Ingo Plag and Laura Winther Balling

Derivational morphology: An integrative perspective on some fundamental questions

Abstract: This chapter tries to answer some central questions in the study of derivational morphology: What are the units of analysis? What are the mechanisms that underlie the creation, as well as the syntagmatic and paradigmatic relationships, of derived words? For each of these questions we discuss a wide variety of approaches in different subdisciplines of linguistics (phonetic, theoretical-linguistic, psycholinguistic, neurolinguistic and computational-linguistic), and see what evidence the diverse approaches have brought forward to support their ideas.

Keywords: Derivational morphology, morpheme, Discriminative learning, Computational modeling, Evidence in morphology, Psycholinguistics, Priming

1 Introduction: What is derivational morphology?

Derivational morphology (or 'derivation' for short) belongs to the realm of word-formation and is usually defined negatively as the kind of word-formation that is not compounding. Compounding is widely understood as the formation of words by concatenating two or more lexemes or bases. Thus, derivation is concerned with the ways in which morphologically complex lexemes are related to, or derived from, other lexemes by affixational or non-affixational means, but not by combination with other lexemes.

This definition of derivation leaves us with a problem of demarcation between compounding and derivation, which hinges on the question of what we understand by 'lexeme' or 'base' as against 'affix'. Additionally, since there is the basic distinction between word-formation and inflection, there is the problem of demarcation of derivation (as part of word-formation) vis-à-vis inflection. Both demarcation problems have been amply discussed in the literature (more recently, for example, by Lieber and Stekauer 2009, ten Hacken 2014),

Ingo Plag, Institut für Anglistik und Amerikanistik, English Language and Linguistics, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany
Laura Winther Balling, Widex A/S, Nymøllevej, Lynge, Denmark

and we will briefly summarize from that literature only what is relevant for the present chapter. We start with inflection vs. derivation.

Traditionally, inflection is considered to be concerned with the encoding of syntactic information, while derivation encodes lexical information. But what do we mean by 'syntactic' and 'lexical'? The definition of these terms is, obviously, theory-dependent but there seems to be a growing consensus that the two notions refer to endpoints on a scale rather than to a clear-cut categorical opposition (see, for example, Dressler et al. 2014). The idea of a cline is, for example, found in the distinctions between contextual inflection, inherent inflection and derivation (Booij 1993, and Chapter 7, Marzi et al. 2020, this volume). Contextual inflection such as agreement morphology is uncontroversially syntactic in nature, while inherent inflection such as plural or tense marking is more ambiguous and may encode categories that in some languages can be taken to be derivational.

The literature on the topic usually puts forward a number of criteria to distinguish between inflection and derivation, such as position (inflection is more peripheral), productivity (inflection is more productive), semantics (inflection encodes grammatical meaning) or transparency (inflection is more transparent). These criteria lead to satisfactory classifications in many cases but borderline cases remain, and this has led some people to reject the distinction altogether. Certain theories (e.g. Beard 1995) dispose of the distinction because it is deemed unnecessary in their framework. We will leave these issues open since to a large extent the debate is irrelevant for the present treatment. Nothing we say below hinges on whether we assume that derivation and inflection need to be separated or not. We simply look at phenomena that are standardly classified as derivational in nature. The reader may feel free to assign these phenomena to the morphological compartment they think appropriate.

With regard to the delimitation of derivation and compounding several problems can be discerned. One frequently discussed problem is how to determine whether a given form should be classified as a base or an affix. If a base, the complex word would be a compound, if an affix, the complex word would be a derived word. Borderline cases are well-known and again there seems to be a cline rather than a dichotomy. If we include non-affixational word-formation processes such as reduplication or blending, we may suspect that these might be processes that involve the concatenation of bases and could thus be regarded as compounding processes. And indeed, blending in English has been explicitly argued to be a kind of compounding (accompanied by the loss of phonological material, Bauer et al. 2013, Arndt-Lappe and Plag 2013). Reduplication, on the other hand and by its very nature, never involves different bases, which is probably the reason why it is generally regarded as a kind of derivation (but see Štekauer et al. 2012, who treat full reduplication as compounding).

Obviously, derivational morphology is a vast field. The list of formal processes or relationships occurring in derivation is long and poses many challenges, both empirical and theoretical, for the analyst: prefixation, suffixation, infixation, circumfixation, transfixation, conversion, reduplication, truncation, back-formation, templatic derivation, etc. In the present chapter we therefore need to be very selective on our coverage, and will focus on very few, but rather fundamental, questions in the study of derivation, approaching these from an interdisciplinary perspective that combines insights from phonetics, theoretical linguistics, psycholinguistics, neurolinguistics and computational linguistics.

In particular we will deal with two long-standing, central issues.[1] The first is what units of analysis we need to assume (Section 2), the second is what kinds of mechanisms manipulate these units (Section 3). An integrative approach allows us to take a fresh look at these issues in order to reassess the debates and open up new research perspectives.

2 What are the units?

There has been a fierce debate in theoretical linguistics about the nature of morphological knowledge and its organization in syntax or lexicon. Numerous theories have been proposed and depending on which theory one looks at, a different unit may be taken as central for an understanding of word structure, while other units are claimed to be non-existing or epiphenomenal. Prominent approaches in this debate have been grouped under the names of 'word-and-paradigm', 'item-and-arrangement' or 'item-and-process'. For example, while word-and-paradigm morphologists believe in the central importance of the word and reject morphemes as independent units, item-and-arrangement theorists may hold the opposite to be true. We will not review all existing approaches here but try to scrutinize commonly proposed units for their usefulness and for the cross-disciplinary evidence that can be put forward for their existence.

---

1 With regard to other issues, the reader is referred to other chapters of this volume, e.g. Chapter 6 on morpho-phonology (Arndt-Lappe and Ernestus 2020, this volume), Chapter 10 on morpho-pragmatics (Merlini Barbaresi and Dressler 2020, this volume), or Chapter 14 on acquisition (Ravid, Keuleers, and Dressler 2020, this volume).

2.1 The word

The central unit in word-formation is the word. However, although this notion is used without hesitation by most researchers, attempts at defining it have shown that it is a rather problematic notion, even if restricted to its morphological aspects. Ultimately, a language-independent definition seems impossible (see the discussion in Haspelmath 2011), and even language-dependent criteria at various levels of description (phonological, morphological, semantic, syntactic etc.) often run into problems by being either hard to apply or yielding contradictory results. It should be noted, however, that computational models of unsupervised word segmentation can reach surprisingly high levels of accuracy (e.g. Synnaeve et al. 2014), which shows that the concept may not be as elusive as some might think.

In spite of these issues, the daily practice in linguistic theory of all flavors and in computational linguistics is to use the word as a basic unit of analysis, which, from a practical point of view, makes a lot of sense, given the many useful insights linguists have arrived at by employing this notion.

But is there any evidence for its reality apart from distributional or theoretical arguments? Research in psycholinguistics and neurolinguistics has basically also followed this traditional strategy, using words as experimental stimuli to test all kinds of theories and hypotheses. Thus, most studies using key methods such as lexical decision, priming, picture naming or self-paced reading are based on the word as a unit of analysis, and by extension, of representation and processing. A partial exception to this otherwise dominant tendency is the study of multi-word units, including both idioms and those units known as 'lexical bundles', i.e. sequences of words that frequently occur together but do not have an idiomatic meaning. When such sequences show frequency effects (e.g. Tremblay et al. 2011), it may indicate that levels higher than the word are represented and processed as lexical (rather than syntactic) units.

However, though such studies show evidence of higher-level units, they do not fundamentally question the psychological reality of the word, and in fact, to our knowledge, there is no study explicitly addressing the question whether the word exists in the minds of the speakers or not. Instead, word-based experiments, although not testing explicitly the existence of words, have provided ample indirect evidence for the psychological or neurological reality of the word as a unit of representation and processing – and one that participants in such experiments accept intuitively when for instance asked to decide whether a string is a word or not. It has to be noted, however, that for languages for which the notion of word seems especially problematic, for example polysynthetic languages, behavioral or neurological evidence is very scarce. Additionally, it

seems that the centrality of the notion of word may be overestimated due to the effect of literacy, but experiments and linguistic theories are so overwhelmingly based on behavior of literate language users that it is difficult to say anything meaningful about the status of the word in a non-literate language.

The fact that the word is so widely accepted as a linguistic unit may be interpreted as good news for approaches in which the word is the central, or indeed the only, unit of analysis, as, for example, in what Blevins (2006) calls 'abstractive' word-based approaches. In such approaches, word forms are the basic units of the system, and recurrent parts are abstractions over these full forms. Crucially, recurrent minimal parts are not listed as independent units. This makes different empirical predictions than a model that assumes that minimal parts are the building blocks for the creation of words (see Blevins 2006: 537ff for more detailed discussion of this point).

In an abstractive word-based system, derivational morphology is thus conceived of as the relation between complex words, and not, for example, between affixes and their bases. Early proponents of modern word-based derivational morphology are Aronoff (1976) and Booij (1977), and Construction Morphology (e.g. Booij 2010, see Section 3.3) is a modern descendent of this approach.

2.2 The morpheme as a minimal sign

Traditionally, the morpheme constitutes a minimal sign, combining a form and a meaning. Roots and affixes are classical instances of morphemes, but so-called 'prosodic templates' as found in Semitic languages have also been subsumed under that notion. In these languages root morphemes are comprised of non-contiguous segments with interleaving elements. Depending on which vowels surface in the interleaving slots, different meanings emerge, for the Arabic root *k-t-b* for instance the words *kitaab* 'book' versus *kataba* 'he wrote' versus *kaatib* 'writer' (Ryding 2005: 46).

The most pervasive problem of the notion of morpheme is the relation of form and meaning itself. There are a number of problems that arise if one tries to identify minimal signs inside words. First, there is the problem of what Hockett named 'total accountability' (Hockett 1947). Thus, every word should exhaustively be analyzable into morphemes. This brings up a number of problems. First, there is the problem of zero morphs, that is meaning without phonological realization, such as the plurals of nouns like *sheep* in English, or *Segel* 'sail' in German, in which plural is not overtly marked. In the realm of derivation, conversion would be a case in point, where for example a verb such as *cook* comes to serve also as a noun, without any morphological marking. Second, there is

phonological material (sometimes called 'empty morphs') that has no meaning, for example *-n-* in *plato-n-ic* or *-in-* in *attitud-in-al* (alternatively, such forms are treated as part of stem allomorphs). Third, there is multifunctionality, by which a single morpheme can have different meanings, for example *-er* in English, which can express an agentive meaning (*writer*), but can also derive inhabitant nouns (e.g. *Londoner*).

Sometimes it is not even clear whether there is a clear meaning associated with a particular form. Consider (1):

(1) a. submit, permit, remit, admit
     b. infer, confer, prefer, refer, transfer

In (1-a) and (1-b), *-mit* and *-fer* are clearly recurrent elements at the level of form, but their meanings are essentially unclear (see, e.g., Plag 2018 for detailed discussion).

Another problem for the mapping of form and meaning with morphemes are most cases of non-concatenative morphology. In conversion, for instance, no (change in) form corresponds to some (change in) meaning. Other cases in point are vowel alternations (as in English *rise* – causative *raise*, Arabic *Maryam* – diminutive *Maryuum*), or phonological truncation (as in English *celeb* – *celebrity*). With such derivatives it is very hard to say what form is being mapped on the given meaning.

Another, related, problem is that the notion of morpheme rests on the assumption that there are clear boundaries that separate from each other the different morphemes that a word is made up of. Even adherents of the morpheme concede that these boundaries may vary in strength, as in *Sound Pattern of English* (Chomsky and Halle 1968) or *Lexical Phonology and Morphology* (Kiparsky 1982), which posit two or more kinds of boundary. More recent research, especially on affix ordering (Hay and Plag 2004; Plag and Baayen 2009), has demonstrated, however, that morphological segmentability is in fact gradient, and that there are no clear cut-off points for positing only two or three degrees of boundary strength.

