the reading research, there are several other lines of research relevant for the processing of derived words. To continue with Semitic, masked priming studies on Hebrew and Arabic show facilitation for prime-target pairs sharing the consonantal root, compared to unrelated controls (Deutsch et al., 1998; Frost et al., 1997, 2000a,b; Bentin and Frost, 2001; Boudelaa and Marslen-Wilson, 2001, 2004; Frost et al., 2005; Boudelaa et al., 2009). For Hebrew (Deutsch et al., 1998), but not for Arabic (Boudelaa and Marslen-Wilson, 2004), primes sharing the vowel pattern but not the consonantal root also facilitate responses. Reseach on lexical processing in Semitic interprets the experimental results on the processing of roots and vowel patterns as straightforward evidence for the cognitive reality of morphemes. Surprisingly, experimental psychologists seem to be unaware of alternative linguistic analyses of non-concatenative morphology such as proposed by Ussishkin (2005, 2006).

One of the striking properties of reading is that it is difficult, for words such as *anwser*, to detect misspellings consisting of letter transpositions. Within monomorphemic words, masked primes with a letter transposition have been found to be almost as effective as identity primes (Perea and Lupker, 2004). For complex words, letter transpositions within constituents are also nearly harmless, but transpositions at the morpheme boundary (e.g., *db* in *sandbank*) have been found to be disruptive in some studies (Christianson et al., 2005; Du abeitia et al., 2007; Lemhöfer et al., 2011), but not in others (Perea and Carreiras, 2006; Rueckl and Rimzhim, 2011). It remains at present unclear to what extent manipulation of the boundary bigram in complex words can serve as a diagnostic for morphological processing.

A final question in reading research asks whether morphology is more than just the coincidence of shared form and shared meaning? Feldman (2000) addressed this question and showed that in primed lexical decision, morphological effects exceeded the individual effects of semantic similarity and of form similarity. Her conclusions find support in recent neuroimaging studies (Bick et al., 2008; Boudelaa et al., 2009) which suggest that there are brain areas that are involved only when morphologically complex words are processed, and not for words related in only form or only meaning. It should be kept in mind, however, that similar conclusions are reached (Pastizzo and Feldman, 2009) for word pairs such as *boat-float* (semantically related and related in form), *swim-float* (only semantically related) and *coat-float* (only related in form). In other words, the sharing of form and meaning seems to be important, and not whether or not this sharing is brought about by means of affixation.

2.1.2 Listening

When reading, morphological processing of derived words is influenced by factors such as where the eye lands, how much of the suffix is visible, and the lengths and frequencies of the constituents. When listening, instead of information about large chunks of words becoming available simultaneously, the speech signal unfolds slowly over time. As a consequence of this slow temporal unfolding, the set of words compatible with the input is winnowed down as more information becomes available. For instance, after having heard just the first two segments of *houseful*, the word *hound* is still a viable continuation, but after having heard the third segment, only words beginning with *house*, including *house* itself, remain in the cohort of lexical competitors.

Marslen-Wilson (1996) proposed a definition of competitors based on a morphological breakdown of the lexicon. Only uninflected monomorphemic words were included, and derived prefixed words. On the basis of this set of potential lexical competitors, he defined a word's uniqueness point as the point in the speech signal at which all of a word's competitors have become incompatible with the speech input. Other things being equal, a word with a uniqueness point earlier in the word tends to be recognized more quickly.

Suffixed and compound words are not considered in cohort theory for two reasons. First, their inclusion would give rise to most words becoming unique after word offset, which would be self-defeating. Second, in the eighties, decompositional theories were dominant, and no experimental evidence was available on, for instance, the importance of whole-word frequency as an independent predictor of the processing complexity of regular complex words in auditory comprehension

Nevertheless, standard cohort has been shown to be too restrictive to be revealing about morphological processing. Following up on work by Wurm (1997), Wurm and Ross (2001), Wurm and Aycock (2003) and Balling and Baayen (2008), Balling and Baayen (2012) defined two uniqueness points. The initial uniqueness point (UP) is reached when morphologically unrelated competitors are no longer compatible with the speech input. The complex uniqueness point (CUP) is reached when morphologically related competitors drop from the cohort. Experiments using the auditory lexical decision task show that both uniqueness points predict shorter response latencies for words with earlier UP and CUP. Balling and Baayen (2012) proposed to understand uniqueness point effects as reflecting changes in surprisal (approximately, amount of information), in parallel to the way changes in surprisal predict processing costs in syntax (Levy, 2008). By the time a uniqueness point has been reached, most of the cognitive costs associated with weeding out unrelated competitors have accrued. As a consequence, subsequent processing can proceed more quickly.

One of the consequences of the distribution of the language signal in time (spoken) rather than space (written) is that morphological family size effects are less robust (see Balling and Baayen, 2012, for detailed discussion). Family size counts are insensitive to position: Complex words are counted irrespective of whether the targeted base word occurs in initial position. However, for auditory comprehension, order does matter. Although the family size count of *house* includes words such as *rehouse* and *roadhouse*, these family members have long dropped out of the cohort when listening to *house* itself. Balling's complex uniqueness point is therefore a more useful construct for gauging paradigmatic morphological structure, as the cohort of competitors that is active between the UP and the CUP consists of morphologically related words. Nevertheless, a family size effect was detected in the EEG signal, starting around 150 ms post stimulus onset, elicited in a normal listening task (with isolated words, but without a decision component) by Kryuchkova et al. (2012).

The uniqueness point effects fit well with the whole-word frequency effects observed across many experiments Meunier and Segui (1999b,a); Baayen et al. (2003, 2007); Balling and Baayen (2012) as well as with research on acoustic reduction (Schuppler et al., 2012), and point to a rich lexicon with semantic representations not only for monomorphemic words, but also for complex words. Interestingly, the computational model for auditory comprehension of Norris and McQueen (2008) is also based on such a rich lexicon. According to this model, the probabilities of form representations for both simple and complex words undergo continuous Bayesian updating as the speech signal unfolds. No specifically morphological processes are involved during listening. It is unclear, however, how this model would handle the understanding of novel complex words that are not in its lexicon.

For cohort theory, it is convenient to think of the speech signal as a series of discrete segments.

However useful for formulating theoretical constructs such as uniqueness points, the actual speech signal is much richer. For understanding morphological processing in auditory comprehension, the richness of fine phonetic detail is an important factor to be taken into account.

A comparison of the orthographic forms of *work* and *worker* would suggest that information about morphological complexity is carried exclusively by the suffix. However, in speech, the prosodic cues of the stem change when the suffix is added. For instance, syllable structure changes, and the stem becomes shorter. As a consequence, listeners can already anticipate upcoming morphological structure while listening to the stem (Kemps et al., 2005a,b).

