Psychonomic Bulletin and Review (2022) 29:1673–1702
https://doi.org/10.3758/s13423-022-02086-0

THEORETICAL/REVIEW

From decomposition to distributed theories of morphological processing in reading

Patience Stevens[1] · **David C. Plaut**[1]

Accepted: 6 March 2022 / Published online: 20 May 2022

Abstract

The morphological structure of complex words impacts how they are processed during visual word recognition. This impact varies over the course of reading acquisition and for different languages and writing systems. Many theories of morphological processing rely on a decomposition mechanism, in which words are decomposed into explicit representations of their constituent morphemes. In distributed accounts, in contrast, morphological sensitivity arises from the tuning of finer-grained representations to useful statistical regularities in the form-to-meaning mapping, without the need for explicit morpheme representations. In this theoretically guided review, we summarize research into the mechanisms of morphological processing, and discuss findings within the context of decomposition and distributed accounts. Although many findings fit within a decomposition model of morphological processing, we suggest that the full range of results is more naturally explained by a distributed approach, and discuss additional benefits of adopting this perspective.

Keywords Morphological processing · Visual word recognition · Cross-linguistic · Reading acquisition

The central goal of reading is comprehension, and arguably the most informative unit of language is the word; thus, the cognitive processes underlying written word recognition are of enormous import, justifying the wealth of research that has been devoted to understanding them over the last century. Traditionally, the relationship between the surface form of a word and its meaning is described as arbitrary (e.g., Hockett & Hockett, 1960): for example, the words BEACH and PEACH look and sound alike but mean very different things, and thus perceptual similarity must be disregarded in deriving their meanings. However, this is not the case for morphologically complex words, such as BEDROOM and BEDDING, or MINIATURE and MINIMIZE. In these words, perceptual similarity does convey similarity in meaning. Complex words are the result of a productive morphology: a system, dependent on the language in question, by which recurring sequences of letters or phonemes can be combined or transformed to convey more varied and nuanced meanings. In some languages, the written form of complex words can provide more information about the words' meaning than the spoken form. In English, for example, the stems of the words MAGICIAN and HEALTH are more salient when read than when heard (see Rastle, 2019a). The morphological structure of complex words makes the form-to-meaning mappings that are learned during reading acquisition more systematic, and this regularity presents opportunities for more efficient access of word meanings.

Ample evidence indicates that, during the time-pressured task of reading, the visual system does make use of morphological structure to process words. For example, after learning a list of words that includes the word CAR, English readers recognize a morphologically related word like CARS more quickly than a word that is merely visually similar, like CARD (Murrell & Morton, 1974). Uncommon complex words in Italian are recognized more quickly if they have high-frequency as opposed to low-frequency stems, suggesting more common morphemes better facilitate the recognition of the word (Burani, Salmaso, & Caramazza, 1984; Taft, 1979). Researchers have found evidence of sensitivity to morphological structure in reading in a wide variety of languages and writing systems, including French (Grainger, Colé, & Segui, 1991), Spanish (Duñabeitia, Perea, & Carreiras, 2008), Dutch (Drews & Zwitserlood, 1995), Hebrew (Frost, Forster, & Deutsch,

✉ Patience Stevens
pstevens@andrew.cmu.edu

1 Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213–3890, USA

1997), Chinese (Tsang & Chen, 2013), Japanese (Clahsen & Ikemoto, 2012), and Serbian (Feldman, 1994). Despite the variability in morphological systems across these languages, morphological structure is consistently learned and used during lexical processing.

Given widespread evidence for morphological sensitivity during visual word recognition, much attention has been paid to how a word's morphological structure affects its underlying representations and processes. Most prominent theories rely on a *decomposition* mechanism that splits words into their constituents (e.g., TEACHER into TEACH and ER) at some phase of processing. As the current review will make apparent, the decomposition perspective has considerable empirical support. However, as will also become apparent, accounting for the full span of evidence across languages and contexts with decomposition as a central mechanism is proving increasingly awkward. This has led some researchers to explore *distributed* accounts, in which morphemes are not represented, but morphological structure is captured nonetheless through learned sensitivity to statistical structure among word forms and their meanings (see the next section for a more detailed description). In our view, a distributed approach to morphological processing holds more promise for providing a comprehensive account of the full range of observed empirical effects, including those that would seem to implicate explicit decomposition. However, important details of the distributed account have yet to be fleshed out.

There are many reviews of the literature on morphological processing (e.g., Amenta & Crepaldi, 2012; Diependaele, Grainger, & Sandra, 2012; Hay & Baayen, 2005; Marelli, Traficante, & Burani, 2020; Milin, Smolka, & Feldman, 2018), in part because the relevant findings are extensive and sometimes inconsistent. The goal of the current review is to provide a comprehensive summary and theoretical discussion of empirical evidence regarding the nature of morphological processing in visual word recognition. The selection of publications to be included was delineated accordingly. First, this review focuses entirely on the visual aspects of word processing. This means that studies aiming to investigate auditory language processing or language processing that is not modality-specific (e.g., results obtained using cross-modal priming) were not included unless they provide an informative contrast with visual effects. Additionally, to focus on processing during the mapping from visual word forms to their meanings, only evidence from tasks that promote semantic access were included. Such tasks include lexical decision (determining a stimulus' status as a word or nonword), determining lexical category (e.g., is this a noun or an adjective?), and sentence reading.[1] Finally, the most useful theoretical model would be one that explained not only how processing occurs but also how those mechanisms emerge with language experience. Thus, the current review will emphasize papers that highlight differences in morphological processing across languages or over the course of reading acquisition.

Theories of morphological processing

Theories of morphological processing are diverse, and have evolved significantly over the past few decades. An enduring divide among these approaches, however, is whether decompositional or distributed mechanisms are relied upon to explain sensitivity to word-form structure. To contextualize the ensuing discussions of established empirical effects, we begin by describing the history and details of both the decomposition and distributed approaches, as well as how they relate to each other.

Decomposition-based theories

When psycholinguists in the '70s and '80s found evidence for a role of morphological structure in word processing, the primary theories for what that role might be were *listing* and *parsing* theories (Marelli et al., 2020). Listing theories posited that all words are stored in their full form, and morphologically related words activate each others' representations after the presented word's representation has been accessed (e.g., Butterworth, 1983; Manelis & Tharp, 1977). *Parsing* or *decomposition* theories, on the other hand, proposed that complex words are decomposed into their constituent morphemes, prior to recognition and access of the word's overall meaning (e.g., Murrell & Morton, 1974; Stanners, Neiser, Hernon, & Hall, 1979; Taft & Forster, 1975). The decompositional approach has considerable theoretical appeal, as it uses a language's morphological systematicity to make the storage and processing of words more efficient. Additionally, this approach better accounts for a majority of the evidence on morphological processing accumulated over subsequent decades, as will be described shortly. Consequently, decomposition theories have generally gained more favor than listing theories among psycholinguists.