The idea of discrete and categorical boundaries is also challenged by work on semantic transparency. While syntax and inflection have a strong tendency to exhibit full semantic compositionality in the sense of Frege (1892), derivational morphology often shows semantic opacity, i.e. forms where the semantics of the derived word is not the sum of its parts (see, among many treatments, Ronneberger-Sibold 2003 for illustration and discussion). There is a consensus that semantic opacity is a gradient phenomenon (e.g. Gonnerman and Anderson 2001, Pastizzo and Feldman 2001), and both the existence and the gradience of semantic opacity is a challenge to the notion of morpheme.

According to Hay (2001, 2003) morphological segmentability is also influenced by the frequential properties of base and derivative, which in turn often go together with gradient phonological and semantic opacity. Consider *government* vs. *discernment*. The derived word *government* is far more frequent than its base govern (the opposite is true for the pair *discern–discernment)*. Notably, *government* is also somewhat opaque in its meaning ('body that governs', rather than 'action of governing') and phonologically considerably reduced. All three properties contribute to its being less easily segmentable than, say, *discernment* (see Plag 2018 for an introduction to the notion of variable morphological segmentability). Another example is the English nominalizing suffix *-th*, which is much less easily separable from its base (e.g. *depth*) than *-ness* (e.g. *soundness*), and even the same affix can be more or less separable in different words (e.g. *soundness* vs. *business*). The evidence for the gradience of boundaries is hard to reconcile with the idea of the morpheme as a discrete unit.

What is more, phonetic studies have shown that adding an affix to a base also affects the acoustic properties of the base, such that a base occurring on its own systematically differs acoustically from its realization as part of a derived word, for example in duration and pitch. For example, the base *help* without a suffix is generally shorter when it occurs in *helper* than if it occurs as a free morpheme (Lehiste 1972, Kemps et al. 2005, Frazier 2006, Blazej and Cohen-Goldberg 2015).

Importantly, two of these studies (i.e. Blazej and Cohen-Goldberg 2015; Kemps et al. 2005) have also shown that the sub-phonemic acoustic information is in fact used by listeners in lexical processing. In Blazej and Cohen-Goldberg's (2015) perception experiment, for example, listeners were able to distinguish between (segmentally identical) free and bound forms (e.g. *clue* as in *clue*, versus *clue* as in *clueless*) without hearing the suffix. Complex words thus do not behave as mere concatenations of distinct units, but significantly blur the alleged separate identity of the morphemes.

While such results may be seen as further evidence that the morpheme as a separable unit is untenable, there is also phonetic evidence that seems to speak in favor of the morpheme. Thus, Plag et al. (2015) have shown, albeit for inflection, that the different suffixal S morphemes in English (and indeed non-morphemic S) have systematically distinct durations. While such findings seem inexplicable for traditional models of the morpheme (which only allow for phonemic formal representations), the differences in duration could nevertheless be taken as indicating different morphs belonging to different morphemes. However, Plag et al. (2015) did not look at the acoustic properties of the stems to which the different S morphemes attached, so that it is still unclear whether the differences between the distinct Ss are accompanied by differences in the

stems. If so, this would tie in with the results by Kemps et al. (2005) and Blazej and Cohen-Goldberg (2015).

The insights from perception described above align nicely with other results from psycholinguistics that have been produced in the context of the debate about the nature of lexical storage and retrieval. Large parts of the literature on morphological processing are concerned with the question of whether complex words are segmented into their constituent morphemes or are instead treated holistically, and which mechanism might be responsible for the assumed differences in processing. This volume devotes a whole chapter to this debate (see Chapter 12, Fábregas and Penke 2020, this volume) and we only briefly summarize it here, restricting ourselves mostly to derivational morphology.

We start, however, with inflection, since the debate about morphemes vs. whole words as units for lexical processing and representation has focused mostly on inflection, with three theoretical positions. Firstly, all inflectional forms may be stored as morphemes and then combined based on rules during production or comprehension. Secondly, all morphologically complex words may be processed associatively, with no rule-based combination of morphemes. Thirdly, rule-based processing may apply to regular words and associative processing to irregular ones. In fact, the first position, usually associated with Generative Phonology (Chomsky and Halle 1968), is not psycholinguistically tenable. Instead, the so-called past-tense debate occurs between the second and the third position.

The Dual-Mechanism or Declarative-Procedural model (Ullman 2001, 2004, Pinker and Ullman 2002) posits two fundamentally distinct cognitive processing systems, one based on rules, which applies to regular inflected words as well as syntax and other supposedly symbolic domains, while the other, based on associative mechanisms, applies to irregular complex words. In opposition to this, various associative models, which are often but not necessarily connectionist, hold that all words, whether simple or complex, regular or irregular, are processed in fundamentally the same way (e.g. Seidenberg and McClelland 1989, Joanisse and Seidenberg 1999). In such models, morphological effects arise not as the result of morphemic representations, but due to the relationships in form and meaning that morphological structure codes.

For derivational morphology, the picture is more complex, because forms that are entirely regular in form and meaning are rarer for derivation than for inflection; in fact, full formal and semantic regularity tends to be seen as a characteristic of inflection as a morphological operation, while derivation is characterized by many-to-many mappings, more gradience and semantic and phonological irregularity. However, some of the same diagnostics of rule vs. analogy are used both in the past-tense debate, which focuses on inflection,

and in the broader word recognition literature, which includes derived words as well as other morphologically complex words.

The most important diagnostic, since (at least) Taft (1979), is probably frequency effects. This is based on the rationale that units for which we observe frequency effects are also units of representation in the mental lexicon. Thus, so runs the logic, if for morphologically complex words, we see that faster recognition of words with higher base frequency, then the base must be the unit on which the processing of such words is based. If for instance the high base-frequency *own-er* is processed faster than the lower base-frequency *send-er*, this would traditionally be evidence of morpheme-based recognition of these words. If, instead, effects of the frequency of whole complex words are observed, the unit of processing for such words is traditionally taken to be the whole word rather than the morphemes.

Countless studies have investigated the effect of base vs. whole-word frequency, usually in factorial designs where the frequencies are manipulated categorically, comparing words with high and low values on base or whole-word frequency. Whole-word frequency effects are the largest and most pervasively found, across a range of languages (e.g. Taft 1979 for English, Baayen et al. 1997 for Dutch, Lehtonen et al. 2006 for Swedish, Balling and Baayen 2012 for Danish, Meunier and Segui 1999 for French, Moscoso del Prado Martín et al. 2004 for Finnish and Moscoso del Prado Martín et al. 2005 for Hebrew), but base frequency effects are also found in a number of studies (Baayen et al. 1997; New et al. 2004; Taft 1979; Vannest et al. 2002). The relevance of the two types of different frequency effects is hypothesized to vary depending on a range of factors, most prominently formal and semantic regularity, with stronger base frequency effects for regular and transparent words and stronger whole-word frequency effects for more irregular and opaque words.

More recently, experiments employing regression designs, where it is possible to investigate graded effects of different frequency measures, have shown more complex patterns of co-existing and sometimes interacting whole-word and morpheme frequency effects (e.g. Baayen et al. 2007; Balling and Baayen 2008; Kuperman et al. 2008; Plag 2009). In addition to these complexities, the case has been made by Baayen et al. (2007) that base frequency may in fact not be a good indicator of morphemic processing, i.e. processing in which the morpheme is the basic unit of representation. The argument is based on the fact that the frequency of a base as an independent word is not really informative about the relevance of that base in the context of a specific complex word. This interpretation is supported by the relative weakness of base frequency effects, even in the presence of other morphemic effects. Conversely, whole-word frequency may in turn not solely indicate word-based processing of complex

words; instead, it may also be understood as the combinatorial probability of the constituent morphemes co-occurring and thus a measure of morphemic processing. Such an understanding is supported by the strength and pervasiveness of whole-word frequency effects, by their early occurrence as measured by eye-tracking of compound reading (Kuperman et al. 2008), and by the interactions between different frequencies referenced above.

The literature on morphological processing is predominantly focused on (visual) word recognition, but evidence from word production does exist and uses similar experimental manipulations. Thus, manipulations of morpheme vs. whole-word frequency are also used in the production literature as a diagnostic of morphemic vs. whole-word representations for complex words. The evidence from this research is mixed, with some studies showing whole-word frequency effects for complex words (e.g. Chen and Chen 2006; Janssen et al. 2008) and others showing morpheme frequency effects for similar words (e.g. Bien et al. 2005; Roelofs 1996).

Despite the similarity in the use of frequency manipulations, the literature on word production deviates in (at least) two ways from that on word recognition: firstly, the stimuli are often compound words rather than derived, at least partly for the practical reason that compound words are easier to depict and thus elicit from participants. Secondly, a central concern for the studies mentioned above (as well as those cited in Section 2.3 below) is the level of the production process at which morphological information becomes relevant; an aspect that we have to ignore here to avoid straying too far from the fundamental questions of derivational morphology. However, an important similarity remains across processing directions, namely that there is evidence of both morphemic and whole-word frequency effects, and that the interpretation of these frequency effects, as we saw above, may not be as clear-cut as has been previously assumed.

The processing literature thus shows a complex picture. However, it seems clear that processing cannot to any pervasive extent be based on combinatorial processing of morphemic representations. There is a growing consensus that all sorts of information is used during processing, cf. for instance Libben's (2006) notion of maximization of opportunity. The types of information used relate both to whole complex words and to the morphological structures of such words, but importantly, this does not necessarily entail morphemic representations.

In conclusion, theories that restrict their inventory of units to morphemes must either stretch this notion to such an extent that it becomes unrecognizable, or they simply ignore important evidence that morphemes are not the neat discrete units that these theories are working with.

2.3 The morpheme as a phonological spell-out: The separation of form and meaning

Both word-based and morpheme-based approaches to morphology can be characterized as sign-based for they assume units of form and meaning. There are, however, also approaches that separate form and meaning in morphology. Proponents of such an approach are Beard (1995), Don (1993), Gussmann (1987a), Szymanek (1985). Distributed Morphology (Halle and Marantz 1993) also adopts a separationist view. We will restrict the following discussion of separationism to aspects that are of special relevance to derivational morphology.

Separationism proposes a radical solution to address the problem of many-to-many relations between form and meaning in morphology: What is traditionally called a morpheme is taken to be merely a phonological spell-out (a 'vocabulary item' in Distributed Morphology lingo) of a grammatical or semantic category. This dissociation of meaning and phonological spell-out supposedly "accounts for asymmetries of affixal sound and meaning such as polysemy, synonymy, zero and empty morphology, which plague other approaches" (Beard 1990: 103).

Unfortunately, there are hardly any studies available on derivational morphology that systematically compare the merits and problems of sign-based vs. separationist accounts of the same derivational phenomenon. Plag (1999), however, provides a comprehensive formal and semantic analysis of derivation into verbs in English and discusses in detail the implications for the two kinds of morphological theory. Existing separationist accounts of English verbal derivation (Gussmann 1987b, Beard 1995) assume that the different productive processes (*-ize*, *-ify*, *-ate* and conversion) are all spell-out operations on only two underlying semantic operations. The semantic analysis in Plag (1999) shows, however, that each of the four verb-deriving processes comes with its own intricate polysemy (see also Bauer et al. 2013; Lieber 1998, 2004) and that the separationists' crucial assumption of absolute synonymy (Beard 1995: 78) of all pertinent forms is wrong.

