11 Models of Lexical Access and Morphological Processing

PETAR MILIN[1], EVA SMOLKA[2], AND LAURIE BETH FELDMAN[3]

[1] The University of Sheffield
[2] University of Konstanz
[3] The University at Albany, State University of New York, and Haskins Laboratories

Introduction

Humans possess the ability to discriminate among meanings for a practically unlimited number of possible utterances. This ability to make and understand new combinations defines productivity and has been extensively studied with respect to combining words into sentences. However, an analogous ability exists at the level of combining morphemes into words.[1] This is the essence of a *combinatorial* approach to language phenomena. Its point of departure is an assumption that language is a hierarchical system, where at each level in the hierarchy one can find distinct units, each of which makes an "imprint" in the mind, that is, brain.

In the combinatorial tradition, many studies in psycholinguistics use words composed of multiple morphemes, defined as morphologically complex, that have an intriguing property that their "meaning" can range from completely semantically compositional or "transparent," as in SUCCESS-FUL to relatively "opaque," as in SUCCESS-OR.[2] Given that the meaning of semantically opaque words cannot be derived from the meaning of their parts, linguistic, psycholinguistic, and neurolinguistic researchers, who adhere to the combinatorial tradition, are intrigued by questions such as: (i) how are the units at each level in the hierarchy represented and organized in the mental lexicon (e.g., are words stored as whole-word or as constituent units), and (ii) whether their combinatorics entail linguistic patterning (syntax) and/or a more general mechanism of association?

The Handbook of Psycholinguistics, First Edition. Edited by Eva M. Fernández and Helen Smith Cairns.

We can compare that approach to learning-based models where the emphasis is on how existing knowledge guides what functions as relevant units and how, with subsequent learning, those early units get better-differentiated one from another. In essence, learning-based models of word recognition emphasize progressive differentiation of larger chunks into smaller units and then use those to control the uncertainty in communication (see, e.g., Ramscar & Baayen, 2013; Ramscar & Port, 2015). Further, depending on what is already known, new relations can be easy or difficult to learn depending on how easy it is to differentiate them as distinct (c.f., Ellis, 2006; Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017).

In the present chapter, we sample from the experimental literature using the lexical decision task with masked and unmasked priming techniques to investigate how we understand morphologically complex words and discuss how those findings inform both lexicon-based and learning-based models of word knowledge (the mental lexicon). Then we focus on behavioral and electrophysiological measures on derived words to discuss experimental findings pertaining to the processing of morphologically complex words with and without so-called semantic transparency. How best to capture knowledge about the patterns that form complex words and the extent to which the underlying process is specific to language or particular types of linguistic units is a theme that recurs throughout the domain of psycholinguistics. Discussion is particularly animated in the domain of morphology, therefore, word recognition for morphologically complex words is where we concentrate.

Languages can differ in the way in which morphemes combine and in the complexity of those patterns. In Indo-European languages with concatenative morphology like English, French, and German, morphemes are appended to one another in a more or less, linear fashion. In the combinatorial tradition, again, the creation of new words from existing stems entails attaching either prefixes or suffixes to a stem that captures the core meaning of the word, as in UN + SUCCESS + FUL + LY. In contrast, in Semitic languages like Hebrew and Arabic, words consist of two abstract morphemes, the root and the word pattern, that are intertwined one within the other. The root conveys semantic aspects of the word, while the word pattern carries phonological and morpho-syntactic information. For example, the root ZMR (relating to "sing") can be inserted into different word patterns to derive a large family of Hebrew words like ZiMeR ("sing"), hiZdaMeR ("sounded like a song"), ZaMaR ("singer"), ZeMeR ("song"), ZiMRa ("singing"), and tiZMoRet ("orchestra"). The way the root and word pattern combine is nonlinear in that the letters of the root are no longer contiguous. Some have claimed that the nonlinear (nonconcatenative) manner in which morphemes combine in a language such as Hebrew affects the way a word is processed and represented (e.g., Frost, Kugler, Deutsch, & Forster, 2005).

Inflection and derivation are two well-studied processes to form morphologically complex words. Inflectional affixes specify, among other small changes, the number or tense of an event, as in CATS or WALKED. That is, they introduce only minimal changes to the content (i.e., traditional "meaning") of the communication (e.g., from one to more in CAT-CATS, and from present to past tense in

WALK-WALKED). Derivational affixes form new words and occur in Indo-European languages like English, French, and German as prefixes or suffixes that append to base words. For example, the suffix -MENT often derives a noun that reflects the process or state of the underlying verb (e.g., DEVELOPMENT, RESENTMENT, ENLARGEMENT). In comparison to inflected forms, the semantic relation between derivations and their base is not as straightforward. For example, the suffix -MENT sometimes produces abrupt changes in the message, as from DEPART to DEPARTMENT.

Below we contrast lexicon-based (i.e., combinatorial) and learning-based approaches to the processing of prefixed and suffixed derived words. The lexicon-based approach to word processing follows in the combinatorial tradition where lexical entries like {WALK} carry the core meaning of a stem, from which further words and meanings are derived by attaching affixes like ER or ABLE. Within the learning-based approach we can distinguish parallel-distributed connectionist models (such as the Triangle model of Harm & Seidenberg, 2004), and discrimination learning models (like the Naive Discrimination Learning model—NDL of Baayen, Milin, Filipović Đurđević, Hendrix, & Marelli, 2011). The later make use of discrimination to understand and explain the dynamics of language, where, in essence, language is learned (acquired) through discrimination of contrastive properties rather than by the "content" of representations. In other words, for the NDL approach invariant meanings cannot be *encrypted* in utterances, and they cannot be mapped combinatorially (or compositionally) onto linguistic units at different levels of granularity. Instead, the essential form to meaning mapping is abstract and relational: elaborated meaning emerges only across communicative contexts (c.f., Ramscar & Baayen, 2013; Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017). The NDL framework does not assume that lexical entries carry an invariant core of meaning, such as a core meaning for WALK plus a modification for past tense in ED. Rather, the listener's uncertainty about the speaker's intentions is reduced or even eliminated when she hears *John walked me to the train station*, and not *John walked me through his new book*. With respect to morphology, an appreciation of the variable form of morphologically related words such as RUN and RAN or RUNNER and RUNNABLE derives from the similarity of the contexts in which they appear relative to those in which, for example, WALK and WALKED or WALKER and WALKABLE are possible.

Methodological issues

The lexical decision task

Much of what we know about morphological processing comes from the outcomes of experiments where participants perform the lexical decision task. In this task, participants judge the lexical status of the letter string presented as the "target" (i.e., they make a "lexical decision" whereby they determine whether the letter string is a real word) and latencies to reach a decision along with judgment accuracy to each target constitute the typical dependent measures. In recent years, more and more often, electrophysiological brain responses (e.g., EEG) to "lexical events" supplement decision latencies and accuracies.

Many describe response latency as encompassing the time it takes to access a word in the lexicon and to ascertain that the lexical entry corresponds to the word that was presented (e.g., Rueckl & Galantucci, 2005). Interpretation of the data provides useful insights into a reader's lexical knowledge with an emphasis on words and how they are represented. Researchers introduce manipulations of context or presentation conditions and those contrasts help them to understand the underlying processes. For example, most typically targets appear after a single word that is called a *prime*.

The contrast of interest in any primed lexical decision task is the same target in the context of a related and an unrelated prime. For example, how does the determination of lexical status for the target SUCCESS differ depending on whether the prime is related, like SUCCESSFUL, or unrelated, like NEGLECTFUL? Any facilitatory or inhibitory effect (i.e., faster versus slower decision latencies relative to the unrelated prime) could occur because SUCCESSFUL and SUCCESS share form, meaning, or both. Hence, to differentiate between form and meaning, a further type of prime, like SUCCESSOR, would be introduced, that shares form but not meaning with the target SUCCESS. When there are multiple types of related primes, typically, "between-target" designs are adopted (e.g., Rastle, Davis, Marslen-Wilson, & Tyler, 2000) in which each target appears with one related and one unrelated prime and different targets appear with different types of related primes. For example, the target SUCCESS would appear with the semantically transparent prime SUCCESSFUL (and a matched unrelated prime), while the target DEPART would appear with the semantically opaque prime DEPARTMENT (and a matched unrelated prime). The typical way to control for differences targets in these experiments has been to eliminate significant differences between condition means for attributes such as word frequency and number of words that are similar in written (orthographic) or spoken (phonological) form in the various conditions and then to compare target decision latencies and accuracy judgments across conditions.

An alternative design applies a "within-target" manipulation and presents the same target in multiple related prime contexts. For example, the target SUCCESS would appear to different participants, with a prime that is transparent and thus semantically related (SUCCESSFUL), with a prime that is opaquely related (SUCCESSOR) and with a prime that is unrelated to the target word (NEGLECTFUL or MANAGER; for details, consult Feldman, 2000). The obvious advantage of comparing decision latencies to the *same* target across the different related prime contexts is that individual (distributional) differences among target words can be better controlled so as to avoid potential confounding of target properties with prime type. In principle, this confounding issue can also be controlled by statistical means. Rigorous implementations are infrequent, however (but see Feldman, O'Connor, & Moscoso del Prado Martín, 2009; Milin *et al.*, 2017).

Prime presentation

Presentation duration for the prime and the presence or absence of a pattern mask further define the processing condition under which a participant makes a lexical decision to the target. Prime-target pairs can be presented visually or auditorily or in a cross-modal format. Particularly well-explored presentation conditions for the prime include manipulations of exposure duration of the prime and the presence or absence of pattern mask that precedes it.