[1] Word naming was not included, due to previous findings suggesting that semantic characteristics of complex words show strong effects for lexical decision but not for word naming, while phonological characteristics play a stronger role in word naming (e.g., Baayen, Wurm, & Aycock, 2007; Burani, Dovetto, Spuntarelli, & Thornton, 1999).

Advocates of decomposition theories have varied in their descriptions of precisely how decomposition is executed. In Rastle and Davis (2008), decomposition is described as “morpho-orthographic”, applied to any word that has the appearance of morphological complexity (e.g., applied indiscriminately to both the pseudo-complex word CORNER and the truly complex word TEACHER). Crepaldi, Rastle, Coltheart, and Nickels, (2010) proposed that, in addition to an initial level of morpho-orthographic representations, a lemma level exists to allow for more abstracted identification of constituent morphemes. Grainger and Ziegler (2011) emphasized the inhibition of affixes to allow for stem recognition as the primary means of decomposition, whereas Taft (1994) argued for the simultaneous activation of both stem and affix representations. Despite such variation in descriptions of decomposition, the common element among them is that, within some phase or route of word processing, the word is represented as a combination of independent morphemic units.

An entirely combinatorial approach to processing complex words, however, would work only for a language with a perfectly systematic morphology. This certainly does not hold for English: several words appear complex but aren’t (such as CORNER and WITNESS), whereas others are morphologically structured but the morphemes’ contributions to word meaning must be interpreted more loosely (e.g., RECALL doesn’t mean “to call again”). Similar challenges arise in most other languages, including French (-ETTE means “little” in FILLETTE, “little girl”, but not in BAGUETTE, “breadstick”) and Chinese (公 means “public” in 公园, “public park”, but not in 公鸡, “rooster”). Irregular words, such as TAUGHT from TEACH, additionally complicate the form-to-meaning mapping by introducing variability in complex words’ appearance. Thus, a system for processing complex words must be able to take advantage of systematicity where relevant, but also handle words where morphological structure is less informative or even misleading.

Decomposition theories satisfying these constraints are mostly combinations of, or compromises between, listing and parsing theories. Taft (1979) refined his prefix-stripping theory to assert that words are stored *both* by their root morpheme, produced by splicing off affixes, and by their whole form. Clahsen (1999) distinguished routes for irregular versus regular processing of inflected words: irregular words are processed via structured lexical entries, whereas regular words are subject to affix-stripping. The Parallel Dual-Route Model proposed by Schreuder and Baayen (1995; also see Baayen, Dijkstra, & Schreuder, 1997 and Grainger & Ziegler, 2011) posits two parallel access routes: one that processes the whole word directly and a second that splits morphological constituents and constructs meaning from them, with reaction times determined by the fastest route for a particular word. Despite differing in important ways, these models all assume that morphemes are explicitly represented during complex word processing, and give a central role to the decomposition of at least some complex words into constituent morphemes (Rueckl, 2010). Decomposition theories have shaped the framing of research on morphological processing for decades (see reviews by Amenta & Crepaldi, 2012; Diependaele et al., 2012; Marelli et al., 2020) and continue to feature prominently in current discussions of this subfield (e.g., Fleischhauer, Bruns, & Grosche, 2021; Ciaccio, Kgolo, & Clahsen, 2020; De Grauwe, Lemhöfer, & Schriefers, 2019).

Distributed theories

Distributed theories of cognition provide an alternative theoretical perspective on morphological processing. Such theories are most commonly implemented in the form of artificial neural networks (also known as *connectionist* models). Neural networks consist of a system of connected units, typically arranged in layers. The activation of one unit is determined by the activations of the units connected to it, which are weighted by the strengths of their respective connections and summed. This sum is then passed through a nonlinear function. The values of connections between units (referred to as the *weights*), are adjusted with training to optimize performance on a given task. Neural networks are often mischaracterized as being purely associative, only able to learn surface-level statistical regularities. In fact, such networks can learn to make use of abstract structural relationships and perform complex operations, if doing so is useful for the task at hand (see Elman, Bates, & Johnson, 1996; Rumelhart, Hinton, & Williams, 1986, for overviews on using neural networks to understand cognition; see Gonnerman, Seidenberg, & Andersen, 2007; Plaut & Gonnerman, 2000; Rueckl, 2010 for efforts to apply them to morphological processing).

Distributed theories that tackle the processing of complex words trace back to a controversial chapter by Rumelhart & McClelland (1986). A neural network free of rules or explicit morpheme representations was presented as a potential account of how English speakers generate the past tense of verbs from base forms. Rumelhart & McClelland’s model, unlike other linguistic accounts of conjugation at that time, did not require discrimination between regular (“wanted”, “sounded”) and irregular (“went”, “threw”) forms, nor any explicit representations of suffixes or phonological transformations. Instead, the only morphological information the model received was implicit in the structure of the task it was trained to perform: phonological representations of the base forms of a small set of verbs were presented to a neural network, which was trained via adjustment of unit-connecting weights to output the phonological representation of the past tense form. Within this simple learning environment, the model learned the base-to-past mappings well and was able to generalize to novel verbs (e.g., "wug" to "wugged"). In some phases of training it overgeneralized the "-ed" rule to irregular forms ("goed"), an observed tendency in children learning language. The chapter spurred an "intellectual firestorm" among linguists and cognitive scientists regarding the nature of linguistic knowledge (for a summary of these exchanges and their impact on theories of language processing, see Seidenberg & Plaut, 2014).

Even with its limitations, Rumelhart & McClelland's work on learning the English past tense provided an excellent example of how distributed mechanisms can illuminate a specific cognitive task, and demonstrated their particular suitability for linguistic phenomena. Thus, as studies and models of morphological processing in the context of word recognition gained increasing attention, the need for a neural-network account of morphological sensitivity in word comprehension (as opposed to in production) became clear. Rueckl and Raveh (1999) trained a neural network with one hidden layer to map from orthographic representations to semantic representations for two artificial languages. In one language, morphological regularity was captured by constructing words from invented stems and suffixes, concatenating stem and suffix orthographic representations and combining their semantic representations. The other language was constructed by shuffling the first language's mappings from form to meaning, such that words' morphological structures did not inform their meanings. Rueckl & Raveh found that the language with morphological regularity required less training to perform well and could accommodate larger vocabularies, demonstrating how morphology enables efficient word learning. Additionally, they found evidence of compositionality in network processing after training: words containing a particular stem had similar patterns of hidden layer activation. Critically, this intermediate sensitivity emerged by applying general learning mechanisms to the whole-word form-to-meaning mappings the authors constructed, and no explicit morphological knowledge was encoded in the model itself.