Synonymy and polysemy thus seem to constitute an interesting difference between inflection and derivation. While in inflection, the assumption of absolute synonymy of different affixes ('spell-outs') is uncontroversial at least for contextual inflection (such as structural case marking or agreement), absolute synonymy of different derivational affixes seems very unusual. The latter point is supported by large-scale empirical evidence across many derivational processes as gleaned by Bauer et al. (2013) in their recent survey of English morphology. Furthermore, polysemy is absent from, at least, contextual inflection (or else goes under the name of syncretism and is viewed as homonymy rather than polysemy).

Given that the major claim of separationist morphology is the lack of morphological signs it is unclear what kind of independent neurological or behavioral evidence could prove this assumption (see Chapter 4, Marangolo and Papagno 2020, this volume, and Chapter 12, Fábregas & Penke 2020, this volume, for discussion). After all, it is hard to prove the non-existence of something. However, two related lines of behavioral research address the separationist rejection of morphemic units.

Firstly, there is a whole literature comparing formal (mainly orthographic), semantic and morphological priming effects and their time courses, i.e. the facilitation of the reading of a target word depending on whether it has been preceded by a word that is orthographically (*twin-twinkle*), semantically (*idea-notion*) or morphologically (*government-govern*) related compared to an unrelated prime word (*idea-tin*). If morphological priming effects can be shown to be the sum of semantic and formal overlap, it would provide evidence against an independent role for morphology in word recognition.

However, such priming studies show that morphological priming effects tend to be stronger and more consistent across different types of priming (subliminal, immediate and delayed) than both semantic and formal priming effects (e.g. Feldman 2000; Marslen-Wilson et al. 1994; Rastle et al. 2000). In fact, purely formal priming effects are surprisingly elusive and, if found, tend to be inhibitory (e.g. Gonnerman et al. 2007; Longtin et al. 2003; Marslen-Wilson and Zhou 1999). This does not entirely rule out that morphological effects are epiphenomenal, arising from interacting effects of semantic and formal relatedness, but it clearly does not support that idea.

The second line of research concerns the question of whether morphological effects are graded depending on semantic and formal overlap between a complex word and its morphological relatives, specifically its base. Although the majority of studies in this area are again priming studies, there are exceptions to this, which is an advantage given the way especially subliminal priming relies on misleading the processing system, by briefly presenting a word that may or may not be related to that target word that participants are asked to focus on.

Starting with the priming literature, the seminal study of Marslen-Wilson et al. (1994) showed morphological priming effects that were the same irrespective of whether the formal relation between the word pairs was transparent, e.g. *punishment-punish*, or more opaque, e.g. *vanity-vain*, *elusive-elude*. By contrast, morphological priming does seem to vary depending on the degree of semantic transparency (Longtin et al. 2003; Marslen-Wilson et al. 1994; Rastle et al. 2000), though there is some debate as to whether this also holds for very early priming effects (Feldman et al. 2015; Rastle and Davis 2008).

Also word recognition studies that do not employ priming show effects of semantic transparency on recognition, with word recognition becoming faster the more transparent the combination of base and affix (Baayen et al. 2007; Balling and Baayen 2008; Wurm 1997). Such graded morphological effects are potentially problematic for models relying on morphemes as units of representation, though some level of morphemic representation may still be postulated if it is posited that whole-word and morpheme representations interact during processing.

Again, the word production literature uses similar manipulations, but, as we saw above, with a skew towards compound words as stimuli. Nonetheless, a brief review of the findings is relevant due to the similarity of manipulations. The picture-word interference paradigm resembles the priming paradigms used to study word recognition. In picture-word interference, a written or spoken context word – essentially a prime – is presented before a picture that the participants are requested to name. Effects of the relation between the context word and the target picture are taken as evidence of relations or overlap between representations in the mental lexicon. The results of these studies show that, also in this domain, morphological effects seem to be qualitatively different from semantic and formal effects: while a semantic relation between context word and target results in interference, at least for categorically related words (Costa et al. 2005; Schriefers et al. 1990), morphological relatedness produces shorter naming latencies for the target, i.e. facilitation (Dohmes et al. 2004; Zwitserlood et al. 2000, 2002). This holds whether or not the morphologically complex context word is semantically transparent or not, and with both immediate and delayed presentation of the target relative to the prime (Dohmes et al. 2004). Facilitation is even observed with morphologically complex pseudo-word primes (Bölte et al. 2013). Formal priming effects are again more elusive than both semantic and morphological effects.

2.4 Phonaesthemes

One particularly problematic potential unit of morphological analysis is the so-called 'phonaestheme'. The debate has centered around the question whether phonaesthemes should be regarded as kinds of morphemes. Phonaesthemes are defined as subparts of roots that show a recurrent sound-meaning pairing, and thus resemble morphemes, but are different from morphemes in at least two important respects. First, their semantics is rather vague and not very consistent (e.g. 'light emitted from a source', as in, for example, *glimmer*, *glow*, *gleam*, *glare* as opposed to *glove*, *glue*, *glum*). Second, once separated out, the

morphological status of the residue is unclear (what is, for example, *-immer* or *-ow* in *glimmer* and *glow*, morphologically speaking?). Here, the problem of total accountability raises its ugly head again.

Theoretical linguists have struggled with the problems posed by phonaesthemes but have not reached a consensus. In a most recent attempt, Kwon and Round (2015) develop a catalogue of seven criteria and systematically compare roots and affixes with phonaesthemes using these criteria. Only one criterion, the presence of a non-recurrent residue is the sole clear difference between phonaesthemes and other derivational entities that are considered morphemic in nature. In view of this result, theories can choose whether they would extend the notion of morpheme to such entities.

However, there is still the question of whether phonaesthemes have any relevance for the speakers of a language. Bergen (2004) conducted a priming experiment which showed that phonaesthemically related primes facilitate processing, with semantic priming and orthographic priming each yielding quantifiably different priming effects from phonaesthemic priming. This is very similar to the priming effects found with morphemes, discussed above, and ties in with earlier psycholinguistic work on phonaesthetic neologisms (e.g. Abelin 1999; Hutchins 1998; Magnus 2000). In a broader perspective this means that language users unconsciously and generally pick up all sorts of recurrent form-meaning pairings and make use of these pairings even when they are not clearly compositional or categorical (see also Pastizzo and Feldman 2009).

2.5 None of the above

All of the morphological building blocks discussed so far try to approach the crucial problem of form-meaning mappings by segmentation. Language is either segmented into units of finer or coarser grain size that carry meaning (phonaesthemes, morphemes, words), or it is segmented into separate units of phonology, and units of semantics that are then mapped onto each other via complex mapping rules.

A radically different approach is taken in network models in which, crucially, morphological effects are emergent from association processes between certain kinds of formal representation and certain kinds of conceptual units. There are two main types of models: distributed connectionist models (e.g. Gonnerman et al. 2007; Harm and Seidenberg 2004; McClelland and Elman 1986; Norris 1994) on the one hand, and Naive Discriminative Learning (NDL) models on the other hand (e.g. Baayen et al. 2011, 2015, see also Chapter 2, Pirrelli et al. 2020 this volume). These two main types differ more in the processing

mechanisms than in the representational units they use. These units ultimately do not seem to have a particular theoretical status apart from the idea that the building blocks are decidedly non-morphemic.

Both types of architecture operate with input and output layers, in NDL known as cues and outcomes, with weighted connections between them determining the degree of activation of a given outcome based on a given input pattern. In connectionist models of language processing, the representational units may be individual letters or phonemes defined by values on a number of binary features (e.g. Joanisse and Seidenberg 1999). The output layer is typically a semantic representation of some sort, which may be distributed such that a word's meaning is represented by its pattern of activation of different semantic features (e.g. Plaut and Gonnerman 2000), or localist with one unit per word meaning (e.g. Joanisse and Seidenberg 1999).

In NDL, the input units are n-grams of varying sizes, pairs (bigrams) or triplets (tri-grams) of letters or phonemes. Crucially, the n-grams have no independent theoretical status. For illustration, consider the phrase *in its context*, which, under a trigram approach, would consist of the trigrams /ɪnɪ/, /nɪt/, /ɪts/, /tsk/ and so on. These trigrams stand for triphones, which in turn simply represent contrasts in the speech signal that are not directly linked to units of meaning (for example, /ɪnɪ/ does not 'carry' any meaning). Rather, the trigrams are used by the learning system as continuously incoming cues for establishing relationships between sound and other experiences. The relationships are not encoded in the form of words, morphemes or anything of that sort, but as weights of the association between cues and outcomes. The choice of triphones as against diphones (or single sounds) as input units is ultimately driven by practical considerations, and it is acknowledged that these n-grams are ultimately not fine-grained enough to capture all subtleties of the phonetic input (see Baayen et al. 2015: 2 for more discussion).

In terms of the outcome units, most recent work in NDL has come up with so-called 'lexomes'. These are defined as an arguably somewhat intangible

> theoretical construct at the interface of language and a world that is in constant flux with the flow of experience. Lexomes are the lexical dimensions in the system of knowledge that an individual acquires and constantly modifies as the outcome of discriminatively learning from experience within a culture. Because lexomic contrasts serve as communicative counterparts to the specific experiences individuals and cultures discriminate for practical and communicative purposes, they can be evoked in context either by language use or real world experience. Accordingly, the more that a lexome is activated in a given context, the greater the degree of confidence that the cues that culturally discriminate it from other outcomes are present in the external world. (Baayen et al. 2015: 5)

Since the theoretical import of these associative models is mainly in their processing mechanism, we discuss them in more detail in Section 3.5, focusing on NDL.

2.6 Synthesis: What does the evidence say?

There is a long tradition in theoretical linguistics to view morphology either as the combination of minimal signs (see Section 2.2.), as the formal-semantic relationship between words (Section 2.1), or as the combination of phonological elements with semantic-functional elements (as described in Section 2.3). While the theoretical debates seem to have reached a dead end with beliefs rather than evidence dominating the individual researcher's concept of morphological units, the evidence from psycholinguistics points towards a solution.

Both strictly morphemic and separationist models are seriously challenged by the empirical evidence. It seems that a gradient view of morphology in which speakers and listeners make use of all sorts of sound-meaning pairings can better account for the intricate experimental results. These results suggest a continuum of variable association strengths between sound patterns and meaning. At one end of the continuum we find the constellation of strong and clear associations that is traditionally referred to as 'morpheme', while towards the other end of the continuum association between meaning and form becomes less and less strong and consistent. There are weaker and stronger multifaceted associations along both the paradigmatic and the syntagmatic dimensions, and these associations structure the lexicon and guide the processing of words.

Such a gradient view is further supported when one looks at the structural and cognitive mechanisms that have been proposed in order to understand derivational morphology. This will be the topic of the next section.

3 What are the mechanisms?

3.1 Lexical or syntactic?

In theoretical discussions of derivational morphology the question looms large whether morphology should be seen as an independent module of grammar, as part of the lexicon, or as simply non-existing as an independent module because it is part of syntax. Much of this debate focuses on the mechanisms, rather than on the units, that are taken to be instantiated in morphology. For example,

proponents of a syntactic view of morphology assume that word-internally the same mechanisms are at work as above the level of the word, i.e. in phrases, clauses and sentences, only with units that happen to be smaller than the word. Although a lot of ink has been spent to decide the issue, no consensus has been reached, and it is unclear what kind of evidence or killing theoretical argument might be able to solve the issue. The problem of the lexicon-syntax divide is intertwined with the problem of what kinds of particular mechanisms play a role in derivation. In the following we will review the different mechanisms proposed by various theories, sometimes also making reference to the more general question of whether morphology exists or not.
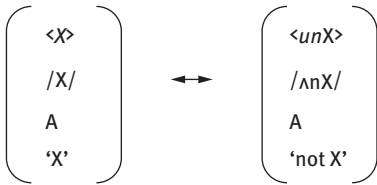
3.2 Rules

One of the most pervasive mechanisms, and one that has been used in different frameworks, is that of the 'rule'. Most approaches conceive of rules as input-output devices that either add structure to given representations or manipulate symbolic representations. Rules have been designed to generate new words in a deterministic fashion, or to formalize generalizations over existing forms. The former have usually been depicted as unidirectional rules, the latter as bidirectional rules (also known as 'redundancy rules', e.g. Jackendoff 1975).