The masked priming procedure

The forward-masked lexical decision task is the most well accepted procedure to study early phases of word recognition. Here, a pattern mask appears for about 500 ms, then the prime appears in a lower-case equal-spaced font for about 48 ms, and finally a target printed in upper case is visible for 500 ms or more (Forster, Davis, Schoknecht, & Carter, 1987). For some, prime durations shorter than about 70 ms after a pattern mask meet the experimental conditions that qualify as "early" processing in this task. As is typical in other variants of the lexical decision task, latency, accuracy, and electrophysiological brain responses (EEG) to judge the lexicality of the target are measured.

Proponents of the forward-masked priming task argue that because the prime is not consciously perceived, participants cannot used it for strategic processing such as anticipating the upcoming target (Forster *et al.*, 1987). However, strategic effects can arise even in the forward-masked lexical decision task. For example, even at prime durations of 48 ms with a mask, the proportion of related prime-target pairs influences the magnitude of the difference in decision latencies between targets after related and unrelated primes. Facilitation increases as the proportion of related trials increases (Feldman & Basnight-Brown, 2008; Bodner & Masson, 2003). The implication is that processing of a masked prime can be informative about the orthographic and semantic properties of the target even though the prime is not available for conscious processing because of the mask. Despite its limitations, the forward-masked priming task provides the evidence for much of what we know about how morphologically complex words are understood during early stages of the word recognition process.

Overt priming

In contrast to masked priming, unmasked or overt priming taps into lexical processing and representation in addition to early recognition. The prime is presented either auditorily or visually at a relatively long exposure duration (230 ms or longer) and is presented either immediately preceding the target (i.e., "immediate repetition priming") or with other items intervening between prime and target (i.e., "long lag priming"). In either case, the long exposure duration for the prime allows that it be consciously perceived. As a result, this procedure is vulnerable to conscious and strategic processing. One means to attenuate the anticipation of a particular target is to drastically reduce the proportion of related prime-target pairs (e.g., 25%) in the set of experimental materials (Napps & Fowler, 1987). Because processing is conscious, overt priming is used to examine not only lexical processing but also deeper, integrative processes across words. As in masked priming, most often response latency, accuracy and brain responses to targets (and sometimes to primes) are measured. In both masked and overt priming techniques, the critical measure refers to the difference between the latency/accuracy/brain response to targets following unrelated vs. related primes, this difference is called the priming effect.

Lexicon-based and learning-based models of word recognition

The productivity of a language derives from the numerous ways in which a finite set of units can combine to form a new message. As noted above, productivity is characteristic not only at the level of combining words into sentences but also at the level of combining morphemes into words. For example, we can easily coin new words by combining morphemes like UNTAPABLE, or by combining words as in the compound SCHOOLHOUSE ROOF COLOR. Models of word recognition diverge on the question of how to represent morphologically complex words and the units from which they are composed and on the potential role that morphological rules play in describing the ways in which units combine. Likewise, models differ as to whether knowledge about morphemes is explicitly represented as lexical knowledge (Butterworth, 1983; Taft & Forster, 1975). Models further differ with respect to whether or not they require language-specific morphological rules to account for differences between experiments conducted with materials from English and from other languages, where the overall tendency to combine morphemes productively or the manner for forming those combinations (such as linear concatenation) differs from that of English (Frost, 2012; Feldman & Moscoso del Prado Martin, 2012; Smolka & Libben, 2017; Smolka, Preller, & Eulitz, 2014).

Below we review several basic types of models. These models differ with respect to the role granted to lexical representations and rules versus learning and the discovery of statistical patterning. Crucially, however, some of them are representative of a combinatorial and others of a discrimination tradition. One obvious manifestation of these differences is the contribution of whole-word units as contrasted with morpheme units and the concomitant role of rules to describe if and how morpheme units combine to form whole words.

Most well accepted are the *lexicon-based models* that assume the explicit representation of morphemes and of a word's morphological structure. Historically, dual-mechanism accounts of lexical representation have been particularly influential, especially in the domain of morphological processing. Their advocates often describe morphological knowledge in terms of a default option that provides for a rule-governed computation of complex linguistic forms from symbols, accompanied by storage of whole-word forms arrayed in lexical space when rules cannot succeed. The most distinct alternative are the *learning-based models.* These do not assume the explicit representation of morphemes, and describe “morphological effects” in terms of patterns of form and meaning that are processed conjointly. Nevertheless, even models within the general learning-based framework differ to a considerable extent in how they define and implement non-symbolic input/output representations. On the one hand, the major parallel-distributed processing models tend to covertly introduce item-specific lexical events, when they are trained on words presented in isolation and one at a time. This, for example, applies to the triangle model (Harm & Seidenberg, 2004). On the other hand, models such as the Naive Discrimination Reader (NDR; Baayen, Milin, Filipović Đurđević, Hendrix, & Marelli, 2011), build from sublexical form units, such as the letter bigrams and/or trigrams that occur in a particular language. As for the outcomes, the model introduces a hypothetical construct of *lexomes*—word-like units that represent neither word forms nor word meanings, but contribute to meaning in relation with other lexomes in the context (c.f., Milin *et al.*, 2017).

Broadly speaking, most of the learning-based models implement a single mechanism, characterized by graded effects whose underlying activation dynamics are based on mapping input to output similarities. In these frameworks, although both prime-target pairs are matched on form similarity, the pair FELL-FALL appears to be more similar in terms of the contexts in which the two appear than is BELL-BALL. The consequence is a greater difference in reaction times relative to a control word paired with the same target. In the same sense, decision latencies to FELL-FALL differ more from FULL-FALL than do BELL-BALL from BULL-BALL in this task.

Lexicon-based accounts focus on the role of rules and representations, and proponents reason that because formation of the inflected form FELL does not follow from a simple rule applied to FALL, it must be represented and greater facilitation for pairs such as FELL-FALL than for BELL-BALL must reflect something about how the lexical representations for the two words in a pair reference each other (Crepaldi, Rastle, Coltheart, and Nickels, 2010). Both Parallel-distributed processing and Naive Learning accounts, focus on what makes learning easy or hard and would thus quantify the processing cost derived from the systematicity of the mapping between form and meaning when trying to differentiate among words whose forms partially overlap with FALL as compared with BALL. Between the extremes, models vary along a continuum from purely rule-based, to fully probabilistic and inferential.

Dual-mechanism accounts

The dual-mechanism class of lexical models for the recognition (or production) of words posits two independent mechanisms associated with different brain areas (Marslen-Wilson & Tyler, 1998; Pinker, 1999; Pinker & Ullman 2002, 2003; Silva & Clahsen 2008; Ullman, 2001, 2004) where the choice of mechanism depends on how adequately rules can describe word formation).[3] In the dual mechanism framework, differences between languages are captured in terms of different types of rules and/or the symbolic representations over which they apply. If morphological formations in a language follow a general pattern, so that word forms are compositional and can be described by rules, then word specific morphological structure need not be stored with each lexical entry. Past tense formation is the traditional domain of investigation where forms such as WALKED and HUMMED can be described by a rule that operates on a stem (WALK, HUM) and affix(es) such as ED, the inflection for past tense. However, for words that are irregular in that they cannot be formed or decomposed by rule, the incorrect application of the rule would produce the overgeneralized forms SPEAKED and RUNNED. This is the justification for a second non-combinatorial mechanism based on associations among full word representations; its function is to store the exceptions to the rules like SPEAK-SPOKE and RUN-RAN.

This type of dual-mechanism model further assumes different processing mechanism for morphological versus semantic processing. By dual-mechanism accounts, morphological facilitation from a regularly formed prime to its morphologically related target arises when the prime is decomposed into stem and affix and the stem of the prime preactivates the target. This results in faster recognition (facilitation) when it is time to recognize that target than when there is no preactivation. Thus ARTIST is decomposed into ART (stem) + IST (affix) and activation of the stem of the prime benefits the target ART. Similarly WALKED is decomposed into the stem WALK and the affix ED and its stem preactivates the target WALK. In this dual-mechanism framework, the mechanism that produces (regular) morphological facilitation entails decomposition and it differs from that for irregular inflection such as RUN-RAN or semantically related words such as CRAFT-ART where there is no shared stem to preactivate. In those cases, activation spreads from the whole prime word to the whole word target.

One challenge to the dual-mechanism account comes from the "word frequency effect", a comparison of words that vary on how frequently they occur: more frequent words are faster to recognize and faster to produce than less frequent words. A classical dual-mechanism interpretation of the whole word frequency effect in tasks such as lexical decision emphasizes access or activation of forms that are stored in the lexicon. According to the original dual-mechanism model whereby regular inflections are stored in terms of their stem while irregulars are stored in the mental lexicon as full words, the difference between high and low frequency words should be larger for irregularly than for regularly inflected forms in recognition tasks such as lexical decision (Alegre & Gordon, 1999) and production (Budd, Paulmann, Barry, & Clahsen, 2013). Frequency effects for regularly inflected forms thus pose a challenge to the dual-mechanism model (e.g., Baayen, Wurm, & Aycock, 2007).

A further test of the dual-mechanism account comes from nonword priming, where nonwords are formed from an illegal combination of existing stems and affixes like TAUGHTEN or SONGED. Dual-mechanism accounts assume that irregular verb forms are stored as whole word units; hence then these irregular stems (participle stems) should not be represented in the lexicon (Clahsen, Prüfert, Eisenbeiss, & Cholin, 2002). Facilitation from nonwords with irregular stems like

GE + SUNG + T or GE + WURF + T in German (analogs in English include SONG + ED or THREW + N) indicate that irregular and semi-regular stems can function similarly to regular verb stems and thus seriously challenge some assumptions of the dual-mechanism models (Smolka, Zwitserlood, & Rösler, 2007).