Plaut and Gonnerman (2000) also trained neural networks on form-to-meaning mappings, and demonstrated that the stronger overt priming by semantically opaque words (i.e., words that appear to be but are not complex, such as DEPARTMENT and CORNER) in Hebrew compared to English can be attributed to the languages' differences in morphological systematicity. Networks were trained to map orthographic representations to semantic representations for the same set of artificial words, together with a remaining vocabulary that was either entirely transparent in how morphological structure informed words' meanings (morphologically *rich*) or entirely opaque (morphologically *impoverished*). When trained in the context of a morphologically rich language, likened to Hebrew, the network was faster to settle on a representation of a word after being presented with a morphologically but not semantically related word (i.e., opaque priming occurred). In the context of a morphologically impoverished language, likened to English, the network did not show such facilitation, as is seen in overt priming contexts in English. Plaut & Gonnerman also simulated morphological priming effects with diminishing magnitude for complex words whose meanings were semantically close versus distant from the target. This provides a concrete account for graded contributions of morphological structure sensitivity, as seen in Jared, Jouravlev, and Joanisse (2017) and Gonnerman et al. (2007). Both Rueckl and Raveh (1999) and Plaut and Gonnerman (2000) demonstrate how theories of morphological processing, as well as the emergence and variation of morphological processing across languages, can be explored and understood using a distributed approach.

In a distinct but related vein of modeling work, Baayen, Milin, Revic, Hendrix, & Marelli, (2011) applied *naive discriminative learning* to the form-to-meaning mappings of real text corpora. This approach uses distributed weight learning mechanisms like those in neural network models, and similarly provides no explicit morphological information to the model. Naive discriminative learning has been used to simulate effects such as family size in English (Milin, Feldman, Ramscar, Hendrix, & Baayen, 2017), inflectional entropy in Serbian (Baayen et al., 2011), and semantic transparency in German (Baayen & Smolka, 2020). A clear benefit of this method is that training can be efficiently applied to actual words, capturing realistic linguistic regularities (as opposed to those in artificial languages which are exaggerated for the sake of demonstration) and improving ease of comparison with empirical results.

The naive discriminative learning approach, though informative, is limited in the complexity of the functions it can learn. Functionally, it is similar to a neural network with no internal or "hidden" representation layers. Such networks can only learn transformations that are linearly separable—which is tantamount to being nearly completely systematic—and many aspects of morphology are not systematic. For example, the semantic role of -ER varies based on the accompanying stem, as in GREATER and TEACHER, as does that of -STICK in CHOPSTICK and UNSTICK: although some morphemes may appear to make consistent, independent semantic contributions to the words in which they appear, the majority do not. The presence of hidden layers is also the reasonable choice from a neurobiological perspective. Thus, as is articulated in discussions of this work (e.g., Milin et al., 2017), naive discriminative learning approaches may be best understood as a useful and interpretable predictive tool built on the foundational concepts of distributed modeling, rather than as a mechanistic implementation of word processing itself.

Distinguishing distributed and decomposition theories

Perhaps the key theoretical contrast between decomposition and distributed accounts is the following: Whereas, on a decomposition account, a given word either does or does not contain a morpheme, on a distributed account, the notion of "containing" a morpheme—as a recurring string of letters or phonemes—or of "sharing" a morpheme with another word, is entirely a matter of degree. Thus, on the latter, the representation of DRESSER doesn't contain that of DRESS; rather, the contribution of the letter string DRESS to the internal representation of DRESSER is highly similar to—but not identical to—the varying contributions DRESS makes to the representations of other words (e.g., DRESSING, REDRESS). In light of this contrast, decomposition and distributed theories are often pitted against each other in discussions of morphological processing mechanisms (e.g., Fleischhauer et al., 2021; Jared et al., 2017; Marelli et al., 2020). It is worth noting, however, they are not always in conflict, and can be difficult to distinguish empirically. In the context of a language with a perfectly systematic morphology, they could even be thought of as describing the same phenomenon at differing levels of detail. As demonstrated by Plaut and Gonnerman (2000) in the morphologically rich condition, a fully-trained distributed system can give rise to decomposition-like phenomena for all complex words. In a perfectly systematic language, hidden layers' activations in response to complex words might be mostly divisible into independent contributions of constituent morphemes, and a morpheme's contribution to these representations could be nearly identical, and hence functionally equivalent, across its appearances in different words. In this case, discriminating between distributed and decomposition theories might feel a bit like splitting hairs—at least with regard to the performance of skilled adult readers—and it would be natural to characterize morphological processing as involving decomposition. The "rule" of decomposition would have emerged from the sensitivity of local learning mechanisms to highly reliable regularities. Differences between the two accounts become more critical, however, when considering the prevalence of quasi-regularity (i.e., rules that are broken to varying degrees, all the way to semantically opaque words like COURTEOUS and DEPARTMENT) and nonlinearity (i.e., context-dependence of morpheme contributions) in morphological systems across languages. Even here, though, it is worth keeping in mind that graded effects can often be captured by differential and dynamic weighting among an ensemble of explicit, discrete representations, as in so-called "localist" connectionist models (e.g., Dell, 1986; McClelland & Rumelhart, 1981; Taft, 1994), although formulating an effective learning procedure for how such representations are identified, weighted, and updated remains a challenge.

Relatedly, sensitivity to statistical regularities present in language is not a distinguishing feature of the distributed view. Both decomposition and distributed perspectives have embraced the idea that morphological sensitivity is driven by features of the language (e.g., Baayen et al., 2011; Dawson, Rastle, & Ricketts, 2021b; Grainger & Beyersmann, 2017; Rastle & Davis, 2008; Plaut & Gonnerman, 2000). In the case of decomposition views, this is discussed in terms of how morpheme units are learned, such that complex words can be identified and decomposed (Grainger & Beyersmann, 2017; Rastle & Davis, 2008); in the case of distributed views, the representations themselves are shaped by regularities to become sensitive to morphological structure in a graded manner (Baayen et al., 2011; Plaut & Gonnerman, 2000). A strength of the distributed view is that theories of how language statistics shape representations and processes have been specified to the degree that this learning can be demonstrated computationally (Tamminen, Davis, & Rastle, 2015).

A final important issue to consider is the nature of assumptions regarding orthographic and semantic representations. Representations are less prominently discussed in the context of decomposition models, mainly because these models are rarely implemented computationally and so the information inherent in a word's form and meaning do not need to be explicitly described. However, this is an especially critical aspect of the distributed account, as the nature of representations has a strong impact on how (and what) distributed networks learn.

For orthographic representations, Rueckl and Raveh (1999) gave each possible letter and position combination a corresponding input unit, whereas Plaut and Gonnerman (2000) used two 15-unit binary patterns, one for the first syllable and one for the second syllable. Because the language in Rueckl and Raveh (1999) was entirely made up of bimorphemic words with 3-letter stems and 1-letter affixes, the inputs of both these models could be viewed as pre-segemented in that each input unit contributes to either the first or second morpheme across all examples (Rastle & Davis, 2008). However, the networks still needed to learn which units correspond to each morpheme and, in the case of Plaut and Gonnerman (2000), which second syllables were morphologically informative.