Furthermore, in colloquial speech, complex words are often produced in highly reduced form. For instance, the Dutch adverb *eigenlijk*, [ɛɪxənlək], is often reduced to single-syllable [ɛxk] (Ernestus, 2000; Keune et al., 2005). Out of context, such strong reductions are difficult if not impossible to understand (Ernestus et al., 2002), whereas succesful understanding of a strongly reduced form appears to come with the percept of a much richer, more canonical, phonological form (Kemps et al., 2004). In some cases, the fine phonetic detail of the reduced form still reflects its polysyllabic origin (Niebuhr and Kohler, 2011), which may help guide the listener to the appropriate meaning.

Acoustic reductions of (complex) words, like the paradigmatic effects discussed in the previous section, challenge the usefulness of the dictionary metaphor for lexical processing. It is, of course, possible to enrich the lexicon with separate auditory access representations for reduced words, but such a move does not help explain why without context [ɛxk] does not activate any semantics. Given frequency effects observed for regular (non-idiomatic) word sequences (Bannard and Matthews, 2008; Arnon and Snider, 2010; Tremblay and Baayen, 2010; Tremblay et al., 2011), it might be argued that reduced forms are part of multiword templates, and that the mental dictionary should be broadened to a repository of both words and phrases. However, this could lead to hundreds of millions of additional entries for canonical n-word combinations (with $n < 5$) alone. A more dynamic approach with context-sensitive anticipation of the acoustic consequences of admissable articulatory shortcuts would not have this disadvantage (cf. Baayen et al., 2012).

2.1.3 Speaking

As holds for theories of language comprehension, models of speech production are also heavily influenced by the dictionary metaphor, with as major change a reversal in direction: The speaker, with some communicative goal in mind, has to find the right words to express herself. Current models posit nodes (representations) for word meanings, and link these meanings up to nodes for word forms, which in turn may link up to nodes for syllables and/or phonemes. The two main models in the literature differ in how activation is passed on from one node to the other. The interactive activation model of Dell (1986) posits both top-down and bottom-up links, which causes lexical processing to become highly interactive. In the WEAVER model of Levelt et al. (1999), connections are strictly top-down, from higher conceptual levels to lower levels of word forms and segments. In both models, various rules and checking mechanisms ensure that at the different levels the proper nodes are selected. What both models also have in common is that morphologically complex words are assumed to be constructed from their constituent morphemes.

In the model of Levelt et al. (1999), morphemes are not conceptualized as the smallest meaning-bearing unit, but as formal planning units. The complete separation of form and meaning, which fits well with the separation hypothesis of Beard (1977, 1981, 1995) as well as with distributed morphology (Halle and Marantz, 1993), is motivated by a series of experiments using the implicit priming paradigm. These experiments suggest that the semantic compositionality of a complex word is irrelevant for speech production: semantically transparent words such as *input* and semantically opaque words such as *invoice* exhibited a priming effect of similar magnitude that was much larger than for monomorphemic controls such as *insect* (Roelofs and Baayen, 2002). Using a long-distance priming paradigm with picture naming, Koester and Schiller (2011) reached the same conclusion, as did Lüttmann et al. (2011) using picture-word interference experiments. For Hebrew, however, Deutsch and Meir (2011) reported effects of morphology that did not reduce to the joint effects of form and semantic similarity.

Consistent with a strictly decompositional approach to speech production, Roelofs (1997) obtained a base frequency effect for particle verbs. On the other hand, experimental evidence for whole-word frequency effects in speech production is mixed. A picture naming study on plural inflection in Dutch failed to find a whole-word frequency effect (Baayen et al., 2008) which was well atttested for similar word materials in the same language for both reading and listening (Baayen et al., 1997a, 2003). Although Bien et al. (2005) observed a U-shaped frequency effect for compounds in a position-response association task, Bien et al. (2011) failed to find whole-word frequency effects for inflected and derived words. However, Tabak et al. (2010) observed effects of form frequency across several picture naming experiments with inflected verbs, and Janssen et al. (2008) found strong support for a whole-word frequency effect for compounds in picture naming in both English and Chinese. As pointed out by Janssen et al. (2008), it is quite possible that more natural tasks such as picture naming are better suited for the detection of whole-word frequency effects than associative memory tasks such as implicit priming or position-response association.

Further challenges to strictly decompositional models of speech production come from two sources. First, for inflected words, the entropy of the inflectional paradigm has been found to predict response latencies in both the picture naming (Baayen et al., 2008; Tabak et al., 2010) and the positional response association task (Bien et al., 2011). The inflectional entropy measure can be thought of as estimating the difficulty of choosing between different inflectional variants. The greater the inflectional entropy, the greater this difficulty is, and the longer response latencies become. Inflectional entropy effects show, albeit indirectly, that the speech production process is sensitive to the relative probabilities of inflected wordforms. Since strictly decompositional models have no representations for inflected wordforms, they cannot predict relative entropy effects.

A second challenge for strictly decompositional models of speech production comes from analyses of the speech signal. Whereas experimental studies in speech production typically work with response latencies, or with the consequences of priming manipulations on the electrophysiological response of the brain (Koester and Schiller, 2011), the phonetic record of what speakers have actually said is also highly informative (see, e.g., Gahl, 2008). Using a large speech corpus, Pluymaekers et al. (2005b) were able to show that the acoustic durations of prefixes and suffixes and/or the durations of segments in these prefixes and suffixes may be co-determined by the frequency of the derived words in which they occur. Pluymaekers et al. (2005a) showed, furthermore, that the acoustic realization of a suffix is co-determined by contextual factors such as the number of times the word was used in the preceding discourse, as well as its predictability from the preceding and following word. In addition, Tremblay and Tucker (2011) observed that the frequency with which combinations of four words occur co-determines acoustic duration. In the model of Levelt et al. (1999), which posits that the segments selected by morphemes are first bundled up into syllables, and which takes these syllable units to drive articulation, it is difficult to see how the frequency of a higher-order unit of a derived word (represented in the model only at higher conceptual and syntactic levels, but not at the wordform level), and contextual probabilities, might affect the articulatory execution of an affix. More in general, the WEAVER model of speech production is challenged by the accumulating evidence that a word's similarity neighborhood co-determines speech production (Munson and Solomon, 2004; Scarborough, 2004; Vitevitch, 2002).