Single mechanism accounts

Single mechanism learning models of morphological processing differ from dual-mechanism accounts in that they posit just one mechanism and that mechanism is sensitive to the frequency and sequential patterning of units in everyday language. Therefore, the foundation of Single Mechanism Accounts is the statistical structure that is present in language rather than distinct mechanism(s) that operate on the symbolic representations for words or the rules that operate on them. The basis of this framework is a serious treatment of the systematic mapping between form and meaning that characterizes the many words that share a base morpheme (Bybee, 1985, 1995; Bybee & McClelland, 2005). For example the form-meaning mapping is stronger in (SALT, SALTY, SALTINE, SALTSHAKER) than in (CALM, CALMNESS, CALMLY) because of the number of different words formed from the stem "morphological family size" (De Jong, Schreuder, & Baayen, 2000).

The most familiar option in the single mechanism research framework is parallel-distributed models (PDP: Gonnerman, Seidenberg, & Andersen, 2007; Kielar, Joanisse, & Hare, 2008; Joanisse & Seidenberg, 1999; Plaut & Gonnerman, 2000; Rueckl & Raveh, 1999; Seidenberg & Gonnerman, 2000, etc.). These models are also known as the connectionist models. The framework permits activation from the systematic mappings between form and meaning to vary in degree and to converge for noncompositional irregulars like RAN and SPOKE and FELL as well as for compositional regulars like WALKED and HUMMED. Stated alternatively, in a single mechanism framework, non-compositional irregulars and compositional regulars vary in degree not in type of morphological processing mechanism. Indeed, electrophysiological brain responses to regular (default), semi-irregular, and completely irregular participles in German have been observed to vary in degree and not in an all-or-none fashion (Smolka, Khader, Wiese, Zwitserlood, & Rösler, 2013).

At their core, PDP accounts are non-symbolic in nature and thus do not refer to stored representations or rules. Rather the organization emerges from distributed patterns of connectivity so that similarity patterns that encompass semantics, orthography and phonology contribute in a graded manner to the recognition and the production of all inflected forms. In the system, contributions of (constraints related to) task and more permanent differences between words are captured in a graded manner. Differences in facilitation in a priming study reflect the underlying dynamics of a system along with its earlier resting state or initial conditions (Rueckl, 2002). There is no shifting between mechanisms. For example, the point of departure for a PDP account of the apparent effect of inflectional regularity would focus on the consequences of greater orthographic overlap between prime and target for regularly than for irregularly inflected verbs (e.g., Bird *et al.*, 2003; Bybee

& McClelland, 2005; Patterson, Lambon Ralph, Hodges, & McClelland, 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996). Also relevant is greater semantic connectedness among irregularly than regularly inflected verbs (devoid of any prime) as this influences the starting point for recognition (Baayen & Moscoso del Prado Martin, 2005).

The finding that in native speakers irregularly inflected verb forms with high form overlap (e.g., DRAWN-DRAW) show facilitation that is comparable to regular verb forms (e.g., GUIDED-GUIDE) and not to change stem irregulars (e.g., FELL-FALL; in Basnight-Brown, Chen, Shu, Kostić, & Feldman, 2007) is consistent with single mechanism accounts based on convergent activation from form and meaning.

Information theoretic approach

Information theory is the basis for another approach to morphological organization and processing. The essence is that information can be quantified, expressed in terms of an amount of information, that can serve as an alternative account of lexical knowledge, one that does not depend on a structure of representations arrayed by similarity in lexical space. A key measure is *entropy*: the average number of bits to communicate a message. The measure can be understood as quantifier of the uncertainty in predicting an outcome from a set of possible outcomes, which may or may not be equiprobable. In Shannon's (1948) own words, the number of discrete states of a system along with the way these states are organized determines the amount of information in that system. The cost of reducing the uncertainty is predictive of response latency in tasks such as lexical decision. Thus, in the information-theoretic approach, uncertainty and/or processing cost rather than a structure provide the organizing framework for lexical knowledge.

In the information theory framework, high information is characteristic of improbable events and low information is characteristic of probable events. Additionally, all events that may occur can be represented jointly by a probability distribution. On one extreme, as we said, this distribution can be equiprobable, representing events that have the same (equal) probability of occurrence. In that case, uncertainty is at its highest—with maximum entropy. On the other extreme, one event can have a probability of 1.0 and all others of 0.0. In that case, there is no uncertainty—entropy is zero—we know what will occur.

Stated crudely, the probability of a word appearing among the set of possible words can be described in terms of uncertainty, and uncertainty is correlated with the processing cost that is typically measured with reaction time. For example, a particular inflected form appears on an experimental trial (WORKING) and not one of its other related forms (WORKED), and this uncertainty can be quantified with information-theoretic measures, like the amount of information (c.f., Kostić, Marković, & Baucal, 2003) or the entropy (c.f., Moscoso del Prado Martín, Kostić, and Baayen, 2004), which are predictive of decision latencies (for an extensive overview see Milin, Kuperman, Kostić, & Baayen, 2009).

Processing rate per information unit increases as the amount of information becomes higher, which has been documented for words presented both in isolation (no context) and in contexts with various experimental manipulations (Kostić, 1991; 1995; Kostić, Marković, & Baucal, 2003). Similarly, as the probabilities of events get more equal, uncertainty gets higher and consequently processing time gets longer (Milin, Filipović Đurđević, Moscoso del Prado Martín, 2009; Milin *et al.*, 2009; Moscoso del Prado Martín, Kostić, & Baayen, 2004).

The probability of an inflected variant of a word (like WORKING), in the information-theoretic framework is based on the frequency of that specific form when compared to the sum of the frequencies of all related forms (e.g., WORK, WORKS, WORKED, WORKING). This sum serves as a normalizing term. The probability distribution of the inflected variants of a particular word's inflected forms may differ from the probability distribution of its inflectional class in general. An analogy from English would be to compare the probability distribution of the inflected variants of a particular word (e.g., WORK: WORK; WORKS; WORKED; WORKING), with the probability distribution of all word endings—inflectional suffixes or exponents (-Ø; -S; -ED; -ING). The extent to which the general and word specific probability distributions differ is quantified by relative entropy, often called Kullback-Leibler divergence. This quantity is predictive of response latencies in lexical decision and other word recognition tasks: the greater the divergence, the longer the reaction time (Milin, Filipović Đurđević, & Moscoso del Prado Martín, 2009; Baayen, Milin, Filipović Đurđević, Hendrix, & Marelli, 2011).

In the information-theoretic framework, differences between regular and irregular verb forms get tied to properties of the words themselves, including their inflectional entropy—the frequency distribution of inflected forms including both regulars and irregulars. At the same time, it includes properties that pertain to semantics such as a word's imageability, number of senses in WordNet (Miller, 1995), contextual diversity compared to other words (Baayen & Moscoso del Prado Martín, 2005; but see also McDonald & Shillcock, 2001 and Adelman, Brown, & Quesada, 2006). In sum, within the information-theoretic framework, when regulars and irregulars incur differences in processing, the explanation is that their statistical properties differ, not that they are assigned different types of representations or different processing mechanisms from the outset. For example, the irregular past tense BUILT will be processed differently from the regular past tense HOUSED, because it differs in its uncertainty in context, that is, with respect to its frequency of occurrence and co-occurrence, distributional semantics and so on (see further discussions in Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014; Baayen, Milin, & Ramscar, 2016).

Naive discriminative learning

Learning represents a counterpart to information processing, as it focuses on the *costs* of "inserting" new items into our memories. It increases the overall uncertainty, as it increases the elements in the system. It pays off, however, by increasing our capabilities to make ever-finer discriminations in our environment, which is the essence of adaptation (for general discussion related to learning and adaptation see Hinton & Nowlan, 1987). Similarly, language can be defined as a complex adaptive (sub)system (Beckner *et al.*, 2009), whose principal function is to facilitate ways in which we interact with our social environment (see Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). Thus, learning more language entails greater discrimination abilities in verbal communication.

The Naive Discriminative Learning (NDL; Baayen Milin, Đurđević, Hendrix, & Marelli, 2011; Baayen, Shaoul, Willits, & Ramscar, 2015; Milin, Ramscar, Baayen, & Feldman, 2015) framework provides an account of morphological processing that, inspired by Word and Paradigm Morphology (Matthews, 1991; Blevins, 2006), eschews the theoretical constructs of stems, morphemes, and affixes as units of form.

For language comprehension, this framework proposes a pair of two-layer networks, with connections between inputs (henceforth cues) and outputs (henceforth outcomes), which are obtained by applying the Rescorla-Wagner equations (Rescorla & Wagner, 1972) to time series of learning events—points in time at which weights between cues and outcomes are updated. The first network of the pair has letter n-grams or n-phones (typically, n is 2 or 3) as cues, and lexomes as outcomes.

The concept of the *lexome* is best explained by analogy to atoms in chemistry. Atoms have two important properties. First, from the perspective of physics, they are not indivisible, yet they suffice for understanding the chemical properties of molecules. Second, the chemical properties of molecules are specific to the molecules, and cannot be derived from the properties of the atoms. Like atoms, then, lexomes are theoretical constructs that have no meaning of their own, but instead their meaning is relational in that it emerges dynamically from the other lexomes with which they collocate. In the spirit of Landauer and Dumais (1997), semantic similarity between lexomes is approximated by the cosine of the angles between these lexomes' weight vectors. In the spirit of adaptive systems it is dynamical and evolves with learning.