Neither Plaut and Gonnerman (2000) nor Rueckl and Raveh (1999) simulate cases such as DISTRUST, UNTRUSTING and TRUSTWORTHY, where morphological structure (TRUST) varies in input position. In Baayen et al. (2011) on the other hand, bigram input units necessitate that at least the first and last letters of a morpheme will be represented differently across instances: The first T in TRUST is represented as ST in DISTRUST, NT in UNTRUSTING and #T in TRUSTWORTHY. However, the single letter input units and any internal bigrams remain consistent across instances. The use of bigram (and trigram; e.g., Milin et al., 2017) input units might also be interpreted as providing the model with pre-segmented morphemes in the case of two- and three-letter morphemes; however, the model is still tasked with learning to differentiate bigram units with morphological status from those without. This likely isn't an unrealistic characteristic of morphological learning, given readers' robust sensitivity to non-morphological letter sequence frequencies (e.g., Broadbent and Gregory, 1968; Rice & Robinson, 1975). Although the lack of morphological structures with non-overlapping positions should be addressed in future distributed simulations for completeness, a neural network can learn sensitivity to informative elements at variable positions within input, provided the network has sufficient layers (Hannagan, Agrawal, Cohen, & Dehaene, 2021).

Semantic representations used in distributed models have been relatively simplistic: both Rueckl and Raveh (1999) and Plaut and Gonnerman (2000) used binary representations for which a transparent complex word's meaning was a superposition of its morphological constituents. Semantic representations used in discriminative models (e.g., Baayen et al., 2011) were localist, mapping to sets of units that accounted for the word's meaning and grammatical role (e.g., HAND and HANDS might both map to a "hand" unit, while HANDS also maps to a "plural" unit). Although these representations have been sufficient to provide numerous insights into how distributed morphological processing might unfold, semantics are in reality much more sparse and dynamic, with a very heterogeneous similarity structure. One benefit of distributed semantic representations relative to localist representations is the ability to capture degrees of semantic similarity, which can allow for relationships not only among morphologically related words but also semantically similar families to impact how processing unfolds. However, this still falls dramatically short of capturing the contribution of a given word to online comprehension in natural cognition.

In the ensuing review, we consider a range of established empirical findings on morphological processing in visual word recognition, and discuss how well they align with the decomposition and distributed accounts.

Review of empirical findings

A rich body of empirical findings exists regarding the recognition of complex words. Below, we discuss results in light of how they contribute to understanding mechanisms of morphological processing. We start with a consideration of factors that characterize individual morphemes.

Morpheme characteristics

Frequency

In a lexical decision task, high-frequency words are responded to more quickly than low-frequency words (Forster & Chambers, 1973; Scarborough, Cortese, & Scarborough, 1977; Shapiro, 1969). Frequency effects are often interpreted as evidence that the representations of words that are seen or heard more frequently are easier to access (e.g., Morton, 1979). Taft (1979) argued that, correspondingly, if morphemes are also represented during complex word processing, a morpheme's cumulative frequency across different words should predict how quickly a reader recognizes a word containing it. To demonstrate this effect in English, Taft carefully selected prefixed words with similarly low surface frequencies[2] (e.g., RECLINE, DEMOTE) whose stems occur in other contexts with either higher frequency (e.g., in INCLINE and DECLINE) or lower frequency (PROMOTE, EMOTE). For such stimuli, readers make lexical decisions to words with high-frequency stems more quickly than words with low-frequency stems. Stem frequency effects—also referred to as base frequency effects—have been found for both prefixed and suffixed words (Taft, 1979), as well as for words that are inflected, derived (Bradley, 1979) and compounded (e.g., HEADSTAND vs. LOINCLOTH; Taft & Forster, 1976). In addition to English, they are found in Italian (Burani et al., 1984), Dutch (Baayen et al., 1997), French (Colé, Beauvillain, & Segui, 1989), Finnish (Kuperman, Bertram, & Baayen, 2008), and Chinese (Myers, Huang, & Wang, 2006).

Stem frequency effects also interact with words' surface frequency: high-frequency stems facilitate recognition for less-frequent words, but slightly inhibit recognition of more-frequent words (Baayen et al., 2007). Along similar lines, surface frequency is less predictive of lexical decision latencies for low-frequency complex words than for low-frequency simple words (Alegre & Gordon, 1999). Suffix

[2] *Surface frequency* refers to the frequency of the whole word, in contrast to frequencies of the word's morphological constituents.

frequency has also been found to moderate stem frequency effects (Burani & Thornton, 2011). These results suggest that information at the word level and at the morpheme level supplement each other during complex word processing: Very familiar words can be recognized rapidly regardless of the identification of sublexical structures, whereas less-familiar words benefit from the facilitation that those structures provide. Such an interpretation aligns with previously described theories for the parallel roles of decomposition and listing access mechanisms (e.g., Schreuder & Baayen, 1995).

However, decomposition is not the only explanation for stem frequency effects. Burani et al. (1984) argued that such effects can be explained within a listing model, in which words are accessed directly by their surface form and not by their stem. In this model, when a word's representation is activated, the morphologically related words are subsequently activated to a lesser degree. High-frequency words will be easiest to access as a result of their regular occurrence (perhaps due to a lowered threshold of activation; Morton, 1979), but words that are morphologically related to high-frequency words will also be affected due to this lateral activation. Thus, both surface and stem frequency would impact the ease of a word's activation, with the relative import of either predictor depending on its frequency and that of its morphological cousins.

Frequency effects can be found not only for morphemes and simple and complex words, but also for short phrases (e.g., "don't have to worry"; Arnon and Snider, 2010). Given the broad range of linguistic grain sizes over which frequency matters, requiring unique representations for all informative linguistic units seems unwieldy and inefficient. The distributed approach resolves this issue by attributing frequency effects to differences in cumulative weight changes, as opposed to differences in activation thresholds of representations. Linguistic entities that are encountered more frequently have a greater impact on the learned weights of a network, and so the weights are more customized to their accurate and speedy retrieval (Seidenberg & McClelland, 1989). When presented with a high-frequency complex word, then, it is likely that the most influential weights are fine-tuned to that specific word. The recognition of a low-frequency word consisting of common morphemes, on the other hand, will make more use of weights tuned to sublexical structures to benefit from the word's frequent constituents.