A similar challenge comes from work on relative frequency. Hay (2003) distinguished between derived forms which are more frequent than their base words (e.g., *illegible, swiftly*) and those derived words for which the base is more frequent (e.g., *illiberal, softly*). Hay observed more t-deletion in English for derived words with a large relative frequency (*swiftly*, derived frequency > base frequency) than for words with a small relative frequency (*softly*, derived frequency < base frequency). A similar effect of relative frequency was reported for Dutch by Schuppler et al. (2012), but not in a reanalysis by Hanique and Ernestus (2012). If the effect of relative frequency turns out to receive more experimental support, it challenges full decomposition production models. By denying a role to whole-word representations for complex words, it becomes impossible to predict segment reduction from form frequency.

Thus far, we have considered the production of speech. Some results are also available for the production of writing. A large series of studies on typing in German, reviewed in Weingarten et al. (2004), investigated inter-keystroke intervals. For letter pairs spanning a morpheme boundary but not a syllable boundary, inter-keystroke intervals did not differ for non-morphological controls. However, when morpheme and syllable boundaries coincide, inter-keystroke intervals were found to be longer compared to controls with only a syllable boundary. Weingarten and colleagues also observed an effect of whole-word frequency, independently of base frequency. Kandel et al. (2012) compared, for handwriting, interletter pauses at the morpheme boundary for prefixed words and suffixed words in French, and compared them with pseudo-affixed controls. They only found a difference for suffixed words, from which they conclude that only suffixed words would be processed decompositionally.

2.2 Lexical processing without a mental lexicon

The theories and models reviewed thus far build on three important assumptions. First, they all accept without question that there are discrete lexical units for morphemes. Questions raised about the validity and usefulness of the morpheme as a theoretical construct, as raised by Matthews (1974); Uhlenbeck (1978); Anderson (1992); Blevins (2003); Stump (2001), have not entered into the awareness of most of the psycholinguistics community.

Second, the models formulated in this framework, irrespective of whether developed only as blueprints or computationally implemented, are declarative models that systematize a large body of knowledge, but, importantly, that do not learn. Irrespective of whether a dictionary theory works with interactive activation or with just a uni-directional flow of activation, the algorithms are designed to work in exactly the same way for a given word, irrespective of how many times that word (and other words in its context) have been encountered.

Third, these models work with a highly idealized and simplified view of the relation between form and meaning. Here, several issues come into play. First, from a linguistic perspective, it makes sense to distinguish between the skeleton and body of a word's meaning (Lieber, 2004), where the skeleton denotes the language structural scaffolding that supports the body, the rich encyclopedic knowledge that is part of a word's meaning. Thus, returning to the above example of *worker*, the skeleton is (simplified) "a subject noun derived from the verb *to work*", whereas part of the body is that the word denotes a particular kind of bee. It is important to realize that theories of lexical processing have to explain, for instance, how a listener comes to a proper understanding of a sentence such as "In the warm afternoon sun, we could see many workers collecting honey". Since no rule can reconstruct the meaning 'bee' from *work* and *-er*, decompositional theories of comprehension can only provide access to the skeleton, but not to the body. Similarly, decompositional theories of production cannot account for the longer acoustic duration of *worker* in the low-frequency sense of 'honey bee' compared to the high-frequency sense of 'participant on the industrial labor market' (cf. Gahl, 2008). Furthermore, it is not the case that in comprehension, the skeleton is accessed first, subsequently to be enriched with its body. Evidence is accumulating that rich information about the body plays an important role already during the earliest stages of comprehension (see Elman, 2009, and references cited there). Finally, words don't have or carry meanings (Ramscar and Baayen, 2013, see) — it is only thanks to the context in which a word occurs that they come into their own (recall, for instance, that strongly reduced derived words, for which we do not have orthographic awareness, are not interpretable out of context, see Ernestus et al., 2002; Kemps et al., 2004).

Two kinds of approaches have been pursued for understanding lexical processing without mediation by form entries for words or morphemes. Both take learning very seriously. Distributed connectionist models (Harm and Seidenberg, 2004; Joanisse and Seidenberg, 1999; Bird et al., 2003; Seidenberg and Gonnerman, 2000; Moscoso del Prado Martín et al., 2003; McClelland and Elman, 1986; Norris, 1994; Moscoso del Prado Martín, 2003) seek to explain morphological effects in the experimental literature as an emergent property of a processing architecture with three interacting banks of units: a bank of orthographic feature units, a bank of phonological feature units, and a bank of semantic feature units. In the TRIANGLE MODEL (Harm and Seidenberg, 2004), each of these banks of units is connected through intervening banks of hidden units. These hidden units serve a dual purpose: They allow for compression of statistical regularities between form and meaning, and as a consequence similarities in patterns of activations over hidden units can come to resemble generalizations over the input space. Seidenberg and Gonnerman (2000); Plaut and Gonnerman (2000); Gonnerman and Anderson (2001) made use of distributed connectionist models to explain processing advantages in priming studies of derived words (*boldly - bold*) vis-a-vis orthographic (*corner–corn*) and semantic (*idea–notion*) controls as arising due to the convergence of form and meaning.

Although distributed connectionist models are learning models, the learning algorithm used, back-propagation, has been criticized for being psychologically and neurobiologically implausible (Crick, 1989; Murre et al., 1992; OReilly, 1998, 2001). Furthermore, designing such models involves choices about the number of banks of hidden units, the numbers of units in the different banks, and the featural representations chosen for orthography, phonology, and semantics. A further criticism of these kind of models is that the behavior of any given model requires detailed statistical analysis of the banks of hidden units.

An alternative to distributed connectionist models is the naive discrimination learning (NDL) model (Baayen et al., 2011). The network structure of an NDL model is extremely simple: the nodes in a first layer of cues are linked up to the nodes in a second layer of outcomes. There are no hidden layers, and both cues and outcomes are straightforward symbolic represenations (e.g., cue nodes for letters and letter pairs, and outcome nodes for meanings). The weights on the links from cues to outcomes are estimated from the equilibrium equations of Danks (2003) for the learning equations developed by Wagner and Rescorla (1972). This makes it possible to estimate the connection weights from large corpora with hundreds of millions or even billions of words. Given the weights, the activation of an outcome is obtained by summation over the weights from the cues in the input to that outcome. The activation of a meaning outcome reflects how well that meaning can be learned given the words, their orthographic forms, and their meanings, in the language as sampled by the corpus.