In the NDL framework, the only form representations are those of letter bi/trigrams (or bi/triphones). There are no representations for stems, words, or affixes. Furthermore, no distinction is made between representations for derived, inflected or compounds words. The pivotal unit in the NDL approach is the *lexome*. Lexomes approximate experiences that are discriminated within a speech community, including not only experiences of different objects and actions, but also more "linguistic" experiences such as aspect, number, and tense. Morphological effects emerge in the first network (with letter or phone trigrams as cues and lexomes as outcomes), as a consequence of the co-occurrence statistics of these trigram or triphone cues and the lexomes (Baayen, Milin, Đurđević, Hendrix, & Marelli, 2011; Baayen, Shaoul, Willits, & Ramscar, 2015; Milin *et al.*, 2017). For example, differences in visual word recognition tasks between primes for PAST (like PASTOR and PASTA), reflect the extent to which sublexical letter trigrams are associated with the target lexome PAST. No decisions pertaining to whether trigrams such as STO, TOR, OR#, or STA and TA# (where # represents a terminal marker), are, or are not, affixes are invoked. No morphological decomposition is invoked whatsoever. Importantly, effects of form similarity of both embedded or embedding words (e.g., Bowers, Davis, & Hanley, 2005) as well as of orthographic neighbors (e.g., Baayen, Feldman, & Schreuder, 2006; Davis & Lupker, 2006; Forster & Taft, 1994; Kinoshita, Castles, & Davis, 2009) occur during learning, not from decomposition at decision time.

Learning is driven by both positive and negative evidence. As a cue occurs more often in contexts where it does not pertain to a given target lexome, the connection strength from this cue to this target lexome will be reduced. As a crucial consequence, measures based on discrimination learning go beyond frequency counts. Whereas frequency counts capture only the co-occurrence frequencies of a cue and a lexome, discrimination measures take into account how often a cue is "unfaithful' to this lexome (i.e., when it supports any other lexome; see more in Ramscar & Yarlett, 2007; Ramscar, Yarlett, Denny, & Thorpe, 2010).

The activation of a lexome is an index of how well it is discriminated from other lexomes, and typically correlates with its frequency of occurrence. A lexome's prior availability represents its degree of entrenchment in the network, and is also correlated with frequency of occurrence. But whereas the activation provides a frequency measure for bottom-up support, the prior availability is a measure of top-down expectation that is independent of the input.

Rescorla-Wagner networks specify the computational engine of the NDL approach, and are only part of a larger conceptual framework (for the specification of a discrimination-based model of auditory comprehension, see Baayen *et al.*, 2015; Hendrix, 2015, shows how NDL activation and NDL prior availability differentiate between the bottom-up and top-down processes guiding eye movements in reading compounds; Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017 address "morpho-orthographic" segmentation in reading).

Lexical access: Morphological and form effects in early visual word recognition[4]

Complex words like UNSUCCESSFULLY include several morphemic constituents, UN, SUCCESS, FUL, LY, that recur in the language in many other words. Since the seventies (Murrell & Morton, 1974; Taft & Forster, 1975) researchers have asserted that morphemic structure affects the recognition of complex words. The much-debated question—how complex words are accessed and represented in lexical memory—is closely related to the definition of a morpheme and to the possibility that morphological structure influences processing in the absence of semantics. Morphemes are units of form and semantics, as defined in traditional linguistic theory. Many interpret morphological structure devoid of semantics (comparable outcomes for semantically transparent SUCCESSFUL-SUCCESS and semantically opaque SUCCESSOR-SUCCESS) as evidence of an early prelexical (prior to access to knowledge stored in the lexicon) process. Experimental methods and measures define what is early and what is late, but whether the absence of semantically informed morphological structure necessarily implicates an earlier form stage that is independent from a later semantic process stage is more contested. We review some of the evidence for early morphological processing with and without a semantic contribution and evidence for form processing with and without a morphological contribution in the remainder of this section.

Semantic contributions to morphological processing: Early or late?

Most variants of the lexicon-based account of morphological processing posit two-stages: Orthographically based but semantically blind morphological decomposition that occurs early during visual word recognition and a semantic interpretation of the decomposed constituents that occurs at a later, lexical stage: The early decomposition process is based on the form of the morpheme without regard to how that unit maps onto the "meaning" of the word in which it appears (hence "form-then-meaning"; e.g., Lavric, Rastle, & Clapp, 2011; Meunier & Longtin, 2007; Rastle, Davis, & New, 2004; Rastle & Davis, 2008). Because primes like UNCOVER and RECOVER are morphologically well structured and thus decomposable and because semantics plays no role until later in the sequence of processing stages, the rational is that they should facilitate the recognition of targets like COVER comparably regardless of whether they share both meaning and form with it, or form but little meaning.

Indeed, under masked priming conditions in many behavioral studies in English and French, different types of morphologically related primes and targets produce facilitation but have failed to show an effect of meaning (Longtin, Segui, & Hallé, 2003; Rastle *et al.*, 2000; Rastle, Davis, & New, 2004). Conversely, faster decision latencies to targets like SUCCESS after forward masked SUCCESSFUL than after SUCCESSOR provide evidence of an early role for meaning that might be characterized as semantically informed decomposition (hence "form-with-meaning"; see Feldman, Milin, Cho, Moscoso del Prado Martín, & O'Connor, 2014; Feldman, Smolka, Cho, & Milin, 2014). Similar effects arise with prefixed words like UNDRESS-DRESS and REDRESS-DRESS (Feldman, Smolka, Cho, & Milin, 2014). One difference between those studies that report an early effect of semantic transparency and those that fail to show the effect is that only the former tend to use a within target design so that the same target appears with a transparent, an opaque and an unrelated prime (for in-depth discussion about differences in the two experimental designs also consult Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017).

Analogous to the behavioral measures, the above models have been tested with EEG measures using masked visual primes presented at short durations (below 50 ms SOA) and the lexical decision task (for a review of morphological EEG effects see Smolka, Gondan, & Rösler, 2015). Results with EEG provided strong N250 and/or N400 attenuations for letter sequences that are semantically related and exhaustively decomposable into morphemes (e.g., word pairs like FARMER-FARM), but results are inconsistent with respect to pairs that are semantically unrelated and differ with respect to exhaustive or partial decomposability into morphemic constituents (e.g., pairs like CORNER-CORN and BROTHEL-BROTH respectively where ER is an affix but EL is not.).

Most studies found no difference in facilitation induced by morphologically related primes and either exhaustively (e.g., + ER) or only partially (e.g., + EL) decomposable primes in either the N250 or the N400 latency range (Morris *et al.*, 2008, 2011, 2013), while one study found more priming by the former than by the latter two types (Morris *et al.*, 2007). Yet other studies revealed similar processing at an early (N250) processing stage combined with a differentiation at a later (N400) processing stage. The inconsistent patterning of words with exhaustively decomposable word primes continues to fuel the discussion about the model—"form-then-meaning" (e.g., Lavric *et al.*, 2007, 2011; Morris *et al.*, 2011), or "form-with meaning" where even early in processing true morphologically related pairs benefit from shared semantics in a way in which pairs that share only form cannot (e.g., Diependaele *et al.*, 2005; Feldman *et al.*, 2009; Morris *et al.*, 2008, 2011, 2013; Holcomb & Grainger, 2006).

Reports that semantic transparency of the prime reliably influences early morphological processing is incompatible with variants of the lexicon-based tradition where form is independently analyzed before meaning can influence recognition. The "form-with-meaning" view is based on conjoint effects of form and meaning, from the onset of the morphological processing (Baayen *et al.*, 2011; Feldman *et al.*, 2009, 2014; Feldman, Kostić, Gvozdenović, O'Connor, & Moscoso del Prado Martín, 2012). Finally, with respect to learning-based models more generally, only NDL anticipates that the effect of semantic transparency of the prime on morphological facilitation depends on the similarity of the target to the other words that constitute its form neighbors (Feldman *et al.*, 2014).

Morphological contributions to early orthographic processing

Different patterns of facilitation for form similar pairs with a fully decomposable or exhaustive morphological structure like CORNER-CORN and for pairs with only a partially decomposable structure like BROTHEL-BROTH are crucial to the first stage of form-then-meaning lexical models. Neighborhood density or number of neighbors is a measure of form similarity. Analogs exist in the auditory as well as the visual domain (see Pisoni, this volume). Neighbors can be formed by letter substitution (Coltheart, Davelaar, Jonasson, & Besner, 1977), letter deletion or letter addition (for overviews, see Perea, Acha, & Fraga, 2008; Davis & Taft, 2005). A word's similarity to many other words as indexed by its orthographic neighborhood density increases single word recognition latencies when other measures are controlled (Baayen, Feldman & Schreuder, 2006; Yarkoni, Balota, & Yap, 2008). Similarly, when targets are preceded by masked primes that are neighbors, word recognition latencies get slower as target neighborhood density increases (Forster, Davis, Schoknecht, & Carter, 1987; Forster & Davis, 1991). Finally, targets from sparse orthographic neighborhoods tend to show stronger facilitation after orthographically similar primes than do targets from dense neighborhoods (Forster & Taft, 1994; Kinoshita, Castles, & Davis, 2009; Perry, Lupker & Davis, 2008).

We have recently documented that orthographic neighborhood density can systematically influence the magnitude of facilitation in morphological studies (Feldman, Smolka, Cho, & Milin, 2014; Milin *et al.*, 2017). For example, a word like

FORM with many neighbors (e.g., FORK, FOAM, FIRM, DORM) will tend to show smaller differences between masked neighbor primes (e.g., DEFORM, PERFORM) than a word with fewer neighbors like DRESS (e.g., PRESS, CRESS). Thus primes for FORM will differ less than primes for DRESS (e.g., UNDRESS, REDRESS).