As noted above, Baayen et al. (2007) found that higher-frequency stems actually slightly inhibit recognition of words with high surface frequency. *Inverse* stem frequency effects can also be prompted by certain lexical processing tasks: Taft (2004) found that complex words with higher frequency stems were more slowly classified than those with medium-frequency stems when nonword foils contained real stems (e.g., MIRTHS, REDLY), but not when nonwords contained nonsense stems (KOSSLED, JUXING). Amenta, Marelli, and Crepaldi (2015) also found inverse stem frequency effects on first fixation durations when words were read in opaque sentence contexts ("His efforts were FRUITLESS") relative to transparent sentence contexts ("The tree was FRUITLESS"). In both cases, the contexts that led to inverse frequency effects were ones in which information from the stem needed to be overridden or disregarded in order to accomplish the task at hand. Taft (2004) interpreted these effects as indication of an obligatory morphological decomposition mechanism, rather than a dual-pathway model in which words are processed either via their decomposed or surface form. Amenta et al. (2015) made a similar proposal, and furthermore posited that morpheme semantics are accessed as part of initial combinatorial processing. Within a distributed framework, mechanisms that are optimized for typical situations (e.g., stem meaning access facilitating sentence comprehension or lexical decision) can be suboptimal, and thus inhibitory, in atypical situations. These effects are stronger for higher frequency stems because the impact of such stems on processing is stronger, as described above. Distributed models that could simulate task-prompted inverse frequency effects would need to move beyond static form-to-meaning mappings to incorporate higher level semantic processes.

Much of the controversy regarding how to interpret stem frequency effects centers around when these effects occur: prior to lexical access (as suggested by decomposition models) or after lexical access (as suggested by listing models). From a distributed perspective, the answer to this question might be best summarized as "a bit of both". Word and morpheme characteristics should determine the degree of morphologically-influenced processing occurring early on, but lateral and feedback activation that is sensitive to morphological factors occurs throughout processing as well.

The timing of frequency effects in complex word processing has been explored in experiments using masked primed lexical decision. In this paradigm, a word or nonword prime is presented extremely briefly (around 50 ms) following a mask (e.g., hash marks; #######) so as to be weakly processed but not consciously perceived. Following the masked prime and a brief delay, a second word is presented as a target for lexical decision. If participants decide on a target more quickly following a related prime than following a control prime, this strengthens the case for facilitation due to visual word processing, rather than due to conscious reasoning following recognition of the word.

Masked priming experiments in French do not detect stem frequency effects on the magnitude of morphological priming (Giraudo, Dal Maso, & Piccinin, 2016; Giraudo and Grainger, 2000). However, primes with high surface frequency do yield stronger morphological priming than primes with low surface frequency (Giraudo & Grainger, 2000). Although a comparable study in English found no such difference (McCormick, Brysbaert, & Rastle, 2009), Amenta, Crepaldi, and Marelli, (2020) recently provided evidence that the effect in English may depend on orthographic-semantic consistency (see section below). In particular, Amenta et al. found that high-consistency targets showed a positive effect of prime frequency on masked priming magnitude, whereas low-consistency targets showed a negative effect. This suggests that, among morphologically related words that share more semantically useful morphemes, the effects reported by Giraudo and Grainger (2000) can be found in English (which is less morphologically rich than French). Overall, these results favor listing models over decomposition models. Additional neural network simulations would be beneficial to determine whether a positive effect of prime frequency on masked morphological priming is also amenable to the distributed perspective.

Family size and entropy

In addition to a morpheme's frequency, the number of different words in which it appears also impacts complex word processing. Variations in *morphological family size*, or the number of compound and derived words sharing a particular stem, predict lexical decision latencies: larger families lead to faster responses (Bertram, Baayen, & Schreuder, 2000; Schreuder & Baayen, 1997). Family size effects are pervasive across languages (e.g., in Dutch: Schreuder & Baayen, 1997; in German: Lüdeling & De Jong, 2002; in Finnish: Moscoso Del Prado Martín, Bertram, Häikiö, Schreuder, & Baayen, 2004; in Hebrew: Moscoso del Prado Martın, 2003).

Schreuder and Baayen (1997) suggested that family size might be the underlying factor driving stem frequency effects, since more frequent stems are also likely to have larger morphological families. In the context of a factorial design with Dutch words, De Jong, Schreuder, and Baayen, (2000) did not find an effect of family frequency (frequencies of morphological relatives minus the frequency of the standalone stem) on lexical decision response latencies when family size is controlled, but did find an effect of family size when family frequency is controlled. However, Ford, Davis, and Marslen-Wilson, (2010) demonstrated with a correlational design that stem frequency and morphological family size independently facilitate lexical decision responses in English, with family size effects being somewhat weaker. De Jong, Feldman, Schreuder, Pastizzo, and Baayen (2002) found similar results for lexical decision with both Dutch and English compounds: family size of the left constituent (e.g., BANK in BANKROLL) predicts response latencies separately from and more weakly than family frequency. De Jong et al. (2000) found that family size estimates are stronger predictors when irregular forms, such as TAUGHT for the stem TEACH, are included; likewise, these estimates improve when semantically opaque forms, like WITNESS from WIT, are excluded (Bertram et al., 2000; Schreuder & Baayen, 1997). Thus, whereas stem frequency could be a more orthographically-driven effect, stem family size may relate more to semantic processing.

Complex word recognition also appears to be impacted by its *secondary family size*: that is, the number of words that share a morpheme with any word in the primary morphological family. For example, TROLLEY is compounded in TROLLEY CAR, TROLLEYBUS, and TEA-TROLLEY; CAR, BUS and TEA occur in 16, 3 and 25 compound words, respectively, so the secondary family size of TROLLEY is 44 (Baayen, 2010). Secondary family size has a slightly inhibitory effect on lexical decision and word naming when the semantically dominant constituent has a small family size and the compound is not generally strongly connected to other compounds (Baayen, 2010; Mulder, Dijkstra, Schreuder, & Baayen, 2014). This is presumably because the activation of information corresponding to secondary family members via spreading activation leads to semantic processes unrelated to that of the target word, which can be inhibitory if not overwhelmed by support from direct relatives.

Family size effects have also been observed cross-linguistically: Moscoso Del Prado Martín et al. (2004) and Moscoso del Prado Martín et al. (2005) found that lexical decision latencies for a word in a language known to participants could be predicted by the morphological family size of the word's translation in a second language that is *unknown to them*. This "isomorphism" effect was found between Finnish and Dutch, and between Dutch and Hebrew, both pairings of very different morphological systems (and, in the latter case, of unrelated etymology). This suggests that morphological family size may be, at least partly, a proxy for richness of the word's semantic domain: larger morphological families are more likely in rich semantic domains, shared across languages and cultures. Thus, a word that happens to have a small morphological family in English yet falls in a semantically rich domain, would be likely to have larger morphological family sizes in other languages and also be responded to more quickly. Modern distributional semantics methods could be useful in testing this explanation (see Amenta, Günther, & Marelli, 2020).