Both morpheme-based and word-based morphology have used rules to account for derivation. Example (2-a) gives a syntactic phrase structure rule and (2-b) illustrates a similar kind of rule for the structure of words.

(2) a. Phrase structure rules
       NP   →   (article) (adjective) noun
       PP   →   preposition NP
     b. Syntactic rules below the word level
       word → root (suffix)
       word → root (suffix)

Figure 1 illustrates a word-based rule in the spirit of, for example, Aronoff (1976).

In rule-based frameworks, word-formation rules are accompanied by morpho-phonological adjustment rules which handle allomorphy or morpho-phonological alternations such as stress shift, assimilation or deletion. Word forms that do not conform to the rules are taken to be idiosyncratic exceptions (or unsystematic analogical formations) that need to be listed in the lexicon, while regular forms are not listed.

$$
\begin{pmatrix}
\langle X \rangle \\
/X/ \\
A \\
\text{'X'}
\end{pmatrix}
\longleftrightarrow
\begin{pmatrix}
\langle unX \rangle \\
/ʌnX/ \\
A \\
\text{'not X'}
\end{pmatrix}
$$

Figure 1: A word-based word-formation rule (from Plag 2018: 182, 'A' stands for 'Adjective').

There are at least three major problems with such deterministic rules. First, the dichotomy between rule and list is a fallacy. Numerous studies have shown that even derived words of the lowest frequency leave traces in memory and show frequency effects that would be impossible if these words were not part of the language user's mental lexicon (e.g. de Vaan et al. 2007; Baayen et al. 2007). Thus, any model that does not take into account the massive storage capacities of our mental lexicon is unconvincing.

Second, the amount of variation is often so large and internally structured that it cannot be satisfactorily accounted for by referring to the notions of rule and exception. For example, Collie (2008) demonstrates that a considerable proportion of the forms in her sample do not show the stress preservation that should regularly result from the pertinent rule (e.g. English *decónsecrate – dèconsecrátion* instead of predicted *decònsecrátion*). (Bauer et al. 2013; Chapter 9, Libben, Gagné and Dressler 2020, this volume) discuss numerous morpho-phonological alternations that show much more variation than previously assumed.

Similarly, affixes are often assumed to attach to a well-defined set of syntactically or semantically defined base categories, for example to verbal bases in the case of English *-able*. However, in reality one often finds minority patterns that can hardly be dismissed as mere exceptions. For example, we find a considerable number of *-able* forms based on nouns whose meaning is very similar to that of the deverbal forms (Plag 2004). Another such case is English *-ee*, which also attaches primarily to verbs but has a significant minority of denominal formations, too (Barker 1998; Mühleisen 2010; Plag 2004). In such cases the problem cannot be solved by simply adding the minority category to the rule as this would lead to massive overgeneration.

Third, rules are usually defined with a particular input being related to a particular output. It has been shown, however, that many morphological patterns are much better accounted for in an output-oriented fashion instead of an input–output relation (see, for example, Bauer et al. (2013: Chapter 9) on phonological aspects, and Plag (2004) on semantic aspects of output-orientedness).

Psycholinguistic and neurolinguistic evidence for the existence of rules in derivational morphology is scarce. In fact, the focus in the psycholinguistic

literature has been almost exclusively on whether or not morphemic units are relevant for processing (as reviewed in Section 2 above), rather than on the mechanisms by which they may be combined for interpretation in language comprehension, an issue which is notoriously difficult for derived words whose interpretation is frequently characterized by some level of opacity. The way the combination problem is usually handled is by postulating whole-word representations in addition to any morphemic representations, and avoiding the need for rules by assuming that the meaning of the combination is part of the whole-word representations. The whole-word representations may be on a later level of processing than the morphemic ones as in the model of Taft (1994), or vice versa as in the Supralexical model of Giraudo and Grainger (2001); alternatively, the two types of units may be on the same level in dual-route competition models (e.g. Frauenfelder and Schreuder 1992).

3.3 Schemas and inheritance

One important alternative to traditional rule-based approaches in derivation is Construction Morphology, as developed by Booij (2010). Although using a different notation, Construction Morphology is very similar to earlier HPSG-inspired approaches, which also used multiple inheritance hierarchies and special types of lexical rules to come up with a system of lexeme formation (see Bonami and Crysmann 2016 for a summary). We will focus on Construction Morphology for practical reasons as this framework uses a less technical formal notation and is more widely known.

In Construction Morphology derivational phenomena are formalized using so-called 'schemas'. A schema expresses a generalization about the form, meaning and syntax of derived words in the lexicon,+ and can serve as the basis for new coinages. The idea that word-formation patterns are abstractions over sets of related words in the lexicon has a long tradition (from Paul 1880 to Bybee 1995, to mention only two prominent thinkers). Example (3) (taken from Booij 2010: 80) illustrates Booij's approach with the Dutch deverbal nominal suffix *-er*. To the left of the double arrow we find a morphological pattern, in this case a word with a base V and the suffix *-er*. To the right of the double arrow we find the semantic interpretation of this pattern, with the subscript indices indicating cross-reference relations with the respective entities to the left of the double arrow. The notation 'SEM' stands for the semantic representation of the subscripted base.

(3) $[Vi -er]_j \leftrightarrow [\text{entity involved in } SEM_i]_j$

Schemas can be related to subschemas in hierarchies of schemas via inheritance. For example, the polysemy of the suffix *-er* can be captured by positing a more general schema, given in (4a), which dominates pertinent subschemas given in (4b) through (4f). The different kinds of syntactic categories that can be inserted for 'X' in (4a) are given in parentheses. The subschemas inherit the properties of the dominant schema and specify a particular semantic pattern among the *-er*-derivatives. An example with English translation is given below each schema. Schema (4d) is a subschema of (4c), as it inherits all properties from (4c), but makes the notion 'person' more specific ('inhabitant'). Example (5) illustrates the inheritance hierarchy for the schemas in (4).

(4) a. $[X_i -er]_j \leftrightarrow [\text{entity with some relation R to } SEM_i]_j$ (X = V, N, QN, Num)
     b. $[V_i -er]_j \leftrightarrow [\text{entity involved in } SEM_i]_j$
         klopp-er 'knocker'
     c. $[N_i -er]_j \leftrightarrow [\text{ person with some relation R to } SEM_i]_j$
         VVD-er 'member of VVD'
     d. $[N_i \text{ -}er]_j \leftrightarrow [\text{ inhabitant of } SEM_i]_j$
         Amsterdamm-er 'inhabitant of Amsterdam'
     e. $[QN_i \text{ -}er]_j \leftrightarrow [\text{ object with property } SEM_i]_j$
         tienponder 'ten-pounder'
     f. $[Num_i -er]_j \leftrightarrow [\text{ entity with some relation R to } SEM_i]_j$
         twintig-er 'person in his twenties'

(5)   4a
         ⟋⟍
     4b   4c |4e   4f
          |
          4d
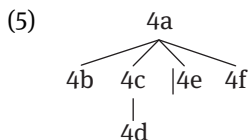
Paradigmatic relationships between formation patterns can be formalized by relating schemas to schemas. For example, the fact that all English nouns in *-ist* have a related form in *-ism* is captured in (6).

(6) $< [X\text{-ism}]_{Ni} \leftrightarrow SEM_i > \approx$
     $< [X\text{-ist}]_{Nj} \leftrightarrow [\text{person with property Y related to } SEM_i]_j >$

In order to coin new words, speakers may use schemas by simply unifying a particular base word ('X') with a schema. As the schemas differ in their degree of abstractness, the coinage of forms can happen at all levels. There is, however, no

explicit mechanism that would predict which schemas are productive and which ones are not, and psycholinguistic evidence for the existence of schemas is also absent.

Booij concedes that it is also possible to coin a word in analogy to a single existing word, sometimes referred to as 'local analogy' (Booij 2010: 88–93) or 'surface analogy' (Motsch 1981), but it remains somewhat unclear how such local analogies can be integrated into the schema-based model. In fact, there are analogical models on the market in which local analogies and wide-ranging generalizations emerge from a single analogical algorithm, which seems to make the postulation of constructional schemas unnecessary. We will discuss such models in the next section.

3.4 Analogy

Analogy is a very old concept in philosophy and linguistics, and one that has attracted a plethora of interpretations and definitions (see Arndt-Lappe 2015 for an overview and discussion of linguistic aspects). We will discuss analogy as a general mechanism by which words can be derived on the basis of similarities to existing words in the lexicon. The relevant sets of similar words (called sets of 'neighbors' or 'analogical sets') that form the basis for the analogical process may be very large or may be as small as one word. The crucial point is that the analogies are based on those exemplars in the lexicon that are informative with respect to the given task. What counts as informative is determined by the model, which also means that, in contrast to traditional applications of the notion of analogy, computational algorithms make testable predictions, and their performance can be compared to that of alternative models, for example rule systems. Work using computational analogical algorithms such as TiMBL ('Tilburg Memory-based Learner', Daelemans et al. 2007) or AM ('Analogical Model of Language', Skousen and Stanford 2007), has shown that many properties of morphological systems can be modeled quite successfully in this way.

Krott and colleagues (Krott et al. 2001, 2004a, 2004b), for example, showed that the notorious variability in the use of linking morphemes in Dutch and German compounds can be accounted for by analogy. In a nutshell, a given compound chooses the linking morphemes that other compounds with the same first or second constituent have. Similarly, English noun-noun constructs vary in their stress pattern, with roughly one third of these compounds in running speech being right-stressed and two thirds left-stressed. As demonstrated by Plag (2010) and Arndt-Lappe (2011), the prominence pattern of a given compound can be successfully predicted on the basis of analogy.

In the realm of derivational morphology, such models have only been sparsely used so far. Eddington (2002), for example, investigated Spanish diminutive allomorphy, Chapman and Skousen (2005) dealt with the competition between different negative prefixes in English and Arndt-Lappe (2014) was concerned with the rivalry between the English nominalizing suffixes *-ness* and *-ity*. Notably, all of these papers focus on phenomena of rival affixation in which there is a kind of variation that does not lend itself easily to a rule-based deterministic analysis. And indeed, the analogical models can deal with this variation quite well and achieve very satisfactory accuracy rates.

What is more, however, is that an analogical algorithm can not only model the behavior of isolated idiosyncratic formations and semi-regular behavior. Quite strikingly, the models are also able to come up with decisions that look categorical for certain, well-defined sets of words. This fact can be seen as the main advantage over rule-based deterministic models, which are characterized by the presence of at least two kinds of mechanisms, the rule and the exception.

But how does the algorithm do its job? We will roughly explain the procedure for AM. Let us assume that there is a candidate word for which we want to predict a certain behavior, for example, whether it will take *-ity* or *-ness* as an affix. First, the model must create an analogical set for the given candidate. To include a word in the analogical set a certain degree of similarity with the candidate is necessary. The similarity is computed over formal, syntactic, or semantic features coded for each word in the lexicon. AM decides this for each candidate individually on the basis of the degree of overlap in its features with the words in the lexicon. The model will always include maximally similar words into the analogical set, and words with lower degrees of similarity will be incorporated only if this does not lead to greater uncertainty with respect to the classification task.