These results are anticipated by NDL, using direct discriminative mappings from letter trigrams as input cues to lexome as outcomes. Contrariwise, the same results require theoretically unmotivated adjustments in lexicon-based accounts. Morphological facilitation after exhaustively decomposable forms (e.g., FARMER-FARM; CORNER-CORN; PASTOR-PAST) but not after partially decomposable forms (e.g., BROTHEL-BROTH; PASTA-PAST; LIMBO-LIMB) is central to the claim that early processing is based on semantically blind but, nonetheless morphological units, that is, not based only on form overlap (e.g., Diependaele, Sandra, & Grainger, 2005; Longtin, Segui, & Hallé, 2003; Marslen-Wilson, Bozic, & Randall, 2008; McCormick, Rastle & Davis, 2009; Rastle, Davis, Marslen-Wilson, & Tyler, 2000; Rastle, Davis, & New, 2004). However in those studies, different targets appeared with semantically unrelated exhaustively and partially decomposable primes and form similarity was only crudely matched. With more rigorous controls, that finding is not always replicable (see Milin *et al.*, 2017; also see Duñabeitia, Kinoshita, Carreiras, & Norris, 2011). The implication is that both capture form-based facilitation.

In summary, evidence is emerging that the effect of semantic transparency and of morphological structure of the prime on morphological facilitation can be linked to neighborhood properties of the target. As currently described, even lexicon-based accounts of morphological facilitation fail to anticipate the interaction of semantic transparency with target neighborhood density. This interaction not only is compatible but also is anticipated by an NDL account. However, as we explained above, NDL takes a different view on the questions of what consists a unit, how (and why) it emerges, and what its properties are. Crucially, it is not consistent with a compositional account and, thus, it avoids postulating constructs such as morphemes, stems and affixes as its units (following "Word and Paradigm Morphology": Matthews, 1991; Blevins, 2006). Accordingly, it also does not hypothesize "meaning" residing in those units. Instead meaning arises from the collocation dynamics between lexomes.

Explaining cross-language differences in morphological processing

Effects of orthographic and morphological similarity that vary with visual neighborhood density of the target may help to explain discrepant priming outcomes across languages. For example, studies in Hebrew and Arabic replicated robust morphological priming by masked semantically transparent and opaque primes across long and short prime durations in the lexical decision task (Boudelaa & Marslen-Wilson, 2005; Deutsch, Frost, & Forster, 1998; Frost *et al.*, 1997). In addition, both form and morphological facilitation are easier to document in English (viz., for targets from sparse neighborhoods) than in Hebrew (Frost *et al.*, 2005). In fact the absence of form priming in Hebrew is sometimes interpreted as evidence for language-specific processing (Frost, Deutsch, & Forster, 2000). However, Hebrew roots are only three or four consonants in length and most vowels are not written in text for adult readers. The implication is that when neighborhood density is defined orthographically, Hebrew words will tend to be shorter than words in English or French and, generally, across languages including Hebrew, shorter words tend to have more neighbors.

Lexical representation: Morphological effects in late visual word recognition

In contrast to masked priming, unmasked or overt priming taps into a later stage of lexical processing and representation. Under auditory priming or visual priming at long SOAs, in languages like English and French, stems were primed by suffixed derivations if they were semantically transparent, as in SUCCESSFUL-SUCCESS. By contrast, facilitation was absent for semantically opaque derivations like SUCCESSOR-SUCCESS (for cross-modal priming see Longtin, Segui, & Hallé, 2003; Marslen-Wilson, Tyler, Waksler, & Older, 1994; for visual priming at long SOAs, see Feldman & Soltano, 1999; Feldman, Soltano, Pastizzo, & Francis, 2004; Rastle, Davis, Marslen-Wilson, & Tyler, 2000; Feldman & Larabee, 2001; Meunier & Segui, 2002). Lexicon-based models posit different processing mechanisms for semantically transparent and opaque words at the lexical level. Semantic information can be integrated (in the two-stage model, e.g., Lavric *et al.*, 2011), shared representations can operate at the morpho-semantic level (in the dual-route model, e.g., Morris *et al.*, 2013), or form-to-meaning mappings can be realized (as in a connectionist network (e.g., Plaut & Gonnerman, 2000). EEG findings show a similar pattern. Attenuations to the N400 were largest when they were induced by true morphologically related words like FARMER-FARM, slightly reduced for pairs like CORNER-CORN and smallest for form pairs like BROTHEL-BROTH (Lavric *et al.*, 2011). Results are traditionally interpreted as consistent with the model of visual word recognition with an orthographically based morphological decomposition followed by a later semantically informed stage.

Similarly, prefixed words in English and Serbian facilitated lexical latencies, when they were semantically transparent as in DISOBEY-OBEY, or PRIVOLE-VOLIM. Characteristic of this task is the absence of facilitation when the derivations were semantically opaque as in RESTRAIN-STRAIN or ZAVOLE-VOLIM (Marslen-Wilson *et al.*, 1994; Feldman, Barac-Cikoja, & Kostic, 2002; Feldman & Larabee, 2001). Similarly with sentence primes, semantically transparent or ambiguous Dutch prefixed verbs (that possess a transparent and an opaque reading) showed facilitation, whereas truly opaque prefixed verbs did not (Zwitserlood, Bolwiender, & Drews, 2005).

The above findings are traditionally explained by assuming that the lexical representations of complex words that underlie performance when primes are overt depend on semantic compositionality: Only semantically transparent words can benefit recognition of their stem. By contrast, semantically opaque words like

SUCCESSOR are not represented as derived. Rather, they are represented as unanalyzed and independent lexical items that do not share a stem with words like SUCCESS or SUCCESSFUL (Marslen-Wilson *et al.*, 1994). In lexicon-based models, morphological decomposition as revealed by overt priming is constrained by semantic knowledge that influences the interrelation among lexical entries (Diependaele, Sandra, & Grainger, 2005; Marslen-Wilson, Bozic, & Randall, 2008; Meunier & Longtin, 2007; Rastle *et al.*, 2000; 2004; Taft & Kougious, 2004; Taft & Nguyen-Hoan, 2010).

In contrast, connectionist learning-based models posit a continuity between early and late tasks and emphasize graded effects of form and meaning, in behavioral (Gonnerman, Seidenberg, & Andersen, 2007) and in EEG (Kielar & Joanisse, 2011) studies. For example, stronger N400 priming effects for semantically transparent word pairs like GOVERNMENT-GOVERN than for less transparent word pairs like DRESSER-DRESS along with no facilitation for semantically opaque pairs like APARTMENT-APART or CORNER-CORN capture graded semantic similarity when form similarity is held constant.

The origin of cross-language differences in late morphological processing

In German and in Hebrew effects of semantic transparency on morphological facilitation have not been detected regardless of processing time for the prime. Studies on prefixed verbs in German have found equivalent morphological facilitation after semantically opaque (VERSTEHEN-STEHEN, "understand-stand") and transparent (AUFSTEHEN-STEHEN, "stand up-stand") verbs with both auditory and visual presentations, even at long (300 ms and 1000 ms) SOAs (Smolka, Komlósi, & Rösler, 2009; Smolka, Preller, & Eulitz, 2014). Together with the absence of facilitation for form controls like VERKLEIDEN-LEIDEN ("disguise-suffer") these morphological effects cannot be attributed to form similarity. Similar to the behavioral findings, N400 attenuations are equivalent for semantically transparent (ANKOMMEN-KOMMEN, "arrive-come") and opaque (UMKOMMEN-KOMMEN, "perish-come") prefixed verbs in German (Smolka, Gondan, & Rösler, 2015). Moreover, these morphological effects were stronger than either pure semantic or pure form effects.

Approximating the findings in German, findings in Hebrew and Arabic also showed morphological effects that failed to vary with semantic transparency. A long term priming study in Hebrew (Bentin & Feldman, 1990) demonstrated that complex words in Hebrew are represented in the lexicon in terms of their root (e.g., GDL). The recognition of a target like miGDaL ("tower") was primed when it was preceded by morphologically related words (same GDL root) both when they were semantically related like GaDoL ("big") or semantically unrelated like GiDuL ("tumor"). Further studies in Hebrew and Arabic replicated robust and equivalent facilitation after both semantically transparent and opaque derivations (Boudelaa & Marslen-Wilson, 2004a, 2004b, 2005; Frost, Deutsch, Gilboa, Tannenbaum, & Marslen-Wilson, 2000). In a traditional framework, the findings in

German, Hebrew and Arabic suggest that lexical representations capture morphological, specifically root structure regardless of meaning compositionality. To elaborate, whereas the lexical representation of a complex verb like UNDERSTAND refers to its base STAND (Smolka *et al.*, 2009; Smolka, Preller, & Eulitz, 2014; Smolka, Gondan, & Rösler, 2015), that of a Hebrew or Arabic word like GiDuL is with reference to its root GDL (Frost *et al.*, 2000; Boudelaa & Marslen-Wilson, 2005). Strong morphological effects without an effect of semantic transparency appear difficult to reconcile with connectionist or other accounts that depend on the general convergence of codes. Some offer these findings in support of rules and representations that *a priori* differ across languages. Underrepresented are attempts to track the role of general statistical properties that characterize words and their similarity to other words and how these tendencies might differ across languages.