In addition to the size of a morphological family, the relative frequencies of words within that family also impact recognition latencies. *Entropy*, a concept from information theory, describes the distribution of usage over the forms of a word and is lower if certain words within the family occur much more frequently than others. For example, the inflectional entropy of ANT is lower than that of WASP, because ANTS appears more frequently than ANT, while WASPS and WASP are used with relatively equal frequency (Baayen, Piepenbrock, & Gulikers, 1996). Higher entropy is correlated with faster lexical decision reaction times in Dutch (Moscoso Del Prado Martín, Kostic, & Baayen, 2004) and in English (Baayen, Feldman, & Schreuder, 2006). Additionally, Milin, Filipovic Durdevic, and Moscoso del Prado Martín (2009) found in Serbian that response latencies are slower for base words with atypical distributions of usage frequency across inflectional forms. In other words, the typicality of a morpheme's inflection usage profile relative to other nouns or other verbs facilitated recognition of a word containing it. This relative entropy effect in Serbian was replicated in a sentence reading task, suggesting that it is robust and relevant in more natural reading contexts (Baayen et al., 2011).

Family size and entropy effects favor a view of morphological processing in which a morpheme is not represented independently, but rather the other words containing it are, to a lesser degree, involved in the given word's processing. This would explain why the number of words in which a morpheme appears, and the frequencies of these words relative to each other, play a role in predicting morphological sensitivity after accounting for morpheme frequency. Such an interpretation aligns most closely with listing and distributed models. However, the nature of secondary family size effects, cross-linguistic family effects, and the contribution of irregular words to family size suggests that some such effects may come into play primarily at the level of semantic processing. If so, understanding the range of known family size and entropy effects may require a more advanced treatment of semantics than those provided in current form-to-meaning models.

That said, the only explicit accounts of family size and entropy phenomena are grounded in distributed processing mechanisms. Moscoso del Prado Martín, Ernestus, and Baayen (2004) trained a neural network to produce past-tense forms from present tense for almost 3,000 Dutch verbs. The number of similar words with a particular ending (akin to family size, for suffixes) predicted how likely their model was to choose that ending for a novel word, regardless of those words' frequencies. Baayen et al. (2011) trained the Naive Discriminative Model to map from orthographic to semantic representations for 3,003 simple and derived English words, and found that stem frequency, morphological family size, affix family size and surface frequency each explained independent variance in the model's response latencies. Frequency, family size, and inflectional entropy explained independent variance after training on 2,314 simple and inflected English words, as did constituent family size and frequency, surface frequency, and compound entropy after training on 921 compound words. Baayen et al. (2011) also found weak evidence of secondary family size effects in their compound word demonstration. Together, these simulations demonstrate that morpheme frequency, family size, and entropy effects can all be explained by a single distributed processing mechanism, "without any explicit parsing process being involved" (Baayen et al., 2011, p. 49).

Semantic consistency

Family size effects have been shown to be particularly driven by the family members for which the stem makes a meaningful contribution (Bertram et al., 2000; Schreuder & Baayen, 1997). For example, when counting the family members of WIT, excluding semantically unrelated words such as WITNESS yields a stronger predictor of lexical decision latencies. This suggests that it is not simply the number of words containing the stem, but the number of words using the stem in a semantically consistent manner, that best predicts morphological facilitation.

Along similar lines, Marelli, Amenta, and Crepaldi (2015) devised a metric of stems' *orthographic-semantic consistency* (OSC), calculated as the frequency-weighted mean cosine similarity between the stem's meaning as a standalone word and the meanings of all words containing it. For example, WHISK has a lower OSC than CHEER, because words containing WHISK (e.g., WHISKING, WHISKEY, WHISKER) are overall less semantically related to WHISK than words containing CHEER (CHEERING, CHEERFUL, CHEERY) are to CHEER. OSC captures the degree to which a simple word is semantically similar to any words containing it, making it a better metric of the simple word's consistency than of the morpheme's consistency. However, OSC explains additional variance in morphological priming magnitudes from masked primed lexical decision experiments after controlling for target family size, orthographic neighborhood, length, and frequency (Amenta et al., 2020). Additionally, the relationship between prime frequency and priming magnitude is positive for high-consistency targets, but negative for low-consistency targets. This suggests that OSC influences the nature of priming that is occurring: morphological facilitation in the context of high-frequency primes with high-OSC stems, and orthographic indifference or slight inhibition in the context of high-frequency primes with low-OSC stems. Variance in unprimed lexical decision reaction times for standalone stems can also be partially explained by their OSC (Marelli et al., 2015; Marelli & Amenta 2018; also see Siegelman et al., 2022).

Although we are not aware of any neural-network simulations of OSC effects on morphological processing, the existence of such effects follows most naturally from a distributed account. If sensitivity to morphological structure emerges during the process of learning to map from written words to their meanings, semantically consistent morphemes would be expected to influence learned weights more strongly than their less-consistent counterparts. Put another way, the presence of a consistent morpheme is more likely to be useful for the task of activating meaning than the presence of an inconsistent morpheme, and thus sensitivity to such morphemes should manifest earlier and more strongly.

The role of orthographic-semantic consistency is less clear in models that rely on decomposition as a primary mechanism. As discussed in Dawson et al. (2021b), more-consistent morphemes may have better consolidated and thus presumably more easily activated representations. However, the details of how less- versus more-consolidated representations differ, and the manner in which such differences impact the magnitude of morphological effects, are not well-specified.

Morphological processing in pseudo-complex nonwords

Another source of evidence on the mechanisms of complex word recognition involves the processing of nonwords (i.e., pronounceable strings of letters that are not real words). Analyses of lexical decision tasks typically focus on how quickly participants can confirm a word's lexical status. Studies that are focused on nonwords instead measure how long it takes participants to reject the target. Taft and Forster's seminal study using this approach in (1975) compared correct rejection latencies for bound stems from prefixed words (e.g., JUVENATE from REJUVENATE) with those from non-prefixed words (PERTOIRE from REPERTOIRE). Rejection latencies were slower for stems from prefixed words, suggesting greater sensitivity to the presence of stems that appear in morphologically informative contexts. Following up on this work, Taft and Forster (1976) demonstrated that participants reject polysyllabic nonwords more slowly when the first syllable is a real word (in other words, FOOTMILGE is rejected more slowly than MOWDFLISK). However, the same effect is not found for the second syllable (TROWBREAK and MOWDFLISK are rejected at comparable speeds). These findings imply the existence of position-weighted sensitivity to embedded strings in a novel word.

Subsequent nonword lexical decision studies have typically investigated morphological processing using nonwords constructed from novel combinations of real morphemes. Italian nonwords composed of a stem and a suffix (e.g., CANTEVI, analogous to a nonword like BUYED in English) show the slowest rejection times relative to a non-stem with a suffix, a stem with a non-suffix, or a non-stem with a non-suffix (Caramazza, Laudanna, & Romani, 1988). Similar delayed latencies for pseudo-suffixed nonwords have been found in Finnish (Leinonen et al., 2009) and English (Crepaldi, Rastle, & Davis, 2010). The presence of morphological effects despite the lack of established whole-word meanings suggests that morphological sensitivity is present prior to or without lexical access, supporting a decomposition view.