On the basis of the analogical set, the model computes a probability of a particular choice based on the distribution of the two suffixes in the analogical set. In calculating these probabilities, AM takes into account the degree of similarity between the words in the analogical set and the candidate, as well as the number of words with a particular set of features. The more similar a word is to the candidate, the more weight it receives. And the more words share a particular set of features, the greater the weight assigned to each of these words. For the distribution of the rival nominalizing suffixes, the model cannot only reach high levels of general accuracy. What is more interesting from a theoretical point of view is that the model can also account for the different degress of productivity in different subsets of the data. Differences in the degrees of specificity of the domains for *-ity* and *-ness* translate into differences between more local and more general analogies. The analogical sets are generally quite small (for example between 10 and 16 words for candidates ending in *-able*, *-ous* or *-y*, with an overall lexicon of

545 pertinent words). Nevertheless, the default status of *-ness* emerges from the fact that analogies predicting *-ness* are based on a lesser degree of similarity than those predicting *-ity* (Arndt-Lappe 2014: 541).

Given that variability phenomena are a kind of home turf for analogical algorithms it is not surprising that the few studies of derivational morphology that have implemented such models have also been occupied with diachronic research questions. The standard procedure of such studies is to predict the distribution of forms in one time period (e.g. the 19th century) on the basis of a lexicon from the preceding time period (e.g. the 18th century). The results of such exercises are again quite impressive. For example, in Arndt-Lappe (2014) 85 percent of the predictions are correct if we want to predict the behavior of 20th century neologisms on the basis of the 19th century lexicon.

While analogical algorithms perform well with cases of affixal rivalry, it remains to be seen whether analogical modeling can be extended to other problems, such as predicting the semantic interpretation of a new derivative based on the semantics of the affix involved and the semantics of its base, a rather challenging issue to model given the elusiveness of semantic features.

3.5 Association of cues and outcomes: Naive Discriminative Learning

The subsequent discussion of associative models will focus on NDL. This is done not only because NDL is the more recent, and perhaps more provocative, addition to the language processing literature. What makes this approach especially interesting is the fact that it is based on an explicit theory of learning that is well established in cognitive psychology (e.g Rescorla 1988a, Pearce and Bouton 2001). The general cognitive mechanisms assumed in this theory have been shown to be able to model a number of important effects observed in animal learning and human learning, for example the blocking effect (Kamin 1968) and the feature-label ordering effect (Ramscar et al. 2010). The approach has recently been extended to language learning and language usage, and has been implemented by Harald Baayen and colleagues to model many different kinds of morphological and syntactic phenomena (e.g. Arnon and Ramscar 2012; Baayen et al. 2011, 2013, 2015; Baayen and Ramscar 2015; Blevins et al. 2015; Ramscar et al. 2010, 2013).

The central idea of associative learning theory is that learning results from exposure to informative relations among events in the environment. These relations, or associations, are used to build an overall representation of the world. Organisms adjust their representations based on new, informative experiences.

Technically, the events which an organism is exposed to, and between which associations are established are conceived as 'cues' and 'outcomes'.

The crucial question already hinted at in Section 2.5 above is how this association process can be conceptualized and modeled. In NDL the association is achieved by using the so-called Rescorla-Wagner equations (Rescorla and Wagner 1972, Rescorla 1988b). The reasoning underlying these equations goes as follows. Cues may be absent or present, and outcomes may be present or absent, which means that a particular cue may co-occur with a particular outcome, or it may not. The association strength or 'weight' of an outcome increases with every time that the cue and the outcome co-occur and decreases whenever the cue occurs without the outcome. The changes of these weights over time are modeled by the Rescorla-Wagner equations such that the weight of a cue to some outcome at time point t+1 equals its weight at point t plus some change (as specified mathematically in the equations).[2]

At the end of the learning process a stable state is reached in which each outcome is associated with its final association strength. This final association strength is conceived as the activation for this outcome on the basis of the training with all cues and can be computed as the sum of all changes in the weights during learning.

Let us see how this works. We start with a set of words, i.e. a toy lexicon, some of them complex, some of them simplex, and we want to know whether the NDL model arrives at something that looks like a sensible morphological analysis without positing any morphological unit or operation.

Our data set is given in Table 1. It contains eight words, given in the column 'Words'. The column 'Cues' contains for each word a sequence of bigrams, i.e. orthographic digraphs (with hash marks representing word boundaries). Each bigram functions as a cue in the modeling process.[3] The column 'Outcomes' lists the meanings (or 'lexomes') corresponding to the words in the first column. These meanings are represented in a very simplified manner by simply listing the orthographic form of the word in inverted commas. For words with

---

2 We spare the reader the mathematical details, as they are not important for our discussion. The reader is referred to the original literature (e.g. Rescorla and Wagner 1972), or to Baayen et al. (2011), who introduce and discuss in more detail Rescorla-Wagner equations using linguistic examples.

3 We use digraphs as bigrams as if we were modeling reading comprehension. For modeling auditory comprehension, bigram cues could be represented by phoneme pairs, for example. For the purpose of our exercise nothing hinges on the choice of digrams instead of unigrams or trigrams.

Table 1: Toy lexicon (Frequencies are taken from the SUBTLEX-US corpus, Brysbaert and New 2009).

Words	Cues	Outcomes	Frequency
baptize	#b, ba, ap, pt, ti, iz, ze, e#	'baptize'	37
chance	#c, ch, ha, an, nc, ce, e#	'chance'	12303
extreme	#e, ex, xt, tr, re, em, me, e#	'extreme'	558
modernize	#m, mo, od, de, er, rn, ni, iz, ze, e#	'modern', 'make'	6
optimal	#o, op, pt, ti, im, ma, al, l#	'optimal'	15
optimize	#o, op, pt, ti, im, mi, iz, ze, e#	'optimal', 'make'	5
sand	#s, sa, an, nd, d#	'sand'	1035
size	#s, si, iz, ze, e#	'size'	2353

more than one clearly discernible meaning component, each meaning is separated from the other by a comma.

As the reader may notice when checking the table, the morphological phenomenon to be investigated here is the expression of causative meaning (paraphrased here as 'make'). Note that we have three words ending in what morphologists would call the suffix *-ize*, with one of the words being rather opaque, *baptize*, which is paraphrased by the OED as '[t]o immerse in water, or pour ... water upon, ... in token of initiation into a religious society, especially into the Christian Church'. The word thus does not straightforwardly combine the meaning of 'make' with the meaning of the base (which would generate something like 'make bapt', whatever 'bapt' might mean). Though somewhat opaque from a syntagmatic perspective, the word *baptize* may nevertheless be argued to be morphologically complex due to its relation to *baptism* and due to the fact that it is a verb, a predictable property if one assumes that the word ends in the suffix *-ize*. The meanings of *modernize* and *optimize*, in contrast, clearly involve two meaning components, as nicely illustrated by the OED paraphrases 'to make modern' and 'to render optimal'.

For each of the 40 different cues, e.g. for each of the bigrams in column 2, the model will compute association strengths with each of the outcomes, i.e. the nine different meanings. Overall, there are 360 connections for which the model arrives at an association weight. A subset of the network of cues and outcomes (with only four bigrams, four meanings and 16 connections) is shown in Figure 2, illustrating all connections but not the respective association weights.

Before turning to the overall results of our modeling exercise let us look at possible association weights between cues and outcomes with the help of Figure 2. As can be easily imagined, the bigram <nd> should be an excellent cue for the
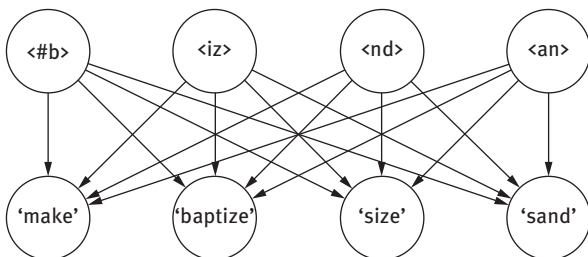
Figure 2: Association from cues to outcomes.

meaning 'sand' because this bigram only co-occurs (with a frequency of 1035) with the meaning 'sand'. In the model we should therefore expect a rather high association weight between this bigram and the meaning 'sand'. We also expect the bigram <an> to be a good cue for the meaning 'sand', but since this cue also co-occurs with one other meaning ('chance'), its association weight with 'sand' should be smaller than that of <nd>. The bigram <#b> should be a very good cue for 'baptize', but should not be a good cue for 'size'. In fact, we would even expect to see a negative association weight here, since its presence at the beginning of a word is likely to be a good cue that the meaning of the word will not be something like 'size'.

Using the ndl package (Arppe et al. 2014) for R (R Core Team 2014) we can compute the association weights after learning has reached a stable state (the R code to reproduce our model is given in the appendix). The result is a matrix which gives the association weight of each cue to each outcome. Four lines of this matrix (corresponding to the network shown in Figure 2) are given for illustration in Table 2.

Table 2: Weight matrix for four bigrams.

	'baptize'	'chance'	'extreme'	'make'	'modern'	'optimal'	'sand'	'size'
#b	0.213	−0.007	−0.004	−0.110	−0.016	−0.089	0.014	−0.062
iz	0.041	−0.032	−0.030	0.106	0.034	0.017	−0.019	0.128
nd	0.014	−0.030	0.006	0.036	0.012	0.006	0.220	−0.069
an	0.007	0.124	−0.009	0.017	0.006	0.003	0.190	−0.076

The weights range from 0.220 to -0.76. We can see that at the end of the learning process <#b> is positively associated with 'baptize' (with a rather high weight of 0.213) and negatively with most other meanings. <iz> is positively

associated with the meanings 'baptize', 'make', 'modern', 'optimal' and 'size'. Of these connections the association with 'make' is strongest (0.106). As expected above, the bigrams <an> and <nd> are excellent cues to the meaning 'sand', with <nd> reaching the highest weight (0.220).

The sum of the association weights of a particular word form can now be conceptualized as the degree of activation of a particular meaning by the string of bigrams corresponding to the word form in question. Full activation amounts to a value of 1. For example, if we take the string of bigrams #b_ba_ap_pt_ti_iz_ ze_e#, we can add up the weights of these bigrams for the meaning 'baptize' to arrive at the activation of this meaning by this word form, as shown in Table 3.

Table 3: Activation weights of bigrams for the meaning 'baptize'.

#b	ba	ap	pt	ti	iz	ze	e#	sum of weights
0.213	0.213	0.213	0.125	0.125	0.041	0.041	0.029	1

Table 4 shows a part of the weight matrix, giving the activation of meanings for four selected bigram strings. The weight sums are mostly extremely small, which means that there is no activation of the respective meaning by the cue. Some of the activation sums reach unity (the latter are given in bold), which means that these meanings are activated by the respective cues. For illustration, consider the joint activation of the meaning of 'chance' by the bigrams #b_ba_ap_pt_ti_iz_ ze_e#, which is vanishingly small. In contrast, the activation of the meaning 'baptize' by the same string is maximal. This is a welcome result.

Table 4: Activations of meanings by bigram strings.

	baptize	chance	make	modern	optimal
#b_ba_ap_pt_ti_iz_ze_e#	**1.00**	2.57e-16	0.00	9.71e-17	6.56e-16
#c_ch_ha_an_nc_ce_e#	−3.82e-17	**1.00**	1.18e-16	1.70e-16	3.56e-17
#m_mo_od_de_er_rn_ni_iz_ze_e#	4.09e-16	2.91e-16	**1.00**	**1.00**	−5.38e-16
#o_op_pt_ti_im_mi_iz_ze_e#	−3.82e-16	1.32e-16	**1.00**	−4.30e-16	**1.00**

From a morphological perspective the most important result is that the two strings #m_mo_od_de_er_rn_ni_iz_ze_e# and #o_op_pt_ti_im_mi_iz_ze_e# (and only those two strings) activate the meaning 'make', as well as the meanings

'modern' and 'optimal'. This result mirrors a classical morphological analysis, but crucially the model has arrived at this analysis without having posited or used any morphological parsing or structure. The model thus captures morphological patterns, but morphology actually remains implicit.