In summary, different portraits of morphological facilitation across languages—restricted by semantic transparency in English, French, Dutch, and Serbian but independent of semantic compositionality in German, Hebrew, and Arabic—highlight the importance of cross-linguistic research. Hebrew and Arabic are morphologically rich languages (e.g., Ravid, 2012) and German represents the morphologically richest language within the Indo-Germanic language family, given that it has retained morphological markers to indicate grammatical functions (De Vogelaer, 2007; Roelcke, 1997). Therefore, one possibility is that the morphological richness of a language influences the representation of its morphological structure. However, within a connectionist framework, a greater effect of semantic transparency in morphologically impoverished than in a morphologically rich languages has been linked to the strength of form-meaning regularities (consult Plaut & Gonnerman, 2000; Raveh & Rueckl, 2000; Rueckl *et al.*, 1997; also, see Mirković, Seidenberg, & Joanisse, 2011 for connectionist model of Serbian noun paradigms). Hence, in a system with abundant form-meaning regularities, these patterns may guide visual word recognition. In a network simulation of a morphologically impoverished environment (a language like English), in which mappings between orthographic surface forms and their meanings are mostly idiosyncratic, morphological regularities played a minor role (Plaut & Gonnerman, 2000). In contrast, in the simulation of a morphologically rich environment (a language like Hebrew) with its dense overlapping mappings between many orthographic forms and meanings, form based morphological regularities dominated processing and simulated morphological priming effects were independent of semantic relatedness. Of course, even in morphologically rich environments, semantically transparent word forms should yield some advantage over semantically opaque ones, but this has not been demonstrated for prefixed words in German.

Within a language, evidence is accruing that morphological effects such as semantic transparency are likely to be more important for some types of words than for others and this may help us understand where and when differences emerge across languages. Semantic transparency can be operationalized in terms from distributional semantics such as the cosine similarity between a stem and a derived-form vector (Marelli & Baroni, 2015). We recently demonstrated in English that the influence of an effect of semantic similarity between morphologically-related prime and target is weaker when primes have many semantically similar words, with a dense semantic neighborhood, than when they have fewer semantically similar words (Feldman, Marelli, Amenta, & Milin, 2015). Whether differences across languages can be linked reliably to general properties like neighborhood density or mappings between form and semantics (Marelli, Amenta, & Crepaldi, 2015; Amenta, Marelli, & Sulpizio, 2016) or whether language-specific properties must be invoked to account for some of the variation in the patterns of morphological facilitation awaits further research.

Summary: Lexical Access and Morphological Processing

Lexicon-based and learning-based models ask different questions about and offer different solutions to challenges in morphological processing. From a general combinatorial position, when words are composed of multiple morphemes, they range in semantic compositionality from semantically transparent, as in SUCCESS-FUL to "opaque," as in SUCCESS-OR and the meaning of semantically opaque words cannot be derived from linguistic rules that combine their meaningful components. Rules that operate on symbols, with storage of more opaque full word forms as a backup, is one popular way to characterize morphological knowledge. Insofar as rules are language specific, some describe differences between languages in terms of differences between rules.

Learning-based approaches look for universal, albeit more complex patterns that can vary with statistical properties of words. Here, we have used tools from distributional semantics and the lexical decision task in English to demonstrate that by differentiating among words that are semantically similar to many versus few other words we can better predict effects of semantic transparency in lexical processing. A better grasp of the variability among words within a language with respect to properties like orthographic and semantic neighborhood size, and how that variation differs across languages, may prove crucial to determine whether or not rules are necessary to capture knowledge about the patterns that form complex words, and the extent to which the underlying process is specific to a language.

Specifically within the NDL framework, meaning loses its encapsulation and becomes relational and contextualized—dependent on the communicative intentions between interlocutors. Form units are also reduced to naive (i.e., theoretically empty) sublexical n-grams of letters or phones, avoiding the burden of representing traditional morphemes and stems and affixes. Within this framework, in particular, there is no simple and direct mapping between form units and meaning units, but rather two interdependent networks, where morphological effects emerge from the network with letter or phone bi/trigrams as cues and lexomes as outcomes, and where word semantics emerge from a second network that takes neighboring lexomes as cues and target lexomes as outcomes, and the "meaning" of a lexome is best approximated by its relational behavior (vector of outgoing weights) with other lexomes.

Author Note

This research was supported in part by the Alexander von Humboldt Foundation to Petar Milin through an Alexander von Humboldt research chair awarded to R. H. Baayen, by the National Institute of Child Health and Human Development HD 01994 to Haskins Laboratories, and by the VolkswagenStiftung, Grant FP 561/11 to Eva Smolka.

NOTES

1 Traditionally, a morpheme is the minimal unit of meaning and morphology refers to the study of word structure for those units. As will become evident (see section 3.4), "meaning" is a term that is compatible with traditional lexical accounts but less so with current learning accounts.
2 This is, of course, conditioned on yet another implicit assumption that morphemes and words are independent "meaning carriers."
3 The term "mechanism" is used differently in this chapter than in the work of the proponents of dual mechanism accounts noted here. That framework posits two distinct brain mechanisms, rather than two different routes (e.g., parsing and retrieval) within a single mechanism.
4 In the discussion below, we revert to the traditional term "meaning" and attempt to remain agnostic with respect to how it should be represented.

REFERENCES

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17(9), 814–823. doi: 10.1111/j.1467-9280.2006.01787.x

Alegre, M. & Gordon, P. (1999). Rule-Based Versus Associative Processes in Derivational Morphology. *Brain and Language*. 68(1), 347–354. doi: 10.1006/brln.1999.2066

Amenta, S., Marelli, M., & Sulpizio, S. (2016). From sound to meaning: Phonology-to-Semantics mapping in visual word recognition. *Psychonomic Bulletin & Review*, 1–7.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 55(2), 290–313. doi: 10.1016/j.jml.2006.03.008

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3), 438–481. doi: 10.1037/a0023851

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, 30(11), 1174–1220. doi: org/10.1080/02687038.2016.1147767

Baayen, R. H., & Moscoso del Prado Martín, F. (2005). Semantic density and past-tense formation in three Germanic languages. *Language*, 81(3), 666–698. doi: 10.1353/lan.2005.0112

Baayen, R. H., Shaoul, C., Willits, J. and Ramscar, M. (2015). Comprehension without segmentation: A proof of concept with naive discrimination learning. *Language, Cognition and Neuroscience*, 31(1), 106–128.doi: 10.1080/23273798.2015.1065336

Baayen R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across task sand modalities. *The Mental Lexicon*, 2(3), 419–463. doi: 10.1075/ml.2.3.06baa

Basnight-Brown, D., Chen, H., Shu, H., Kostić, A. & Feldman, L. B. (2007). Monolingual and Bilingual Recognition of Regular and Irregular English Verbs: Does Sensitivity to Word Form Vary with Language Experience? *Journal of Memory and Language*, 57(1), 65–80. doi: 10.1016/j.jml.2007.03.001

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language learning*, 59(1), 1–26. doi: 10.1111/j.1467-9922.2009.00533.x

Bentin, S., & Feldman, L. B. (1990). The contribution of morphological and semantic relatedness to repetition priming at short and long lags: Evidence from Hebrew. *The Quarterly Journal of Experimental Psychology*, 42(4), 693–711. doi: 10.1080/14640749008401245

Bird, H., Lambon Ralph, M.A., Seidenberg, M.S., McClelland, J. L., & Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection?. *Journal of Memory and Language*, 48(3), 502–526. doi: 10.1016/s0749-596x(02)00538-7

Blevins, J. P. (2006). Word-based morphology. *Journal of Linguistics*, 42(3), 531–573. doi: 10.1017/S0022226706004191

Bodner, G.E., & Masson, M. E. J. (2003). Beyond spreading activation: An influence of relatedness proportion on masked semantic priming. *Psychonomic Bulletin & Review*, 10(3), 645–652. doi: 10.3758/BF03196527

Boudelaa, S., & Marslen-Wilson, W. (2004a). Abstract morphemes and lexical representation: The CV-Skeleton in Arabic. *Cognition*, 92(3), 271–303. doi: 10.1016/j.cognition.2003.08.003

Boudelaa, S., & Marslen-Wilson, W. (2004b). Allomorphic variation in Arabic: Implications for lexical processing and representation. *Brain and Language*, 90(1), 106–116. doi: 10.1016/S0093-934X(03)00424-3

Boudelaa, S. & Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: Incremental masked priming in Arabic. *Language and Cognitive Processes*, 20(1-2), 207–260. doi: 10.1080/01690960444000106

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Automatic semantic activation of embedded words: Is there a "hat" in "that"?. *Journal of Memory and Language*, 52(1), 131–143. doi: 10.1016/j.jml.2004.09.003

Budd, M.-J., Paulmann, S., Barry, C. & Clahsen, H. (2013). Brain potentials during language production in children and adults: An ERP study of the English past tense. *Brain and Language*, 127(3), 345–355. doi: 10.1016/j.bandl.2012.12.010

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Language Production*, pp. 257–294. London: Academic Press.

Bybee, J. (1985). *Morphology*: A Study of the Relation between Meaning and Form (Vol. 9).John Benjamins Publishing. doi: 10.1075/tsl.9

Bybee, J. (1995). Diachronic and typological properties of morphology and their implications for representation. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (109–129). Hillsdale, NJ. Erlbaum.

Bybee, J. and McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22(2-4), 381–410. doi: 10.1515/tlir.2005.22.2-4.381

Clahsen, H., Prüfert, P., Eisenbeiss, S. & J. Cholin (2002). Strong stems in the German mental lexicon: Evidence from child language acquisition and adult processing. In: I. Kaufmann & B. Stiebels (Eds.), *More than Words: a Festschrift for Dieter Wunderlich*. Akademie-Verlag: Berlin, pp. 91–112.