Thus far, NDL modeling studies are available only for reading. The model of Baayen et al. (2011) comprises a network trained on a quarter of the British National Corpus, using letter unigrams and bigrams as input cues, and symbolic representations for meanings (e.g., 'work' and 'agent' for *worker*) as outcomes. At the level of semantics, the model therefore is a full decomposition model. Interestingly, the model correctly captures whole-word frequency effects, morpheme frequency effects, and family size effects observed with the visual lexical decision task, even though there are no representations for whole words, morphemes, or morphological families in the model's architecture. Although distributional measures such as word frequency, constituent frequency, and family size, are often interpreted as diagnostic measures for cognitive representations for whole words, for constituents, and links between morphologically related words, the NDL model shows that a very different interpretation is possible, an interpretation in which these effects reflect learnability.

It is important to keep in mind that the NDL model is not a model of the full reading process. To the contrary, the model captures only the very first stage of the reading process, namely the activation of meaning from low-level visual information (represented in the model by letter unigrams and bigrams). Often, more than one fixation will be necessary for understanding a complex word, and the higher-order cognitive processes further guiding interpretation (Yeung et al., 2006; Ramscar and Gitcho, 2007) constitute an essential part of reading that is not captured by the NDL (see Kuperman et al., 2013b).

The NDL model has thus far been applied not only to English but also to Serbian and to Hebrew. The modeling results of Baayen (2012) suggest that skilled reading of Hebrew may not require a non-concatenative decomposition into morphemes as argued for by (McCarthy, 1981). Theoretically, the NDL model is much closer to the phonotactic approach of Ussishkin (2005, 2006).

Naive discrimination learning makes a prediction concerning the role of infrequent phoneme sequences straddling morpheme boundaries that is exactly opposite to what connectionist models as well as symbolic models with morphemic decomposition predict. Hay (2002, 2003) argued that infrequent letter bigrams straddling a morpheme boundary (e.g., *tl* in *swiftly*) would make the complex word more parseable. Likewise, Seidenberg (1987) argued that for a connectionist network to learn word-specific meanings, higher-frequency boundary digraphs are required. As a consequence, words with low-frequency boundary digraphs would depend more on the mappings of form to meaning in the stem and in the affix, thereby giving rise to processing effects (in the network) that in a symbolic framework would be understood as the effects of parsing. By contrast, in naive discrimination learning, the lower the frequency of a boundary bigram is, the better its cue value becomes for the complex word's own meaning (see Kuperman et al., 2013b, for a modeling study in which derived words and compounds have their own meaning outcomes). Above, we distinguished, following (Lieber, 2004), between the skeleton and the body of a word's meaning. Specifically, accessing the body, a word's idiosyncratic senses such as 'bee' for *worker*, depends in the NDL model on the boundary bigram. To see this, consider the letter bigram *qa* that appears in the scrabble word *qaid*. As long as this is the only word with *qa* known to a reader, the presence of *qa* is a perfect cue to *qaid*. However, the more other words with *qa* exist in a speaker's lexicon (e.g., *qanat*), the less good *qa* is as a cue for *qaid*. In the same way, boundary bigrams that have a low frequency, indicating that they occur in relatively few other words, have a high cue validity for those words. The other side of the same coin is that a low-frequency bigram does not interfere negatively during learning with the activation of the meaning of the base, hence for consonant-initial suffixes, base frequency effects are more likely to be detected (for experimental evidence, see Järvikivi and Pyykkönen, 2011; Vannest et al., 2011).

The re-evaluation of the functionality of low-frequency digraphs in reading suggested by the NDL approach may also shed light on the comprehension of highly-reduced derived words such as [ɛxk] for [ɛɪxənlək]. Whenever acoustic reduction results in rare sequences of segments (such as [xk]), these sequences become excellent cues to meaning (see Baayen, 2010a, for simulations and detailed discussion).

The hypothesis of complexity-based ordering for English derivational suffixes (Hay, 2003; Hay and Plag, 2004; Plag and Baayen, 2009; Baayen, 2010b) is also challenged by the NDL approach. Hay's original hypothesis was that suffixes that are more parsable must occur outside of suffixes that are less parsable. It is not self-evident why a parsability constraint of this form should be in force. An alternative description tapping into the same phenomenon is that productivity decreases as one moves from the right edge to the stem (Krott et al., 1999). Since the more productive suffixes tend to be consonant-initial, these suffixes are more likely to create low-frequency boundary diphones/digraphs, which, if the NDL approach is on the right track, would make these words easier to understand. That is, from an onomasiological perspective, consonant-initial suffixes would create words that are more memorable, and hence have higher probabilities of becoming entrenched in the language.

Possibly, stem and suffix allomorphy likewise enhance the discriminability of the different meanings indexed by combinations of stems and affixes (for detailed experimental studies of allomorphy, see Järvikivi and Niemi, 2002; Järvikivi et al., 2006).

Naive discrimination learning also offers a new perspective on the interactions that often emerge in regression studies of lexical decision and eye-tracking (e.g. Kuperman et al., 2010; Miwa et al., 2013) between measures such as, for instance, whole-word frequency and base frequency. For visual lexical decision, Baayen et al. (2007) observed the strongest effect of whole-word frequency for words with the lowest base frequencies. Conversely, the effect of base frequency was facilitatory for words with low whole-word frequencies, but inhibitory for words with high whole-word frequencies. Within interactive activation frameworks, this pattern of results suggests a tug of war between whole word and base. However, in a learning approach, this tug of war unfolds during the (continuously ongoing) learning process, with cues competing for meanings. Importantly, in real time during reading, there is no actual competition between the meanings of the derived word and its stem, at least during the initial stages of visual processing, as in the NDL model there is just a single forward pass of activation from the orthographic cues to the semantics.

3 Concluding remarks

There are several critical challenges for research on lexical processing for the coming years. First, the field needs corpora that come closer to actual language experience. Corpora of what people actually say, hear, and read, would be ideal, but since such corpora are prohibitively expensive to develop, corpora using film subtitles are a good approximation (for empirical evidence, see, e.g., Brysbaert and New, 2009). The reason for subtitle corpora working well probably is that they approximate more accurately the colloquial use of everyday spoken language.

Second, it will be important to move away from the lexical decision task, as it may tell us more about a metalinguistic judgement task than about actual language processing. However, as psychologists have discovered crowd sourcing and have developed apps for smartphones that can easily harvest millions of lexical decisions, this method will become more instead of less popular in the coming years.

Third, to advance the field, computational implementation is essential. The verbal models of the last 40 years (prelexical decomposition, postlexical decomposition, dual-route models) fail to predict the complex patterns present in the experimental data. If language shares essential properties with complex dynamic systems, which is what the experimental data suggest, then linguistics and psycholinguistics will need to start using the tools and techniques developed in other domains of scientific inquiry for studying complex dynamic systems, and to give up the static dictionary metaphor that still guides many current models of lexical processing.