However, Günther, Marelli, and Bölte (2020) demonstrated that at least when the nonword is presented overtly, the presence of independent morphemes is not the sole source of interference. Using compositional distributional semantic models (Marelli, Gagné, & Spalding, 2017), they quantified the semantically interpretability of morphological combinations, and found that this metric had a significant, inhibitory effect on how quickly novel German compounds composed of two real words were rejected. This finding illustrates that although morphological sensitivity does not require access to lexical representations, it does not stop at accessing constituents' independent meanings. Instead, the degree to which that morpheme *combination* is meaningful impacts how difficult it is to reject, suggesting that even in pseudo-complex nonwords, morphological information is processed with respect to its context as would be predicted by a distributed account. Semantic representations in published distributed models of morphological processing are too simplistic to capture such nuanced effects.

One challenge for evidence from unprimed lexical decision studies is that nonwords may be processed in a distinct manner from real words (as emphasized by Burani et al., 1984). If so, investigating effects for pseudo-complex nonwords in a task that encourages identification of lexical status may not provide the most relevant insights into morphological processing mechanisms.

Longtin and Meunier (2005) used the masked primed lexical decision paradigm to look more closely at the time-courses for word and nonword processing in French. Lexical decisions were made on real, simple target words (RAPIDE) following briefly presented suffixed real words (RAPIDEMENT, similar to RAPIDLY in English), pseudo-suffixed nonwords (RAPIDIFIER, like RAPIDIFY), and non-suffixed nonwords (RAPIDUIT, like RAPIDEL), relative to unrelated controls. Note that participants never made a lexical decision on, or even consciously perceived, the nonwords of interest, mitigating concerns regarding distinct nonword processes. Longtin and Meunier (2005) found that pseudo-suffixed nonwords primed their stems as strongly as suffixed real words, but found no significant priming by non-suffixed nonwords. Beyersmann, Cavalli, Casalis, and Colé, (2016) also found equivalent morphological masked priming with word and nonword primes, and extended the finding to include prefixed and pseudo-prefixed primes. Masked priming experiments in English also show significant priming by pseudo-complex nonwords, similar in magnitude to that by complex words, for both pseudo-suffixed (McCormick et al., 2009) and pseudo-compound (Fiorentino, Politzer-Ahles, Pak, Martínez-García, & Coughlin, 2015) primes.

Contrary to the initial findings by Longtin and Meunier (2005), Beyersmann and colleagues (Beyersmann et al., 2016; Beyersmann & Grainger, 2018; Grainger & Beyersmann, 2020) found significant non-suffixed nonword priming (RAPIDUIT) under certain conditions (also see Morris, Porter, Grainger, & Holcomb, 2011; De Rosa & Crepaldi, 2021). Beyersmann et al. (2016) showed that non-suffixed nonword priming is more strongly related to the participant's word reading proficiency than other priming effects, possibly because it is a weaker phenomenon that only emerges with a great deal of word recognition practice. In Beyersmann and Grainger (2018), such effects were modulated by the family size of the stem. Grainger and Beyersmann (2020) demonstrated that non-suffixed nonword priming is stronger for stems that are more frequently followed by a derivational suffix. Thus, although the pairing of a stem and a suffix best primes the stem's recognition in nonword priming contexts, a stem followed by non-suffix letters can also be somewhat facilitatory, depending on characteristics of the stem and the reader.

The pattern of nonword priming effects varies between masked and overt priming paradigms. For example, suffixed nonword masked priming appears equally strong regardless of whether the stem-suffix was compatible (e.g., SPORTLESS, where -LESS can follow nouns) or incompatible (SPORTATION, where -ATION usually modifies verbs; Longtin & Meunier, 2005). This is not the case with similar stimuli in a cross-modal priming context: an auditorily presented suffixed nonword prime significantly primes recognition of its written stem only if the stem-suffix combination is compatible (Meunier & Longtin, 2007). In the case of compound words, novel compounds (DRUGRACK) and novel pseudo-embedded words (SLEGRACK) are similarly strong primes in a masked priming paradigm, but novel compounds facilitate significantly more strongly when priming is overt (Fiorentino et al., 2015). Contrasting results in overt and masked priming contexts demonstrate a key trend in morphological processing studies: word-level semantic properties, such as stem-suffix compatibility and lexical status, appear to matter more in contexts where the word is presented for a longer duration and is consciously perceived. This phenomenon is also found in investigations of semantic transparency effects, discussed below.

Most results from studies using nonwords align well with decomposition models. In particular, masked priming results with pseudo-suffixed nonwords demonstrate that, for at least a brief period of time, their componential structure facilitates subsequent word processing to a degree that is comparable with real suffixed words, despite having no previous surface form exposure or established meaning.

The presence of weaker non-suffixed nonword priming effects, moderated by stem derivability, stem family size and reader proficiency, tips the evidence scale somewhat away from certain models within the decomposition view. If decomposition is triggered in instances of morphological structure but not in the case of embedded words, as suggested by Rastle, Davis, and New (2004) and Longtin, Segui, and Hallé, (2003), priming by non-suffixed nonwords must be explained by a separate mechanism. Beyersmann and Grainger (2018) suggested that this type of priming proves that embedded stems can be extracted without the need for morpho-orthographic decomposition or affix splicing, by means of recognition and activation of edge-aligned words and their corresponding morphological representations. When edge-aligned words appear in real words such as CASHEW, they clarified, activation of the full word inhibits the embedded word and prevents this type of priming from occurring.

Evidence from morphological processing studies using nonwords is also compatible with the distributed account. Neural networks learn to make use of componential information as much as possible, while maintaining good performance in cases where such information is not useful (Plaut & Gonnerman, 2000). Thus, the processing of novel morpheme combinations in a manner that hinders their rejection as words and facilitates recognition of related words is plausible within this framework. Facilitation by nonsuffixed nonword primes can be explained similarly. As an added benefit, meaning conveyed by sub-lexical orthographic features that might not even be regarded as morphological (e.g., see Hendrix & Sun, 2021) also has the opportunity to impact nonword processing within a distributed account, as information at multiple grain-sizes may play simultaneous roles. However, neural network simulations of known effects with pseudo-complex nonwords have not yet been conducted.