Coltheart, M., Davelaar, E., Jonasson, J. F., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention & Performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, 32(3), 668. doi: 10.1037/0096-1523.32.3.668

Davis, C., & Taft, M. (2005). More Words in the Neighborhood: Interference in Lexical Decision Due to Deletion Neighbors. *Psychonomic Bulletin & Review*, 12(5), 904–910. doi: 10.3758/BF03196784

De Jong, N., Schreuder, R., and Baayen, R. H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes*, 15(4-5), 329–365. doi: 10.1080/01690960050119625

Deutsch, A., Frost, R., & Forster, K. I. (1998). Verbs and nouns are organized and accessed differently in the mental lexicon: Evidence from Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(5), 1238–1255. doi: 10.1037/0278-7393.24.5.1238

De Vogelaer, G. (2007). Extending Hawkins' comparative typology: Case, word order, and verb agreement in the Germanic languages. *Nordlyd*, 34(1), 167–182. doi: 10.7557/12.93

Diependaele, K., Sandra, D., & Grainger, J. (2005). Masked cross-modal morphological priming: Unravelling morpho-orthographic and morpho-semantic influences in early word recognition. *Language and Cognitive Processes*, 20(1-2), 75–114. doi: 10.1080/01690960444000197

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

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied linguistics*, 27(1), 1–24.

Feldman, L. B. (2000). Are morphological effects distinguishable from the effects of shared meaning and shared form? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6), 1431–1444. doi: 10.1037/0278-7393.26.6.1431

Feldman, L. B., Barac-Cikoja, D., & Kostić, A. (2002). Semantic aspects of morphological processing: Transparency effects in Serbian. *Memory and Cognition*, 30(4), 629–636. doi: 10.3758/BF03194964

Feldman L. B., & Basnight-Brown D. M. (2008). List context fosters semantic processing: parallels between semantic and morphological facilitation when primes are forward masked. *Journal of Experimental Psychology: Learning, Memory and Cognition*. May; 34(3), 680–7. doi: 10.1037/0278-7393.34.3.680.

Feldman, L. B., & Larabee, J. (2001). Morphological facilitation following prefixed but not suffixed primes: Lexical architecture or modality-specific processes?. *Journal of Experimental Psychology: Human Perception and Performance*, 27(3), 680–691. doi: 10.1037//0096-1523.27.3.680

Feldman, L. B., Marelli, M., Amenta, S. and Milin, P. (2015). Effects of target density on morphological processing at long and short SOAs. First Quantitative Morphology Meeting. Belgrade, Serbia. 11-12 July 2015.

Feldman, L. B., Kostić, A. Gvozdenović, V., O'Connor, P. A. and Moscoso del Prado

Martín, F. (2012). Semantic similarity influences early morphological priming in Serbian: A challenge to form-then-meaning accounts of word recognition. *Psychonomic Bulletin and Review*, 19(4), 668–676. doi: 10.3758/s13423-012-0250-x.

Feldman, L. B., Milin, P., Cho, K., Moscoso del Prado Martín, F. & O'Connor, P. A. (2014). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, 9. doi: 10.3389/fnhum.2015.00111

Feldman, L. B. & Moscoso del Prado Martín, F. (2012). Does a focus on universals represent a new trend in word recognition?. *Behavioral and Brain Science*, 35(5). doi: 10.1017/S0140525X12000295.

Feldman, L. B., O'Connor, P. A., & Moscoso del Prado Martín, F. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, 16(4), 684–691. doi: 10.3758/PBR.16.4.684

Feldman, L. B., Smolka, E., Cho, K., & Milin, P. (2014). Effects of semantic transparency on morphological processing: Prefixed English Verbs. Paper presented at the Ninth International Meeting of the Mental Lexicon. Niagara on the Lake, ON. Oct. 2014.

Feldman, L. B., & Soltano, E. G. (1999). Morphological priming: The role of prime duration, semantic transparency, and affix position. *Brain and Language*, 68(1-2), 33–39. doi: 10.1006/brln.1999.2077

Feldman, L. B., Soltano, E. G., Pastizzo, M. J., & Francis, S. E. (2004). What do graded effects of semantic transparency reveal about morphological processing?. *Brain and Language*, 90(1), 17–30. Brln.1999.1077 doi: 10.1016/S0093-934X(03)00416-4

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, 30(1), 1–25. doi: 10.1016/0749-596X(91)90008-8

Forster, K. I., Davis, C., Schoknecht, C., and Carter, R. (1987). Masked priming with graphemically related forms: repetition or partial activation?. *Quarterly Journal of Experimental Psychology*, 39(2), 211–251. doi: 10.1080/14640748708401785

Forster, K. I., and Taft, M. (1994). Bodies, antibodies, and neighborhood density effects in masked form-priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 844–863. doi: 10.1037/0278-7393.20.4.844

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, 355, 263–279. doi: 10.1017/S0140525X11001841

Frost, R., Deutsch, A. & Forster, K. I. (2000). Decomposing Morphologically Complex Words in a Nonlinear Morphology. *Journal of Experimental Psychology Learning, Memory, and Cognition*, 26(3), 751–765. doi: 10.1037//0278-7393.26.3.751

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., & Marslen-Wilson, W. (2000). Morphological priming: Dissociation of phonological, semantic and morphological factors. *Memory and Cognition*, 28(8), 1277–1288. doi: 10.3758/BF03211828

Frost, R., Forster, K. I., & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked-priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4), 829–856. doi: 10.1037//0278-7393.23.4.829

Frost, R., Kugler, T., Deutsch, A., & Forster, K. I. (2005). Orthographic structure versus morphological structure: Principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1293–1326. doi: 10.1037/0278-7393.31.6.1293

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, 136(2), 323–345. doi: 10.1037/0096-3445.136.2.323

Harm, M. W., & Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review*, 111(3), 662–720. doi: 10.1037/0033-295X.111.3.662

Hendrix, P. (2015). *Experimental explorations of a discrimination learning approach to language processing*. Doctoral dissertation. Eberhard Karls University Tübingen. doi: 10.15496/publikation-9333

Hinton, G. E., & Nowlan, S. J. (1987). How learning can guide evolution. *Complexsystems*, 1(3), 495–502.

Holcomb, P., and Grainger, J. (2006). On the time course of visual word recognition: an event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience*, 18(10), 1631–1643. doi: 10.1162/jocn. 2006.18.10.1631

Kielar, A., Joanisse, M. F., and Hare, M. L. (2008). Priming English past tense verbs: rules or statistics? *Journal of Memory and Language*, 58(2), 327–346. doi: 10.1016/ j.jml.2007.10.002

Kielar, A., & Joanisse, M. F. (2011). The role of semantic and phonological factors in word recognition: An ERP cross-modal priming study of derivational morphology. *Neuropsychologia*, 49(2), 161–177. doi: 10.1016/j. neuropsychologia.2010.11.027

Kinoshita, S., Castles, A., & Davis, C. (2009). The role of neighbourhood density in transposed-letter priming. *Language and Cognitive Processes*, 24(4), 506–526. doi: 10.1080/01690960802289213

Kostić, A. (1991). Informational approach to processing of inflected morphology: standard data reconsidered. *Psychological Research*, 53(1), 62–71. doi: 10.1007/ BF00867333

Kostić, A. (1995). Informational load constraints on processing inflected morphology. In L. B. Feldman (Ed.), *Morphological Aspects of Language Processing*, Hillsdale,NJ: Erlbaum. doi: 10.1007/BF00867333

Kostić, A., Marković, T., & Baucal, A. (2003). Inflectional morphology and word meaning: orthogonal or co-implicative cognitive domains? R. H. Baayen & R. Schreuder (Eds.): *Morphological Structure in Language Processing*. Mouton de Gruyter.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211. doi: 10.1037/0033-295X.104.2.211

Lavric, A., Clapp, A., and Rastle, K. (2007). ERP evidence of morphological analysis from orthography: a masked priming study. *Journal of Cognitive Neuroscience*, 19(5), 866–877. doi: 10.1162/jocn. 2007.19.5.866

Lavric, A., Rastle, K., & Clapp, A. (2011). What do fully visible primes and brain potentials reveal about morphological decomposition?. *Psychophysiology*, 48(5), 676–686. doi: 10.1111/j.1469-8986.2010.01125.x

Longtin, C., Segui, J., & Hallé, P. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, 18(3), 313–334. doi: 10.1080/01690960244000036

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, 44(3), 295–322. doi: 10.1177/00238309010440030101

Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of Orthography-Semantics Consistency on word recognition. *The Quarterly Journal of Experimental Psychology*, 68(8), 1571–1583.

Marelli, M. & Baroni, M. (2015). Affixation in Semantic Space: Modeling Morpheme Meanings With Compositional Distributional Semantics. *Psychological Review*, 122(3), 485–515. doi: 10.1037/a0039267

Marslen-Wilson, W., Bozic, M., & Randall, B. (2008). Early decomposition in visual word recognition: Dissociating morphology, form, and meaning. *Language and Cognitive Processes*, 23(3), 394–421. doi: 10.1080/01690960701588004

Marslen-Wilson,W., & Tyler, L.K. (1998). Rules, representations, and the English past tense. *Trends in Cognitive Sciences*, 2(11), 428–435. doi: 10.1016/S1364-6613(98)01239-X

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 10(1), 3–33. doi: 10.1037/0033-295X.101.1.3

Matthews, P. H. (1991). Morphology. Cambridge: Cambridge University Press.