Our modeling exercise was meant to illustrate how an NDL approach can achieve morphologically interpretable results without assuming any morphological units, and without assuming any operations that are defined by making use of such units. Whether such a radical approach to word structure is feasible on a larger scale has been tested for some problems of derivational morphology in Baayen et al. (2011). In that paper, Baayen and colleagues model a wide range of inflectional, syntactic and word-formation phenomena, and we will concentrate here on two issues relating to derivational morphology. The first is the processing costs of derived words, the second the problem of phonaesthemes.

Traditionally, a number of diagnostic lexical measures are assumed to reflect morphological processes, and these measures crucially refer to entities that are taken to be morphologically relevant. Among these are, for example, frequency of the derived word, frequency of the base, morphological family size of the derivative,[4] morphological family size of the base, and the frequency of the letter bigram straddling the morpheme boundary between base and affix. In many experiments it has been shown that these measures can account for the distribution of behavioral measures in experimental tasks, such as reaction times (see again Section 2 for discussion).

The general question for us now is whether an NDL model can simulate the results observed in behavioral experiments correctly even though this model has no access to the morphological entities that underlie the traditional diagnostic measures. This question can be answered by relating reaction times to the activation weights derived in NDL models. Large positive activation weights can be taken to negatively correlate with reaction times since higher activation goes together with shorter reaction times.[5]

Using different data sets, Baayen et al. (2011) show that the simulated reaction times generated by the NDL model nicely correlate with the reaction times observed in experiments. NDL models thus closely approximate the effects of the lexical-morphological variables without assuming the morphological entities and morphology-based measures that underlie such variables.

---

4 Morphological family size is a type count of the number of morphologically related forms of a base (see, for example, Schreuder and Baayen 1997).

5 Depending on the properties of the data set at hand, and to achieve a better fit, it may be necessary to mathematically transform the association weights. See Baayen et al. (2011: 451) for discussion.

With regard to morphological productivity in derivation, NDL makes interesting and quite realistic predictions. Thus, in their study of processing and productivity Baayen et al. (2011) investigate the activation levels of the productive nominalizing suffix -*ness* (as in *randomness*) and the unproductive suffix -*th* (as in *truth*). While the activation for the productive suffix is higher, the activation for -*th* is not zero. This suggests two things. First, the model correctly predicts reduced processing costs for neologisms with productive suffixes. Second, new formations in -*th* might occasionally occur, which is also correct (cf. *strong*, *slowth*, *firmth*, and *oldth*, all from the Urban Dictionary, http://www.urbandiction ary.com).

To see how NDL can deal with phonaesthemes, Baayen et al. (2011) used data from the priming study by Bergen (2004), discussed in Section 2.4 above. The NDL model was able to simulate reaction times which strongly correlated with the reaction times in Bergen's experiment. This shows that priming effects also emerge under the conditions of discriminative learning, and they do so again in the absence of any morphologically defined unit.

To summarize, we can say that in NDL effects emerge that run under the name of 'morpheme' or 'phonaestheme' in other models, but without the need of assuming the existence of such units. The fact that NDL implements an established theory of learning makes this approach especially attractive. There are, however, also some problems. The first concerns the cues and outcomes. The ontological nature of the outcomes as "symbolic focal points mediating between linguistic form and experience of the world" (Baayen and Ramscar 2015: 9) is not quite clear. There must be some abstraction going on with regard to establishing 'symbolic' focal points that is not built into the system. Similarly, on the cue level, it is left open how the n-grams themselves are learned. Baayen and colleagues are aware of these problems and work around them by assuming that these elements are "simply available to the learner" (Baayen and Ramscar 2015: 9).

Another reason for worry is that an NDL approach ultimately seems to lead to the abandonment of what is commonly known as the 'mental lexicon', and it is unclear how an NDL model could capture the many different insights that people have gained under this perspective.

3.6 Summary

We have seen in this section that researchers have come up with a wide range of mechanisms to account for the mapping of sound and meaning in derived words. In view of the results emerging from empirical studies and computational implementations, it appears that widely established concepts such as

'rule', or even 'mental lexicon' are somewhat simplistic and in need of an over-haul. Especially the variation found in the data is a matter of concern that future research must address.

4 Where are we going?

Answering the question 'where are we going from here?' is hardly an easy one in any research field. For derivational morphology, several of the issues that we have discussed in this chapter have become matters of conviction in addition to being research questions, which does not make the task easier. Nonetheless, we shall try to sketch some key issues and avenues of research as we perceive them.

In general, it can be said that the findings from different subfields can hardly be reconciled with the reductionist views of most current theories. These views are reductionist in the sense that they try to reduce the often complex and gradient mappings of form and meaning to a manageable, even if simplistic, set of units and necessary mechanisms of combination. Formal modeling is one avenue for exploring how the multitude of effects arise and co-exist, sometimes in interactions. Here, we have focused on Naive Discriminative Learning, which has the advantage of being an application to language of a well-known learning model that has been used to account for a range of other phenomena in cognitive psychology. However, interesting and testable predictions may also be derived from other algorithms.

One important future challenge is to ensure that the models proposed are sufficiently able to account for different processing circumstances and a wide range of linguistic phenomena. Typically, linguists have been guilty of ignoring processing-related issues, while psychologists and psycholinguists have ignored the complexity of the linguistic system. A case in point of the former is the emphasis on problems of zero morphs and empty morphs (outlined in Sections 2.2 and 2.3), which are very much theory-internal linguistic problems without major implications for processing. An example of psycholinguists ignoring the complexity of the system is the focus of much experimental work on whether or not morphemes are relevant units of processing, with little attention paid to how such morphemes might then be combined to produce whole-word meaning, no trivial issue in derivation.

Another challenge is the fact that the psycholinguistic study of derivational morphology, and inevitably to some extent also this chapter, has often focused to a large extent on the visual recognition of morphologically complex words, with auditory comprehension and both spoken and written word production

being confined to secondary roles. This is in noteworthy contrast to theoretical morphology, which often focuses on word 'formation' and thus implicitly on word production. The ideal must surely be models that account for both auditory and visual processing in both comprehension and production with a certain degree of systematicity and consistency. A related concern is the interaction between visual and auditory modalities and the impact of literacy not only on the processing that we study in our experiments but also on our definitions of words and morphemes, and our very thinking about morphology.

Finally, derivational morphology and morphological processing of course do not exist in a vacuum, but in the context of many other linguistic and extra-linguistic influences, which should ideally be accounted for in the same integrated model. Some such influences are speaker and register variability. It has been shown, for example, that age plays a significant role in shaping the lexicon over time. Not only does age of acquisition have an important effect on long-term representation of words (e.g. Brysbaert et al. 2000), there are also significant changes in the size of the lexicon over time, with important consequences for, among other things, processing time (e.g. Ramscar et al. 2014). Morphological knowledge itself seems to vary across speakers as a function of education, and may interact with differences in orthographic knowledge or register. Plag et al. (1999), for example, show that productive derivation is very restricted in spoken registers, unlike in written registers.

Obviously, we are still far away from devising models that can cope with all sources of variation and calling for such models moves beyond the optimistic into downright daydreaming. However, whether daydream or not, it is an ideal that should be kept in mind to avoid getting lost in the complexity and minute details of derivational morphology.

Appendix: R code for NDL example

```
> install.packages("ndl")
> library(ndl)
```

```
> Word <- c("baptize", "chance", "extreme", "modernize", "optimal",
"optimize", "sand", "size")
> Outcomes <- c("baptize", "chance", "extreme", "modern make", "optimal",
"optimal make", "sand", "size")
> Frequency <- c(37, 12303, 558, 6, 15, 5, 1035, 2353)
> ize <- data.frame(Word, Outcomes, Frequency)
> ize$Cues <- orthoCoding(ize$Word, grams=2)
> ize
```

Word	Outcomes	Frequency	Cues
baptize	baptize	37	#b-ba_ap_pt_ti_iz_ze_e#
chance	chance	12303	#c_ch_ha_an_nc_ce_e#
extreme	extreme	558	#e_ex_xt_tr_re_em_me_e#
modernize	modernize	6	#m_mo_od_de_er_rn_ni_iz_ze_e#
optimal	optimal	15	#o_op_pt_ti_im_ma_al_l#
optimize	optimize	5	#o_op_pt_ti_im_mi_iz_ze_e#
sand	Sand	1035	#s_sa_an_nd_d#
size	Size	2353	#s_si_iz_ze_e#

```
> ize.w = estimateWeights(ize)
> round(ize.w, 3)
```

	baptize	chance	extreme	make	modern	optimal	sand	size
#b	0.213	-0.007	-0.004	-0.110	-0.016	-0.089	0.014	-0.062
#c	-0.007	0.154	-0.015	-0.019	-0.006	-0.003	-0.030	-0.006
#e	-0.004	-0.015	0.131	-0.011	-0.003	-0.002	0.006	-0.015
#m	-0.016	-0.006	-0.003	0.102	0.130	-0.007	0.012	-0.052
#o	-0.089	-0.003	-0.002	0.131	-0.007	0.199	0.006	-0.026
#s	-0.048	-0.036	-0.009	-0.126	-0.040	-0.020	0.150	0.284
al	0.006	0.010	0.006	-0.145	0.021	0.061	-0.019	0.084
an	0.007	0.124	-0.009	0.017	0.006	0.003	0.190	-0.076
ap	0.213	-0.007	-0.004	-0.110	-0.016	-0.089	0.014	-0.062
ba	0.213	-0.007	-0.004	-0.110	-0.016	-0.089	0.014	-0.062
ce	-0.007	0.154	-0.015	-0.019	-0.006	-0.003	-0.030	-0.006
ch	-0.007	0.154	-0.015	-0.019	-0.006	-0.003	-0.030	-0.006
d#	0.014	-0.030	0.006	0.036	0.012	0.006	0.220	-0.069
de	-0.016	-0.006	-0.003	0.102	0.130	-0.007	0.012	-0.052

(continued)

	baptize	chance	extreme	make	modern	optimal	sand	size
e#	0.029	0.106	0.085	0.076	0.024	0.012	-0.043	0.107
em	-0.004	-0.015	0.131	-0.011	-0.003	-0.002	0.006	-0.015
er	-0.016	-0.006	-0.003	0.102	0.130	-0.007	0.012	-0.052
ex	-0.004	-0.015	0.131	-0.011	-0.003	-0.002	0.006	-0.015
ha	-0.007	0.154	-0.015	-0.019	-0.006	-0.003	-0.030	-0.006
im	-0.089	-0.003	-0.002	0.131	-0.007	0.199	0.006	-0.026
iz	0.041	-0.032	-0.030	0.106	0.034	0.017	-0.019	0.128
l#	0.006	0.010	0.006	-0.145	0.021	0.061	-0.019	0.084
ma	0.006	0.010	0.006	-0.145	0.021	0.061	-0.019	0.084
me	-0.004	-0.015	0.131	-0.011	-0.003	-0.002	0.006	-0.015
mi	-0.094	-0.013	-0.007	0.276	-0.028	0.138	0.025	-0.110
mo	-0.016	-0.006	-0.003	0.102	0.130	-0.007	0.012	-0.052
nc	-0.007	0.154	-0.015	-0.019	-0.006	-0.003	-0.030	-0.006
nd	0.014	-0.030	0.006	0.036	0.012	0.006	0.220	-0.069
ni	-0.016	-0.006	-0.003	0.102	0.130	-0.007	0.012	-0.052
od	-0.016	-0.006	-0.003	0.102	0.130	-0.007	0.012	-0.052
op	-0.089	-0.003	-0.002	0.131	-0.007	0.199	0.006	-0.026
pt	0.125	-0.010	-0.006	0.021	-0.022	0.110	0.020	-0.088
re	-0.004	-0.015	0.131	-0.011	-0.003	-0.002	0.006	-0.015
rn	-0.016	-0.006	-0.003	0.102	0.130	-0.007	0.012	-0.052
sa	0.014	-0.030	0.006	0.036	0.012	0.006	0.220	-0.069
si	-0.062	-0.006	-0.015	-0.162	-0.052	-0.026	-0.069	0.353
ti	0.125	-0.010	-0.006	0.021	-0.022	0.110	0.020	-0.088
tr	-0.004	-0.015	0.131	-0.011	-0.003	-0.002	0.006	-0.015
xt	-0.004	-0.015	0.131	-0.011	-0.003	-0.002	0.006	-0.015
ze	0.041	-0.032	-0.030	0.106	0.034	0.017	-0.019	0.128

References

Abelin, Asa. 1999. *Studies in sound symbolism*. Göteborg University Dissertation.