Fourth, current research on the processing of derived words (and of lexical processing in general) is typologically severely limited, with strong research traditions restricted to selected Indo-European languages, to Semitic, to Finnish, and to Chinese and Japanese. In all these cases, we are dealing with societies with long traditions of literacy, and with experimental research with a strong bias for the study of reading. It goes without saying that the generality of the results reviewed in this chapter is severely limited by this bias. Finally, the consequences of continued learning throughout the life time and the concomitant accumulation of knowledge (including lexical knowledge) has profound consequences for individual differences in language processing (Ramscar et al., 2013b). The field will need to abandon convenience sampling of university students, and to commit to sampling from broader cross-sections of the population if we are to obtain a realistic view of how language really works in our societies.

References

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

Andrews, S. and Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, 68:279–296.

Arnon, I. and Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1):67–82.

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

Baayen, R. H. (2010a). Assessing the processing consequences of segment reduction in Dutch with naive discriminative learning. *Lingue & Linguaggio*, 9:95–112.

Baayen, R. H. (2010b). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In Olsen, S., editor, *New impulses in word-formation (Linguistische Berichte Sonderheft 17)*, pages 383–402. Buske, Hamburg.

Baayen, R. H. (2010c). A real experiment is a factorial experiment? *The Mental Lexicon*, 5(1):149–157.

Baayen, R. H. (2012). Learning from the Bible: computational modelling of the costs of letter transpositions and letter exchanges in reading classical hebrew and modern english. *Lingue e Linguaggio*, 11(2):123–146.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997a). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, 36:94–117.

Baayen, R. H., Hendrix, P., and Ramscar, M. (2012). Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. *Language and Speech*, page in press.

Baayen, R. H., Kuperman, V., and Bertram, R. (2010). Frequency effects in compound processing. In Scalise, S. and Vogel, I., editors, *Compounding.* Benjamins, Amsterdam/Philadelphia.

Baayen, R. H., Levelt, W., Schreuder, R., and Ernestus, M. (2008). Paradigmatic structure in speech production. *Proceedings Chicago Linguistics Society 43*, 1:1–29.

Baayen, R. H., Lieber, R., and Schreuder, R. (1997b). The morphological complexity of simplex nouns. *Linguistics*, 35:861–877.

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., Milin, P., Filipovic Durdjevic, D., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3):438–481.

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. (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. *The Mental Lexicon*, 2:419–463.

Balling, L. and Baayen, R. (2012). Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition*, 125:80–106.

Balling, L. and Baayen, R. H. (2008). Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes*, 23:1159–1190.

Balota, D., Cortese, M., and Pilotti, M. (1999). Visual lexical decision latencies for 2906 words. *[On-line], Available: http://www.artsci.wustl.edu/∼dbalota/lexical_decision.html*.

Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D., and Yap, M. (2004). Visual word recognition for single-syllable words. *Journal of Experimental Psychology:General*, 133:283–316.

Balota, D. A., Pilotti, M., and Cortese, M. J. (2001). Subjective frequency estimates ofr 2,938 monosyllabic words. *Memory & Cognition*, 29:639–647.

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.

Bannard, C. and Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19:241–248.

Beard, R. (1977). On the extent and nature of irregularity in the lexicon. *Lingua*, 42:305–341.

Beard, R. (1981). On the question of lexical regularity. *Journal of Linguistics*, 17:31–37.

Beard, R. (1995). *Lexeme-morpheme base morphology: A general theory of inflection and word formation*. State University of New York Press, Albany, NY.

Bentin, S. and Frost, R. (2001). Linguistic theory and psychological reality: a reply to Boudelaa and Marslen-Wilson. *Cognition*, 81(1):113–118.

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, 80:290–311.

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

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*, 26:419–511.

Bick, A., Goelman, G., and Frost, R. (2008). Neural correlates of morphological processes in Hebrew. *Journal of cognitive neuroscience*, 20(3):406–420.

Bien, H., Baayen, R., and Levelt, W. (2011). Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes*, 27:683–715.

Bien, H., Levelt, W., and Baayen, R. (2005). Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the USA*, 102:17876–17881.

Bird, H., Lambon Ralph, M., Seidenberg, M., McClelland, J., and Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory and Language*, 48:502–526.

Blevins, J. P. (2003). Stems and paradigms. *Language*, 79:737–767.

Bloomfield, L. (1933). *Language*. Allen and Unwin, London.

Boudelaa, S. and Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, 81(1):65–92.

Boudelaa, S. and Marslen-Wilson, W. D. (2004). Abstract morphemes and lexical representation: the cv-skeleton in arabic. *Cognition*, 92(3):271–303.

Boudelaa, S., Pulvermüller, F., Hauk, O., Shtyrov, Y., and Marslen-Wilson, W. (2009). Arabic morphology in the neural language system. *Journal of cognitive neuroscience*, 22(5):998–1010.

Bozic, M., Marslen-Wilson, W. D., Stamatakis, E. A., Davis, M. H., and Tyler, L. K. (2007). Differentiating morphology, form, and meaning: Neural correlates of morphological complexity. *Journal of cognitive neuroscience*, 19(9):1464–1475.

Brysbaert, M. and New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4):977–990.

Burani, C. and Caramazza, A. (1987). Representation and processing of derived words. *Language and Cognitive Processes*, 2:217–227.

Burani, C., Dovetto, M., Thornton, A. M., and Laudanna, A. (1997). Accessing and naming suffixed pseudo-words. In Booij, G. E. and Van Marle, J., editors, *Yearbook of Morphology 1996*, pages 55–72. Kluwer, Dordrecht.

Caramazza, A., Laudanna, A., and Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28:297–332.

Carp, J. (2012). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fmri experiments. *Frontiers in Neuroscience*, 6:1–13.

Christianson, K., Johnson, R., and Rayner, K. (2005). Letter transpositions within and across morphemes. *Journal of Experimental Psychology: Learning Memory and Cognition*, 31(6):1327–1339.

Crick, F. H. C. (1989). The recent excitement about neural networks. *Nature*, 337:129–132.

Dambacher, M. and Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain research*, 1155:147–162.

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *Journal of Mathematical Psychology*, 47(2):109–121.

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.

Dell, G. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, 93:283–321.

Deutsch, A., Frost, R., and Forster, K. I. (1998). Verbs and nouns are organized and accessed differently in the mental lexicon: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24:1238–1255.

Deutsch, A. and Meir, A. (2011). The role of the root morpheme in mediating word production in hebrew. *Language and Cognitive Processes*, 26(4-6):716–744.