McCormick, S. F., Rastle, K., & Davis, M., H. (2009). Adore-able not adorable? Orthographic underspecification studied with masked repetition priming. *European Journal of Cognitive Psychology*, 21(6), 813–836. doi: 10.1080/09541440802366919

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, 44(3), 295–322.

Meunier, F., & Longtin, C.-M. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, 56(4), 457–471. doi: 10.1016/j.jml.2006.11.005

Meunier, F., & Segui, J. (2002). Cross-modal priming in French. *Brain and Language*, 81(1), 89–102. doi: 10.1006/brln.2001.2509

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., & Baayen, R. H. (2017). Discrimination in lexical decision. *PLoS One*, 12(2), e0171935.

Milin, P., Filipović Đurđević, D. & Moscoso del Prado Martín, F. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, 60(1), 50–64. doi: 10.1016/j.jml.2008.08.007

Milin, P., Kuperman, V., Kostić, A., & Baayen, H. R. (2009). Words and paradigms bit by bit: An information-theoretic approach to the processing of inflection and derivation.In: J. P. Blevins and J. Blevins (Eds.), *Analogy in Grammar: Form and Acquisition* (pp. 214–252). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199547548.003.0010

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38(11), 39–41. doi: 10.1145/219717.219748

Mirković, J., Seidenberg, M. S., & Joanisse, M. F. (2011). Rules versus statistics: Insights from a highly inflected language. *Cognitive Science*, 35(4), 638–681. doi: 10.1111/j.1551-6709.2011.01174.x

Morris, J., Franck, T., Grainger, J., P. J. Holcomb, P. J. (2007.) Semantic transparency and masked morphological priming: An ERP investigation. *Psychophysiology*, 44(4), 506–521. doi: 10.1111/j.1469-8986.2007.00538.x

Morris, J., Grainger, J., & Holcomb, P., J. (2008). An electrophysiological investigation of early effects of masked morphological priming. *Language and Cognitive Processes*, 23(7-8), 1021–1056. doi: 10.1080/01690960802299386

Morris, J., Grainger, J., & Holcomb, P., J. (2013). Tracking the consequences of morpho-orthographic decomposition using ERPs. *Brain Research*, 1529, 92–104. doi: 10.1016/j.brainres.2013.07.016

Morris, J., Porter, J., H., Grainger, J., & Holcomb, P., J. (2011). Effects of lexical status and morphological complexity in masked priming: an ERP study. *Language and Cognitive Processes*, 26(4-6), 558–599. doi: 10.1080/01690965.2010.495482

Moscoso del Prado Martín, F., Kostic, A., & Baayen, R. H. (2004). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition*, 94(1), 1–18. doi: 10.1016/j.cognition.2003.10.015

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, 102(6), 963–968. doi: 10.1037/h0036551

Napps, S. E., & Fowler, C. A. (1987). Formal relationships among words and the organization of the mental lexicon. *Journal of Psycholinguistic Research*, 16(3), 257–272. doi: 10.1007/BF01067546

Patterson K., Lambon Ralph M.A., Hodges, J.R., and McClelland. J.L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia*, 39(7), 709–24. doi: 10.1016/S0028-3932(01)00008-2

Perea, M., Acha, J., & Fraga, I. (2008). Lexical competition is enhanced in the left hemisphere: Evidence from different types of orthographic neighbors. *Brain and Language*, 105(3), 199–210. doi: 10.1016/j.bandl.2007.08.005

Perry, J. R., Lupker, S. J. & Davis, C. J. (2008). An evaluation of the interactive-activation model using masked partial-word priming. *Language and Cognitive Processes*, 23(1), 36–68. doi: 10.1080/01690960701578112

Pinker, S. (1999). *Words and Rules*.New York, NY: Harper Perennial.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6, 456–463.

Pinker, S., and Ullman, M. (2003). Beyond one model per phenomenon. *Trends in Cognitive Sciences*, 7(3), 108–109. doi: 10.1016/S1364-6613(03)00021-4

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing?. *Language and Cognitive Processes*, 15(4-5), 445–485. doi: 10.1080/01690960050119661

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56–115. doi: 10.1037//0033-295X.103.1.56

Ramscar, M., & Baayen, H. (2013). Production, comprehension, and synthesis: a communicative perspective on language. *Frontiers in Psychology*, 4, 233. doi: 10.3389/fpsyg.2013.00233

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science*, 6(1), 5–42. doi: 10.1111/tops.12078

Ramscar, M. & Port, R. (2015). Categorization (without categories). In E. Dawbroska & D. Divjak (Eds.), *Handbook of Cognitive Linguistics*. Berlin, DE: De Gruyter Mouton.

Ramscar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science*, 31(6), 927–960. doi: 10.1080/03640210701703576

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34(6), 909–95. doi: 10.1111/j.1551-6709.2009.01092.x

Rastle, K., Davis, M., H., Marslen-Wilson, W., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, 15(4-5), 507–537. doi: 10.1080/01690960050119689

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, 23(7-8), 942–971. doi: 10.1080/01690960802069730.

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review*, 11(6), 1090–1098. doi: 10.3758/BF03196742

Raveh, M., & Rueckl, J. (2000). Equivalent effects of inflected and derived primes: Long-term morphological priming in fragment completion and lexical decision. *Journal of Memory and Language*, 42(3), 103–119. doi: 10.1006/jmla.1999.2673

Ravid, D. (2012). *Spelling morphology: The psycholinguistics of Hebrew spelling*. New York: Springer. doi: 10.1007/978-1-4419-0588-8_13

Rescorla, R. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43(3), 151–160. doi: 10.1037//0003-066X.43.3.151

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

Roelcke, T. (1997). Sprachtypologie des Deutschen: historische, regionale und funktionale Variation [Language typology of German: historical, regional, and functional variation] (Vol. 2812) Berlin: Walter de Gruyter.

Rueckl, J. G. (2002). The dynamics of visual word recognition. *Ecological Psychology*, 14(1-2), 5–19. doi: 10.1080/10407413.2003.9652750

Rueckl, J. G., & Galantucci, B. (2005). The locus and time course of long-term morphological priming. *Language and Cognitive Processes*, 20(1-2), 115–138. doi: 10.1080/01690960444000188

Rueckl, J. G., Mikolinski, M., Raveh, M., Miner, C. S., & Mars, F. (1997). Morphological priming, fragment completion, and connectionist networks. *Journal of Memory and Language*, 36(3), 382–405. doi: 10.1006/jmla.1996.2489

Rueckl, J., & Raveh, M. (1999). The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language*, 68(1-2), 110–117. doi: 10.1006/brln.1999.2106

Seidenberg, M., & Gonnerman, L. M. (2000). Explaining derivational morphology asthe convergence of codes. *Trends in Cognitive Sciences*, 4(9), 353–361.

Shannon's (1948). doi: 10.1016/S1364-6613(00)01515-1

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, 11(2), 245–260. doi: 10.1017/S1366728908003404

Smolka, E., Gondan, M., & Rösler, F. (2015). Take a stand on understanding: Electrophysiological evidence for stem access in German complex verbs. *Frontiers in Human Neuroscience*, 9, 62. doi: 10.3389/fnhum.2015.00062.

Smolka, E., Khader, P., Wiese, R., Zwitserlood, P., & Rösler, F. (2013). Neurobiological evidence for the continuous processing of linguistic categories of regular and irregular verb inflection in German. *Journal of Cognitive Neuroscience*, 25(8), 1284–1304. doi: 10.1162/jocn_a_00384

Smolka, E., Komlósi, S., & Rösler, F. (2009). When semantics means less than morphology: Processing German prefixed verbs. *Language and Cognitive Processes*, 24(3), 337–375. doi: 10.1080/01690960802075497

Smolka, E., & Libben, G. (2017). "Can you wash off the hogwash?" – semantic transparency of first and second constituents in the processing of German compounds. *Language, Cognition and Neuroscience*, 32(4), 514–531. doi: 10.1080/23273798.2016.1256492

Smolka, E., Preller, K., & Eulitz, C. (2014). 'Verstehen' 'understand' primes 'stehen' 'stand': Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language*, 72, 16–36. doi: 10.1016/j.jml.2013.12.002

Smolka, E., Zwitserlood, P., & Rösler, F. (2007). Stem access in regular and irregular inflection: Evidence from German participles. *Journal of Memory and Language*, 57(3), 325–347. doi: 10.1016/j.jml.2007.04.005

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 638–647. doi: 10.1016/S0022-5371(75)80051-X

Taft, M., & Kougious, P. (2004). The processing of morpheme-like units in monomorphemic words. *Brain and Language*, 90(1), 9–16. doi: 10.1016/S0093-934X(03)00415-2

Taft, M., & Nguyen-Hoan, M. (2010). A sticky stick? The locus of morphological representation in the lexicon. *Language and Cognitive Processes*, 25(2), 277–296. doi: 10.1080/01690960903043261

Ullman, M. T. (2004). Contributions of neural memory circuits to language: The declarative/procedural model. *Cognition*, 92*(*1-2*)*. 231–270. doi: 10.1016/j.cognition.2003.10.008

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research*, 30(1). 37–69. doi: 10.1023/A:1005204207369

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15, 971–979. doi: 10.3758/PBR.15.5.971

Zwitserlood, P., Bolwiender, A., & Drews, E. (2005). Priming morphologically complex verbs by sentence contexts: Effects of semantic transparency and ambiguity. *Language and Cognitive Processes*, 20(1-2), 395–415. doi: 10.1080/01690960444000160