Arndt-Lappe, Sabine. 2011. Towards an exemplar-based model of stress in English noun-noun compounds. *Journal of Linguistics* 47 (11). 549–585.

Arndt-Lappe, Sabine. 2014. Analogy in suffix rivalry: the case of English -ity and -ness. *English Language and Linguistics* 18 (3). 497–548. doi:10.1017/S136067431400015X.

Arndt-Lappe, Sabine. 2015. Word-formation and analogy. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, 822–841. Berlin: de Gruyter Mouton.

Arndt-Lappe, Sabine & Mirjam Ernestus. 2019. Morpho-phonological alternations: the role of lexical storage. In Vito Pirrelli, Ingo Plag & Wolfgang Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon*, 185–220. De Gruyter.

Arndt-Lappe, Sabine & Ingo Plag. 2013. The role of prosodic structure in the formation of blends. *English Language and Linguistics* 17. 537–563.

Arnon, Inbal & Michael Ramscar. 2012. Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition* 122 (3). 292–305.

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, Mass: MIT Press.

Arppe, Antti, Peter Hendrix, Petar Milin, Harald R. Baayen & Cyrus Shaoul. 2014. ndl: Naive Discriminative Learning. http://CRAN.R-project.org/package=ndl.

Baayen, R. Harald, Ton Dijkstra & Robert Schreuder. 1997. Singulars and Plurals in Dutch: Evidence for a Parallel Dual-Route Model. *Journal of Memory and Language* 37. 94–117.

Baayen, R. Harald., Peter Hendrix & Michael Ramscar. 2013. Sidestepping the Combinatorial Explosion: An Explanation of n-gram Frequency Effects Based on Naive Discriminative Learning. *Language and Speech* 56 (3). 329–347. doi:10.1177/0023830913484896.

Baayen, R. Harald, Petar Milin, Dusica Filipovic Durdevic, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118 (3). 438–481.

Baayen, R. Harald & Michael Ramscar. 2015. Abstraction, storage and naive discriminative learning. In Ewa Dabrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistic*s, vol. 39, 100–120. Walter de Gruyter GmbH & Co KG.

Baayen, R. Harald, Shaoul C. J. Willits, Jon Willits & Michael Ramscar. 2015. Comprehension without segmentation: A proof of concept with naive discrimination learning. *Language, Cognition, and Neuroscience* 31 (1). 106–128.

Baayen, R. Harald, Lee H. Wurm & Joanna Aycock. 2007. Lexical Dynamics for Low-Frequency Complex Words: A Regression Study Across Tasks and Modalities. *The Mental Lexicon* 2. 419–463.

Balling, Laura Winther & R. Harald Baayen. 2008. Morphological effects in auditory word recognition: Evidence from Danish. *Language and Cognitive Processes* 23. 1159–1190.

Balling, Laura W. & R. Harald Baayen. 2012. Probability and surprisal in auditory comprehension of morphologically complex words. *Cognition* 125. 80–106.

Barker, Chris. 1998. Episodic -ee in English: A thematic role constraint on a new word formation. *Language* 74 (4). 695–727.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.

Beard, Robert. 1990. The nature and origins of derivational polysemy. *Lingua* 81. 101–140.

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

Bergen, Benjamin K. 2004. The psychological reality of phonaesthemes. *Language* 80. 290–311.

Bien, Heidrun, Willem M. J. Levelt & R. Harald Baayen. 2005. Frequency effects in compound production. *Proceedings of the National Academy of Sciences of the USA* 102. 17876–17881.

Blazej, Laura J. & Ariel M. Cohen-Goldberg. 2015. Can we hear morphological complexity before words are complex? *Journal of Experimental Psychology: Human Perception and Performance* 41 (1). 50–68.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42 (3). 531–573.

Blevins, James P., Farrell Ackerman & Robert Malouf. 2015. Morphology as an adaptive discriminative system. In Heidi Harley & Daniel Siddiqi (eds.), *Morphological metatheory*, 271–302. Amsterdam and Philadelphia: John Benjamins.

Bölte, Jens, Petra Dohmes & Pienie Zwitserlood. 2013. Interference and Facilitation in Spoken Word Production: Effects of Morphologically and Semantically Related Context Stimuli on Picture Naming. *Journal of Psycholinguistic Research* 42. 255–280.

Bonami, Olivier & Berthold Crysmann. 2016. The role of morphology in constraint-based lexicalist grammars. In Andrew Hippisley & Gregory T. Stump (eds.), *Cambridge Handbook of Morphology*, 609–656. Cambridge: Cambridge University Press.

Booij, Geert E. 1977. *Dutch Morphology. A Study of Word Formation in Generative Grammar*. Dordrecht: Foris.

Booij, Geert E. 1993. Against split morphology. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1993*, 27–50. Dordrecht: Kluwer Academic Publishers.

Booij, Geert E. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Brysbaert, Marc, Marielle Lange & Ilse van Wijnendaele. 2000. The effects of age-of-acquisition and frequency-of-occurence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology* 12. 65–85.

Brysbaert, Marc & Boris New. 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.

Bybee, Joan L. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10. 425–455.

Chapman, Don & Royal Skousen. 2005. Analogical modeling and morphological change: the case of the adjectival negative prefix in English. *English Language and Linguistics* 9 (2). 333–357.

Chen, Train-Min. & Jenn Yeu. Chen. 2006. Morphological encoding in the production of compound words in Mandarin Chinese. *Journal of Memory and Language* 54. 491–514.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Collie, Sara. 2008. English stress preservation: The case for 'fake cyclicity'. *English Language and Linguistics* 12 (3). 505–532.

Costa, Alberto, F.-Xavier Alario & Alfonso Caramazza. 2005. On the categorical nature of the semantic interference effect in the picture-word interference paradigm. *Psychonomic Bulletin and Review* 12. 125–131.

Daelemans, Walter, Jakub Zavrel, Ko van der Sloot & Antal van den Bosch. 2007. TiMBL: Tilburg Memory Based Learner, version 6.0, *Reference Guide: LK Technical Report 04-02*. Tilburg: ILK. available from http://ilk.uvt.nl/timbl.

Dohmes, Petra, Pienie Zwitserlood & Jens Bölte. 2004. The impact of semantic transparency of morphologically complex words on picture naming. *Brain and Language* 90. 203–212.

Don, Jan. 1993. *Morphological Conversion*. Utrecht: OTS, Rijksuniversiteit.

Dressler, Wolfgang, Gary Libben & Katharina Korecky-Kröll. 2014. Conflicting vs. convergent vs. interdependent motivations in morphology. In Brian MacWhinney, Andrej Malchukov & Edith Moravcsik (eds.), *Competing Motivations in Grammar and Usage*, 181–196. Oxford: Oxford University Press.

Eddington, David. 2002. Spanish diminutive formation without rules or constraints. *Linguistics* 40. 395–419.

Fábregas, Antonio & Martina Penke. 2019. Word storage and computation. In Vito Pirrelli, Ingo Plag & Wolfgang Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon*, 444–494. Berlin/Boston: De Gruyter.

Feldman, Laurie Beth. 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, Laurie Beth, Petar Milin, Kit W. Cho, Fermín Moscoso del Prado Martín & Patrick O'Connor. 2015. Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers of Human Neuroscience* 9. 111. doi: 10.3389/fnhum.2015.00111.

Frauenfelder, Ulrich H. & Robert Schreuder. 1992. Constraining Psycholinguistic Models of Morphological Processing and Representation: The Role of Productivity. In Geert E. Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1991*, 165–183. Dordrecht: Kluwer Academic Publishers.

Frazier, Melissa. 2006. Output-output faithfulness to moraic structure: Evidence from American English. In *Proceedings of NELS*, vol. 36, 1.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100. 25–50.

Giraudo, Hélène & Jonathan Grainger. 2001. Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review* 8. 127–131.

Gonnerman, Laura M. & Elaine S. Anderson. 2001. Graded semantic and phonological similarity effects in morphologically complex words. In Sabrina Bendjaballah, Wolfgang U. Dressler, Oskar E. Pfeiffer & Maria D. Voeikova (eds.), M*orphology 2000: Selected papers from the 9th Morphology meeting*, 137–148. Amsterdam: John Benjamins.

Gonnerman, Laura M., Mark S. Seidenberg & Elaine S. Anderson. 2007. Graded Semantic and Phonological Similarity Effects in Priming: Evidence for a Distributed Connectionist Approach to Morphology. *Journal of Experimental Psychology: General* 136. 323–345.

Gussmann, Edmund (ed.). 1987a. *Rules and the Lexicon*. Lublin: Catholic University of Lublin.

Gussmann, Edmund. 1987b. The lexicon of English de-adjectival verbs. In Edmund Gussmann (ed.), *Rules and the Lexicon*, 79–101. Lublin: Catholic University of Lublin.

ten Hacken, Pius. 2014. Delineating derivation and inflection. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*. Oxford Handbooks in linguistics, 10–25. Oxford: Oxford University Press.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & Samuel J. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, vol. 24 Current Studies in Linguistics, 111–176. Cambridge and Mass: MIT Press.

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

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45 (1). 31–80.

Hay, Jennifer. 2001. Lexical frequency in morphology: is everything relative? *Linguistics* 39 (6). 1041–1070.

Hay, Jennifer B. 2003. *Causes and Consequences of Word Structure*. New York and London: Routledge.

Hay, Jennifer & Ingo Plag. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory* 22 (3). 565–596.

Hockett, Charles F. 1947. Problems of Morphemic Analysis. *Language* 23 (4). 321–343. http://www.jstor.org/stable/410295.

Hutchins, Sharon Suzanne. 1998. *The psychological reality, variability, and compositionality of English phonesthemes*. Atlanta: Emory University Dissertation.

Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51 (3). 639–671.

Janssen, Niels, Yanchao Bi & Alfonso Caramazza. 2008. A tale of two frequencies: Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes* 23. 1191–1223.

Joanisse, Mark F. & Mark S. Seidenberg. 1999. Impairments in verb morphology after brain injury: a connectionist model. *Proceedings of the National Academy of Science* 96. 7592–7597.

Kamin, Leon J. 1968. Attention-like processes in classical conditioning. In M. R. Jones (ed.), *Miami symposium on the prediction of behavior*, 9–31. Miami: Miami University Press.

Kemps, Rachèl J. J. K, Lee H. Wurm, Mirjam Ernestus, Robert Schreuder & R. Harald Baayen. 2005. Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes* 20 (1–2). 43–73.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In In-Seok Yang (ed.) *Linguistics in the Morning Calm: Selected Papers from SICOL*, 3–91. Seoul: Hanshin.

Krott, Andrea, R. Harald Baayen & Robert Schreuder. 2001. Analogy in morphology: Modeling the choice of linking morphemes in Dutch. *Linguistics* 39. 51–93.

Krott, Andrea, Peter Hagoort & R. Harald Baayen. 2004a. Sublexical units and supralexical combinatorics in the processing of interfixed Dutch compounds. *Language and Cognitive Processes* 19. 453–471.