Devlin, J. T., Jamison, H. L., Matthews, P. M., and Gonnerman, L. M. (2004). Morphology and the internal structure of words. *Proceedings of the National Academy of Sciences of the United States of America*, 101(41):14984–14988.

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

Du abeitia, J. A., Perea, M., and Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? evidence for morpho-orthographic decomposition. *Cognition*, 105(3):691–703.

Dunabeitia, J., Kinoshita, S., Carreiras, M., and Norris, D. (2011). Is morpho-orthographic decomposition purely orthographic? evidence from masked priming in the same–different task. *Language and Cognitive Processes*, 26(4-6):509–529.

Eklund, A., Andersson, M., Josephson, C., Johannesson, M., and Knutsson, H. (2012). Does parametric fMRI analysis with SPM yield valid results? an empirical study of 1484 rest datasets. *NeuroImage*.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive science*, 33(4):547–582.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface.* LOT, Utrecht.

Ernestus, M., Baayen, R. H., and Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81:162–173.

Feldman, L. and Pastizzo, M. (2003). Morphological facilitation: The role of semantic transparency and family size. In Baayen, R. H. and Schreuder, R., editors, *Morphological Structure in Language Processing*, pages 233–258. Mouton de Gruyter, Berlin.

Feldman, L. B. (2000). Are morphological effects distinguishable from the effects of shared meaning and shared form? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6):1431–1444.

Feldman, L. B., O'Connor, P. A., and Moscoso del Prado Martin, F. (2009). Early morphological processing is morpho-semantic and not simply morpho-orthographic: Evidence from the masked priming paradigm. *Psychonomic Bulletin & Review*, 16(4):684–691.

Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., and Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42(2):488–496.

Ford, M. A., Davis, M. H., and Marslen-Wilson, W. D. (2010). Derivational morphology and base morpheme frequency. *Journal of Memory and Language*, 63(1):117–130.

Forster, K. (1999). The Microgenesis of Priming Effects in Lexical Access. *Brain and Language*, 68(1-2):5–15.

Forster, K. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, 28:1109–1115.

Francis, B., Robson, J., and Read, B. (2001). An analysis of undergraduate writing styles in the context of gender and achievement. *Studies in Higher Education*, 26(3):313–326.

Francis, G. (2013). Publication bias in "Red, Rank, and Romance in Women Viewing Men" by Elliot et al. (2010). *Journal of Experimental Psychology: General*, 142(1):292–296.

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.

Frost, R., Deutsch, A., and Forster, K. (2000a). Decomposing morphologically complex words in a nonlinear morphology. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3):751–765.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., and Marslen-Wilson, W. D. (2000b). Morphological priming: Dissotiation of phonological, semantic and morphological factors. *Memory and Cognition*, 28(8):1277–1288.

Frost, R., Forster, K. I., and Deutsch, A. (1997). What can we learn from the morphology of hebrew? A masked priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23:829–856.

Frost, R., Kugler, T., Deutsch, A., and Forster, K. (2005). Orthographic structure versus morphological structure: Principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6):1293–1326.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3):474–496.

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

Giraudo, H. and Grainger, J. (2003). On the role of derivational affixes in recognizing complex words: Evidence from masked priming. In Baayen, R. and Schreuder, R., editors, *Morphological Structure in Language Processing*, pages 211–234. Mouton, Berlin.

Gonnerman, L. and Anderson, E. (2001). Graded semantic and phonological similarity effects in morphologically complex words. In Bendjaballah, S., Dressler, W., Pfeiffer, O. E., and Voeikova, M. D., editors, *Morphology 2000: Selected papers from the 9th Morphology meeting*, pages 137–148. John Benjamins, Amsterdam.

Grainger, J. and Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103:518–565.

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.

Haller, S. and Bartsch, A. J. (2009). Pitfalls in fMRI. *European Radiology*, 19(11):2689–2706.

Hanique, I. and Ernestus, M. (2012). The role of morphology in acoustic reduction. *Lingue e Linguaggio*, 11(2):147–164.

Harm, M. W. and Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, 111:662–720.

Hartshorne, J. K. and Ullman, M. T. (2006). Why girls say 'holded' more than boys. *Developmental Science*, 9:21–32.

Hauk, O., Davis, M., Ford, M., Pulvermller, F., and Marslen-Wilson, W. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage*, 30:1383–1400.

Hay, J. B. (2002). From speech perception to morphology: Affix-ordering revisited. *Language*, 78:527–555.

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

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.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, 19:640–648.

Ioannidis, J. P. A. and A., T. T. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, 4:245–253.

Janssen, N., Bi, Y., and Caramazza, A. (2008). A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes*, 23(7-8):1191–1223.

Järvikivi, J., Bertram, R., and Niemi, J. (2006). Affixal salience and the processing of derivational morphology: The role of suffix allomorphy. *Language and cognitive processes*, 21(4):394–431.

Järvikivi, J. and Niemi, J. (2002). Allomorphs as paradigm indices: On-line experiments with finnish free and bound stems. *SKY journal of linguistics*, (15):119–143.

Järvikivi, J. and Pyykkönen, P. (2011). Sub-and supralexical information in early phases of lexical access. *Frontiers in psychology*, 2.

Jescheniak, J. D. and Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20(4):824–843.

Joanisse, M. F. and Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences*, 96:7592–7597.

Kandel, S., Spinelli, E., Tremblay, A., Guerassimovitch, H., and Álvarez, C. J. (2012). Processing prefixes and suffixes in handwriting production. *Acta psychologica*, 140(3):187–195.

Kastovsky, D. (1986). Productivity in word formation. *Linguistics*, 24:585–600.

Kazanina, N. (2011). Decomposition of prefixed words in Russian. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6):1371.

Kemps, R., Ernestus, M., Schreuder, R., and Baayen, R. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 19:117–127.

Kemps, R., Ernestus, M., Schreuder, R., and Baayen, R. (2005a). Prosodic cues for morphological complexity: The case of Dutch noun plurals. *Memory and Cognition*, 33:430–446.

Kemps, R., Wurm, L. H., Ernestus, M., Schreuder, R., and Baayen, R. (2005b). Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes*, 20:43–73.

Keuleers, E. and Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, 42(3):627–633.

Keuleers, E., Diependaele, K., and Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono-and disyllabic words and nonwords. *Frontiers in Psychology*, 1.

Keuleers, E., Lacey, P., Rastle, K., and Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior research methods*, 44(1):287–304.

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.

Kimura, D. (2000). *Sex and Cognition.* The MIT Press, Cambridge, MA.

Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., and Sommer, W. (2012). Eye movements and brain electric potentials during reading. *Psychological research*, 76(2):145–158.