Krott, Andrea, Gary Libben, Gonia Jarema, Wolfgang U. Dressler, Robert Schreuder & R. Harald Baayen. 2004b. Probability in the grammar of German and Dutch: Inter-fixation in tri-constituent compounds. *Language and Speech* 47. 83–106.

Kuperman, Victor, Raymond Bertram & R. Harald Baayen. 2008. Morphological Dynamics in Compound Processing. *Language and Cognitive Processing* 23. 1089–1132.

Kwon, Nahyun & Erich R. Round. 2015. Phonaesthemes in morphological theory. *Morphology* 25 (1). 1–27. doi:10.1007/s11525-014-9250-z.

Lehiste, Ilse. 1972. *Suprasegmentals*. Cambridge and Mass: MIT Press.

Lehtonen, Minna, Helga Niska, Erling Wande, Jussi Niemi & Matti Laine. 2006. Recognition of Inflected Words in a Morphologically Limited Language: Frequency Effects in Monolinguals and Bilinguals. *Journal of Psycholinguistic Research* 35. 121–146.

Libben, Gary. 2006. Why study compound processing? In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, 1–22. Oxford: Oxford University Press.

Libben, Gary, Christina Gagné & Wolfgang U. Dressler. 2019. The representation and processing of compounds words. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental*Lexicon, 329–345. Berlin/Boston: De Gruyter.

Lieber, Rochelle. 1998. The suffix *-ize* in English: Implications for morphology. In Steven Lapointe, Diane Brentari & Patrick Farrell (eds.), *Morphology and its Relation to Phonology and Syntax*, 12–34. CSLI publications.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

Lieber, Rochelle & Pavol Štekauer. 2009. Introduction: Status and definition of compounding. In Rochelle Lieber & Pavol Stekauer (eds.), *The Oxford handbook of compounding Oxford Handbooks in linguistics*, 3–18. Oxford, New York: Oxford University Press.

Longtin, Catherine-Marie, Juan Segui & Pierre A. Hallé. 2003. Morphological priming without morphological relationship. *Language and Cognitive Processes* 18. 313–334.

Magnus, Margaret. 2000. *What's in a Word? Evidence for Phonosemantics*. Trondheim: University of Trondheim Dissertation.

Marangolo, Paola & Costanza Papagno. 2020. Neuroscientific protocols for exploring the mental lexicon: evidence from aphasia. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon*, 124–163. Berlin/Boston: De Gruyter.

Marslen-Wilson, William, Lorraine K. Tyler, Rachelle Waksler & Lianne Older. 1994. Morphology and Meaning in the English Mental Lexicon. *Psychological Review* 191. 3–33.

Marslen-Wilson, William & Xiaolin Zhou. 1999. Abstractness, Allomorphy, and Lexical Architecture. *Language and Cognitive Processes* 14. 321–352.

Marzi, Claudia, James P. Blevins, Geert Booij & Vito Pirrelli. 2020. Inflection at the morphology-syntax interface. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon*, 221–287. Berlin/Boston: De Gruyter.

McClelland, James L. & Jeffrey L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18. 1–86.

Merlini Barbaresi, Lavinia & Wolfgang U. Dressler. 2020. Pragmatic Explanations in Morphology. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon*, 397–443. Berlin/Boston: De Gruyter.

Meunier, Fanny & Juan Segui. 1999. Frequency Effects in Auditory Word Recognition: The Case of Suffixed Words. *Journal of Memory and Language* 41. 327–344.

Moscoso del Prado Martín, Fermín, Raymond Bertram, Tuomo Häikiö, Robert Schreuder & R. Harald Baayen. 2004. Morphological Family Size in a Morphologically Rich Language: The Case of Finnish Compared With Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30. 1271–1278.

Moscoso del Prado Martín, Fermín, Avital Deutsch, Ram Frost, Robert Schreuder, Nivja H. de Jong & R. Harald Baayen. 2005. Changing places: A cross-language perspective on frequency and family size in Hebrew and Dutch. *Journal of Memory and Language* 53. 496–512.

Motsch, Wolfgang. 1981. Der kreative Aspekt in der Wortbildung. In Leonhard Lipka (ed.), *Wortbildung*, 94–118. Darmstadt: Wissenschaftliche Buchgesellschaft.

Mühleisen, Susanne. 2010. *Heterogeneity in word-formation patterns*. Amsterdam / Philadelphia: John Benjamins.

New, Boris, Marc Brysbaert, Juan Segui, Ludovic Ferrand & Kathleen Rastle. 2004. The Processing of singular and plural nouns in French and English. *Journal of Memory and Language* 51. 568–585.

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

OED. 2013. *The Oxford English Dictionary online*. Oxford: Oxford University Press.

Pastizzo, Matthew John & Laurie Beth Feldman. 2001. Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. *Journal of Experimental Psychology: Learning, Memory and Cognition* 28. 244–249.

Pastizzo, Matthew John & Laurie Beth Feldman. 2009. Multiple dimensions of relatedness among words: Conjoint effects of form and meaning in word recognition. *The Mental Lexicon* 4 (1). 1–25. doi:10.1075/ml.4.1.01pas.

Paul, Herrmann. 1880. *Prinzipien der Sprachgeschichte*. Halle: Max Niemeyer.

Pearce, John M. & Mark E. Bouton. 2001. Theories of associative learning in animals. *Annual review of psychology* 52 (1). 111–139.

Pinker, Steven & Michael T. Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences* 6. 456–463.

Pirrelli, Vito, Claudia Marzi, Marcello Ferro, R. Harald Baayen & Petar Milin. 2020. Psycho-computational modelling of the mental lexicon. A discriminative learning perspective. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon*, 21–80. Berlin/Boston: De Gruyter.

Plag, Ingo. 1999. More on infinitives in creole. The nature of Sranan fu and its complements. In Pauline Christie, Barbara Lalla, Velma Pollard & Lawrence Carrington (eds.), Studies in Caribbean Language II. *Papers from the Ninth Biennial Conference of the Society for Caribbean Linguistics, 1992*, 250–264. St. Augustine: Society for Caribbean Linguistics.

Plag, Ingo. 2004. Review of Carlin, Eithne B. and Jacques Arends. (eds.): *Atlas of the Languages of Suriname*. Leiden: KITLV Press 2000. Zeitschrift f¨ur Sprachwissenschaft 23. 144–145.

Plag, Ingo. 2010. Compound stress assignment by analogy: the constituent family bias. *Zeitschrift für Sprachwissenschaft* 29 (2). 243–282.

Plag, Ingo. 2018. *Word-formation in English. 2nd ed*. Cambridge: Cambridge University Press.

Plag, Ingo & Harald Baayen. 2009. Suffix ordering and morphological processing. *Language* 85 (1). 109–152.

Plag, Ingo, Christiane Dalton-Puffer & R. Harald Baayen. 1999. Morphological productivity across speech and writing. *English Language and Linguistics* 3 (2). 209–228.

Plag, Ingo, Julia Homann & Gero Kunter. 2015. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53 (1). 181–216. (doi:10.1017/S0022226715000183).

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

R Core Team. 2014. R: *A Language and Environment for Statistical Computing*. http://www.R-project.org/.

Ramscar, Michael, Melody Dye & Stewart M. McCauley. 2013. Error and expectation in language learning: The curious absence of mouses in adult speech. *Language* 89 (4). 760–793.

Ramscar, Michael, Peter Hendrix, Cyrus Shaoul, Petar Milin & R. Harald Baayen. 2014. The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in cognitive science* 6 (1). 5–42.

Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny & Kirsten Thorpe. 2010. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34 (6). 909–957.

Rastle, Kathleen & Matthew H. Davis. 2008. Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes* 23. 942–971.

Rastle, Kathleen, Matthew H. Davis, William D. Marslen-Wilson & Lorraine K. Tyler. 2000. Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes* 15. 507–537.

Ravid, Dorit, Emmanuel Keuleers & Wolfgang U. Dressler. 2020. Emergence and early development of lexicon and morphology. In Vito Pirrelli, Ingo Plag & Wolfgang U. Dressler (eds.), *Word Knowledge and Word Usage: a Cross-disciplinary Guide to the Mental Lexicon*, 582-622. Berlin/Boston: De Gruyter.

Rescorla, Robert A. 1988a. Behavioral studies of Pavlovian conditioning. *Annual Review of Neuroscience* 11 (1). 329–352.

Rescorla, Robert A. 1988b. Pavlovian conditioning: It's not what you think it is. *American Psychologist* 43 (3). 151–160.

Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black & William F. Prokasy (eds.), *Classical conditioning II: Current research and theory*, 64–99. New York: Appleton-Century-Crofts.

Roelofs, Ardi. 1996. Serial order in planning the production of successive morphemes of a word. *Journal of Memory and Language* 35. 854–876.

Ronneberger-Sibold, Elke. 2003. On useful darkness. Loss and destruction of transparency by linguistic change, borrowing, and word creation. In Jaap van Marle & Geeert E. Booij (eds.), *Yearbook of Morphology 2002*, 81–104. Dordrecht: Kluwer Academic Publishers.

Ryding, Karin C. 2005. *A reference grammar of modern standard Arabic*. Cambridge University Press.

Schreuder, Robert & R. Harald Baayen. 1997. How complex simplex words can be. *Journal of Memory and Language* 37. 118–139.

Schriefers, Herbert, Antje S. Meyer & Willem J. M. Levelt. 1990. Exploring the time course of lexical access in production: Picture-word interference studies. *Journal of Memory and Language* 29. 86–102.

Seidenberg, Mark S. & James L. McClelland. 1989. A distributed, developmental model of word recognition and naming. *Psychological Review* 96. 523–568.

Skousen, Royal & Thereon Stanford. 2007. AM:: Parallel. available from http://humanities. byu.edu/am/.

Štekauer, Pavol, Salvador Valera & Livia Kortvélyessy. 2012. *Word-formation in the world's languages: a typological survey*. Cambridge: Cambridge University Press.

Synnaeve, Gabriel, Isabelle Dautriche, Benjamin Börschinger, Mark Johnson, Emmanuel Dupoux et al. 2014. Unsupervised word segmentation in context. In *25th International Conference on Computational Linguistics: Technical Papers*, Dublin: Dublin City University and Association for Computational Linguistics.

Szymanek, Bogdan. 1985. *English and Polish adjectives. A study in lexicalist word-formation*. Lublin: Catholic University of Lublin.

Taft, Marcus. 1979. Lexical Access via an Orthographic Code: The Basic Orthographic Syllabic Structure (BOSS). *Journal of Verbal Learning and Verbal Behavior* 18. 21–39.

Taft, Marcus. 1994. Interactive-activation as a Framework for Understanding Morphological Processing. *Language and Cognitive Processes* 9. 271–294.

Tremblay, Antoine, Bruce Derwing, Gary Libben & Chris Westbury. 2011. Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning* 61 (2). 569–613.

Ullman, Michael T. 2001. The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30. 37–69.

Ullman, Michael T. 2004. Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92. 231–270.

de Vaan, Laura, Robert Schreuder & R. Harald Baayen. 2007. Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon* 2. 1–23.

Vannest, Jennifer, Raymond Bertram, Juhani Järvikivi & Jussi Niemi. 2002. Counterintuitive Cross-Linguistic Differences: More Morphological Computation in English than in Finnish. *Journal of Psycholinguistic Research* 31. 83–106.

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

Zwitserlood, Pinie, Jens Bölte & Petra Dohmes. 2000. Morphological effects on speech production: evidence from picture naming. *Language and Cognitive Processes* 15. 563–591.

Zwitserlood, Pinie, Jens Bölte & Petra Dohmes. 2002. Where and how morphologically complex words interplay with naming pictures. *Brain and Language* 81. 358–367.