Knuth, D. E. (1973). *The Art of Computer Programming. Vol. 3: Sorting and Searching.* Addison-Wesley, Reading, Mass.

Koester, D. and Schiller, N. O. (2011). The functional neuroanatomy of morphology in language production. *NeuroImage*, 55(2):732–741.

Krott, A., Schreuder, R., and Baayen, R. H. (1999). Complex words in complex words. *Linguistics*, 37(5):905–926.

Kryuchkova, T., Tucker, B. V., Wurm, L., and Baayen, R. H. (2012). Danger and usefulness in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, 122:81–91.

Kuperman, V., Bertram, R., and Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, 23:1089–1132.

Kuperman, V., Bertram, R., and Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *Journal of Memory and Language*, 62:83–97.

Kuperman, V., Drieghe, D., Keuleers, E., and Brysbaert, M. (2013a). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, page in press.

Kuperman, V., Ramscar, M., Shaoul, C., and Baayen, R. (2013b). Decomposition makes things worse: A discrimination learning approach to the time course of understanding compounds in reading. *Manuscript, McMaster University & University of Tuebingen.*

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2009). Reading of multimorphemic Dutch compounds: Towards a multiple route model of lexical processing. *Journal of Experimental Psychology: HPP*, 35:876–895.

Kuperman, V. and Van Dyke, J. (2011). Individual differences in visual comprehension of morphological complexity. In Carlson, L., Hoelscher, C., and Shipley, T., editors, *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, pages 1643–1648. Cognitive Science Society, Austin, TX.

Kuperman, V. and Van Dyke, J. (2013). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language*, 65(1):42–73.

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.

Laudanna, A., Burani, C., and Cermele, A. (1994). Prefixes as processing units. *Language and Cognitive Processes*, 9:295–316.

Lavric, A., Clapp, A., and Rastle, K. (2007). ERP evidence of morphological analysis from orthography: A masked priming study. *Journal of Cognitive Neuroscience*, 19(5):866–877.

Lavric, A., Elchlepp, H., and Rastle, K. (2012). Tracking hierarchical processing in morphological decomposition with brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, 38(4):811.

Lemhoefer, K., Dijkstra, A., Schriefers, H., Baayen, R., Grainger, J., and Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34:12–31.

Lemhöfer, K., Koester, D., and Schreuder, R. (2011). When bicycle pump is harder to read than bicycle bell: Effects of parsing cues in first and second language compound reading. *Psychonomic bulletin & review*, 18(2):364–370.

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.

Lewis, G., Solomyak, O., and Marantz, A. (2011). The neural basis of obligatory decomposition of suffixed words. *Brain and language*, 118(3):118.

Lieber, R. (2004). *Morphology and Lexical Semantics*. Cambridge University Press.

Lüttmann, H., Zwitserlood, P., Böhl, A., and Bölte, J. (2011). Evidence for morphological composition at the form level in speech production. *Journal of Cognitive Psychology*, 23(7):818–836.

Marslen-Wilson, W. D. (1996). Function and process in spoken word recognition. In *Attention and performance: Control of language processes*, volume X, pages 125–150. Lawrence Erlbaum Associates, Hillsdale, NJ.

Matthews, P. H. (1974). *Morphology. An Introduction to the Theory of Word Structure*. Cambridge University Press, London.

McCarthy, J. J. (1981). A prosodic theory of non-concatenative morphology. *Linguistic Inquiry*, 12:373–418.

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

McClelland, J. L. and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of the basic findings. *Psychological Review*, 88:375–407.

Meunier, F. and Segui, J. (1999a). Frequency effects in auditory word recognition: The case of suffixed words. *Journal of Memory and Language*, 41:327–344.

Meunier, F. and Segui, J. (1999b). Morphological priming effect: The role of surface frequency. *Brain and Language*, 68(1):54–60.

Milin, P., Filipović Durdević, D., and Moscoso del Prado Martín, F. (2009a). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, pages 50–64.

Milin, P., Kuperman, V., Kostić, A., and Baayen, R. (2009b). Paradigms bit by bit: An information-theoretic approach to the processing of paradigmatic structure in inflection and derivation. In Blevins, J. P. and Blevins, J., editors, *Analogy in grammar: form and acquisition*, pages 214–252. Oxford University Press, Oxford.

Miwa, K., Libben, G., Dijkstra, T., and Baayen, R. (2013). The time-course of lexical activation in Japanese morphographic word recognition: Evidence for a character-driven processing model. *Quarterly Journal of Experimental Psychology*, in press.

Morris, J., Frank, T., Grainger, J., and Holcomb, P. J. (2007). Semantic transparency and masked morphological priming: An ERP investigation. *Psychophysiology*, 44(4):506–521.

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

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.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., and Baayen, R. H. (2005). Changing places: A cross-language perspective on frequency and family size in Hebrew and Dutch. *Journal of Memory and Language*, 53:496–512.

Moscoso del Prado Martín, F., Ernestus, M., and Baayen, R. H. (2003). Do type and token effects reflect different mechanisms: Connectionist modelling of dutch past-tense formation and final devoicing. *Brain and Language*, 90:287–298.

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

Murre, J. M. J., Phaf, R. H., and Wolters, G. (1992). Calm: Categorizing and learning module. *Neural Networks*, 5:55–82.

Niebuhr, O. and Kohler, K. J. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics*, 39(3):319–329.

Norris, D. and Kinoshita, S. (2008). Perception as evidence accumulation and bayesian inference: Insights from masked priming. *Journal of Experimental Psychology*, 137(3):434–455.

Norris, D. and McQueen, J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2):357–395.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52:189–234.

OReilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in Cognitive Science*, 2:455–462.

OReilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and hebbian learning. *Neural Computation*, pages 1199–1242.

Pastizzo, M. J. and Feldman, L. B. (2009). Multiple dimensions of relatedness among words: Conjoint effects of form and meaning in word recognition. *The mental lexicon*, 4(1):1.

Perea, M. and Carreiras, M. (2006). Do transposed-letter effects occur across lexeme boundaries? *Psychonomic bulletin & review*, 13(3):418–422.

Perea, M. and Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51(2):231–246.

Pham, H. and Baayen, R. H. (2013). Semantic relations and compound transparency: A regression study in CARIN theory. *Psychologia*, to appear.

Plag, I. and Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language*, 85:106–149.

Plaut, D. C. and Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, 15(4/5):445–485.

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. *Journal of the Acoustical Society of America*, 118:2561–2569.

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.

Pylkkänen, L., Feintuch, S., Hopkins, E., and Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: an MEG study. *Cognition*, 91:B35–B45.

Ramscar, M. and Baayen, R. (2013). Production, comprehension and synthesis: A communicative perspective on language. *Frontiers in Language Science*, in press.

Ramscar, M., Dye, M., Gustafson, J., and Klein, J. (2013a). Dual routes to cognitive flexibility: Learning and response conflict resolution in the dimensional change card sort task. *Child Development*, page doi: 10.1111/cdev.12044.

Ramscar, M. and Gitcho, N. (2007). Developmental change and the nature of learning in childhood. *Trends In Cognitive Science*, 11(7):274–279.

Ramscar, M., Hendrix, P., and Baayen, R. (2013b). Nonlinear dynamics of lifelong learning: the myth of cognitive decline. *Manuscript, University of Tübingen.*

Ramscar, M. and Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, 31(6):927–960.

Rastle, K. and Davis, M. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23(7-8):942–971.

Rastle, K., Davis, M. H., and New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11:1090–1098.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422.

Roelofs, A. (1997). Morpheme frequency in speech production: Testing WEAVER. In Booij, G. E. and Van Marle, J., editors, *Yearbook of Morphology 1996*, pages 135–154. Kluwer, Dordrecht.

Roelofs, A. and Baayen, R. H. (2002). Morphology by itself in planning the production of spoken words. *Psychonomic bulletin and review*, 9:132–138.

Rubin, G. and Turano, K. (1992). Reading without saccadic eye movements. *Vision Research*, 32(5):895–902.

Rueckl, J. G. and Rimzhim, A. (2011). On the interaction of letter transpositions and morphemic boundaries. *Language and cognitive processes*, 26(4-6):482–508.

Sander, P. and Sanders, L. (2006). Rogue males: Sex differences in psychology students. *Eletronic Journal of Research in Educational Psychology*, 8(4):1.

Scarborough, R. A. (2004). Coarticulation and the structure of the lexicon. UCLA dissertation.

Schreuder, R. and Baayen, R. H. (1994). Prefix-stripping re-revisited. *Journal of Memory and Language*, 33:357–375.

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.

Schuppler, B., van Dommelen, W. A., Koreman, J., and Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final/t: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics*, pages 595–607.

Segalowitz, S., Zheng, X., et al. (2009). An ERP study of category priming: Evidence of early lexical semantic access. *Biological Psychology*, 80(1):122.

Seidenberg, M. (1987). Sublexical structures in visual word recognition: Access units or orthographic redundancy. In Coltheart, M., editor, *Attention and Performance XII*, pages 245–264. Lawrence Erlbaum Associates, Hove.

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

Smolka, E., Komlosi, S., and Rösler, F. (2009). When semantics means less than morphology: The processing of german prefixed verbs. *Language and Cognitive Processes*, 24(3):337–375.

Solomyak, O. and Marantz, A. (2010). Evidence for early morphological decomposition in visual word recognition. *Journal of Cognitive Neuroscience*, 22(9):2042–2057.

Spieler, D. H. and Balota, D. A. (1998). Naming latencies for 2820 words. *[On-line], Available: http://www.artsci.wustl.edu/∼dbalota/naming.html.*

Sproat, R. (1992). *Morphology and Computation.* The MIT Press, Cambridge, Mass.

Stump, G. (2001). *Inflectional Morphology: A Theory of Paradigm Structure.* Cambridge University Press.

Tabak, W., Schreuder, R., and Baayen, R. H. (2005). Lexical statistics and lexical processing: semantic density, information complexity, sex, and irregularity in Dutch. In Kepser, S. and Reis, M., editors, *Linguistic Evidence — Empirical, Theoretical, and Computational Perspectives*, pages 529–555. Mouton de Gruyter, Berlin.

Tabak, W., Schreuder, R., and Baayen, R. H. (2010). Producing inflected verbs: A picture naming study. *The Mental Lexicon*, 5(1):22–46.

Taft, M. (1981). Prefix stripping revisited. *Journal of Verbal Learing and Verbal Behavior*, 20:289–297.

Taft, M. (1991). *Reading and the mental lexicon*. Psychology Press.

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, 9:271–294.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *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.

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.

Tremblay, A. and Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In Wood, D., editor, *Perspectives on Formulaic Language: Acquisition and communication*, pages 151–173. The Continuum International Publishing Group, London.

Tremblay, A., Derwing, B., Libben, G., and Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*.

Tremblay, A. and Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon*, 6(2):302–324.

Uhlenbeck, E. M. (1978). *Studies in Javanese Morphology. (Studies in Javanese Morphology)*. Nijhoff, The Hague.

Ullman, M. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92:231–270.

Ullman, M. T., Estabrooke, I. V., Steinhauer, K., Brovetto, C., Pancheva, R., Ozawa, K., Mordecai, K., and Maki, P. (2002). Sex differences in the neurocognition of language. *Brain and Language*, 83:141–143.

Ussishkin, A. (2005). A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language & Linguistic Theory*, 23(1):169–218.

Ussishkin, A. (2006). Affix-favored contrast inequity and psycholinguistic grounding for non-concatenative morphology. *Morphology*, 16(1):107–125.

Vannest, J., Newport, E. L., Newman, A. J., and Bavelier, D. (2011). Interplay between morphology and frequency in lexical access: The case of the base frequency effect. *Brain research*, 1373:144–159.

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.

Vul, E., Harris, C., Winkielman, P., and Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4(3):274–290.

Wagner, A. and Rescorla, R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. and Prokasy, W. F., editors, *Classical Conditioning II*, pages 64–99. Appleton-Century-Crofts, New York.

Weingarten, R., Nottbusch, G., and Will, U. (2004). Morphemes, syllables and graphemes in written word production. In Pechmann, T. and Habel, C., editors, *Multidisciplinary approaches to speech production*, pages 529–572. Mouton de Gruyter, Berlin.

Wood, S. N. (2006). *Generalized Additive Models*. Chapman & Hall/CRC, New York.

Wurm, L. H. (1997). Auditory Processing of Prefixed English Words is Both Continuous and Decompositional. *Journal of Memory and Language*, 37:438–461.

Wurm, L. H. and Aycock, J. (2003). Recognition of spoken prefixed words: The role of early conditional root uniqueness points. In Baayen, R. H. and Schreuder, R., editors, *Morphological Structure in Language Processing*, pages 259–286. Mouton de Gruyter, Berlin.

Wurm, L. H. and Ross, S. E. (2001). Conditional root uniqueness points: Psychological validity and perceptual consequences. *Journal of Memory and Language*, 45:39–57.

Yeung, N., Nystrom, L., Aronson, J., and Cohen, J. (2006). Between-task competition and cognitive control in task switching. *The Journal of Neuroscience*, 26(5):1429–1438.