Semantic transparency

Semantic transparency refers to how directly a complex word's meaning can be derived from its morphological structure. Transparent words include DARKNESS and UNHAPPY because the meanings of these words can easily be inferred from their constituent morphemes. In contrast, as the meanings of RELEASE and DEPARTMENT are unrelated to those of LEASE and DEPART, respectively, these items are semantically opaque.[3] As pointed out by Marslen-Wilson, Tyler, Waksler, and Older (1994), semantic transparency is an important lexical-level characteristic for morphological processing in English: in a cross-modally primed lexical decision task, auditorily-presented complex primes facilitate lexical decision for their stems only if the primes are semantically transparent. For example, hearing "punishment" primes recognition of PUNISH but hearing "casualty" does not prime recognition of CASUAL. Transparent priming is also stronger than opaque priming when primes are overt (not masked) and presented visually (Feldman & Soltano, 1999; Rastle, Davis, Marslen-Wilson, & Tyler, 2000; Rueckl & Aicher, 2008).

When visual prime duration is brief and the word is not consciously perceived (i.e., 35–60 ms masked priming; see Forster & Davis, 1984), semantically opaque prime-target pairs exhibit significant priming. Results originally found by Longtin et al. (2003) and Rastle et al. (2004) suggest that opaque primes facilitate recognition of their pseudo-stems (CORNER-CORN) as much as transparent primes facilitate recognition of their actual stems (TEACHER-TEACH), and significantly more than orthographic primes (BROTHEL-BROTH). A subsequent meta-analysis conducted by Rastle and Davis (2008) found that this pattern of results held across 19 masked morphological priming studies in Indo-European languages. Researchers concluded from these findings that word forms with apparent morphological structure are treated similarly to actual complex words during initial visual processing, supporting morphological decomposition as an early word processing phase prior to recognition. In other words, interpreting the opaque masked priming effect as evidence that CORN is represented briefly when the word CORNER is shown implies the existence of a semantics-blind morphological decomposition mechanism (Rastle & Davis, 2008). Similar patterns of effects have been shown in English with derived-derived priming (Feldman & Soltano, 1999), stem-derived priming (Marslen-Wilson, Bozic, & Randall, 2008), and compound word priming (Fiorentino & Fund-Reznicek, 2009).

As Rastle et al. (2004) found in English, readers in French and Serbian show equivalent morphological priming with transparent and opaque primes when the prime is masked, but transparent priming is stronger when the prime is overt (Feldman, Barac-Cikoja, & Kostic, 2002; Longtin et al., 2003). Complex Chinese words with a common morpheme that makes either a similar semantic contribution (e.g., 公 in 公园, "public park", and 公众, "the public") or a distinct semantic contribution (公 in 公园, "public park" and 公鸡, "rooster") showed equivalently strong facilitation in masked, but not overt, priming contexts (Tsang & Chen, 2013). Stem homographs in romance languages, such as CERRO ("hill") and CERRAR ("to close") in Spanish, appear to be forms of the same stem, but they are semantically and morphologically unrelated. Such words inhibit each others' recognition relative to merely orthographically related words (CERDO, "pig") when the prime is overt (Laudanna, Badecker, & Caramazza, 1989), but are facilitatory relative to orthographic when the prime is masked (Badecker & Allen, 2002). Masked priming studies in Dutch and Bangla found opaque priming to be greater than orthographic, but significantly weaker than transparent priming, although these results may be attributable to slightly longer masked prime durations (Dasgupta, Sinha, & Basu, 2015; Diependaele, Sandra, & Grainger, 2009). In the cases of Hebrew (Frost, Deutsch, Gilboa, Tannenbaum, & Marslen-Wilson, 2000), Arabic (Boudelaa & Marslen-Wilson, 2001), and German (Smolka, Komlósi, & Rösler, 2009), facilitatory opaque morphological priming continues to be strong even when the prime is overt.[4] Cross-linguistic differences in semantic transparency effects will be discussed in more depth later in this review. Generally speaking, studies in languages other than English have also found significant evidence of facilitatory opaque priming beyond what would be expected from orthographic similarity, particularly in a masked priming context. The remainder of this section focuses on English experiments, as a great deal of the discussion regarding opaque priming is grounded in these results.

Initial results in English (Rastle et al., 2004) suggested that facilitation due to opaque morphological priming is not only greater than orthographic priming, but equivalent in magnitude to that of transparent priming. Some subsequent studies have found similar results: for example, Marslen-Wilson et al. (2008) found no significant difference in facilitation by transparent, quasi-transparent, and opaque priming, and this held for prime durations of 36, 48 and 72 ms. Lavric, Clapp, and Rastle (2007) and Lavric, Elchlepp, and Rastle (2012) found ERP evidence for similar initial treatment of opaque and transparent stimuli. However, some studies have found effects from opaque primes to be significantly weaker than those from transparent primes.

[3]The term *semantically opaque* typically refers to complex words that were historically morphologically related to their stems, although modern usage does not make this relationship clear (e.g., COURTEOUS). In contrast, *pseudomorphological* words only appear to be morphologically structured and have no historical morphological relatedness (e.g., CORNER was never morphologically related to CORN). In this review we treat opaque and pseudomorphological words as one category, referred to as opaque, for simplicity.

[4]Semantically opaque morphological priming by unmasked primes has been found in English (Libben, Gibson, Yoon, & Sandra, 2003) and Dutch (Zwitserlood, 2022), in the context of constituent priming during a lexical decision task with compound words. Thus, the type of complex stimuli and task in question may also impact whether opaque morphological priming effects are observed within a language.

Morris, Grainger, and Holcomb (2008) observed stronger effects, both behaviorally and with ERP measures, for transparent primes relative to opaque (next strongest) and orthographic (weakest). Similarly, Diependaele, Duñabeitia, Morris, and Keuleers, (2011), Andrews and Lo (2013) and Feldman, Milin, Cho, Moscoso del Prado Martín, and O'Connor (2015) all found intermediate effects of opaque priming relative to transparent and orthographic priming. Such differences may be found inconsistently in part because they are small: in the meta-analysis conducted by Rastle and Davis (2008), 8 of the 12 studies in English that included a transparent condition showed numerically stronger priming for transparent than for opaque conditions, suggesting that perhaps this effect is present but not easily detected statistically.

Similar effect sizes for opaque and transparent priming may also be due to confounding factors across the prime-target pairs in these conditions. Feldman, O'Connor, and del Prado Martín (2009), finding significantly stronger priming for transparent than opaque pairs when affixes are matched across conditions, posited that lack of accounting for affixes during stimulus selection may contribute to confusing results. These authors suggested that, in multiple studies, a disproportionate use of affixes with low productivity (-ILE) or an inconsistent semantic role (-ER; as in GREATER versus TEACHER) in the transparent condition may have decreased those words' priming, making them more comparable to opaque priming magnitudes.[5] Another potential confound in studies of transparency effects in lexical decision was pointed out by Marelli et al. (2015), relating to the orthographic-semantic consistency (OSC) of target words. They demonstrated that the targets used in several previous masked priming studies of semantic transparency differ significantly in OSC across conditions, and argued that this drives commonly observed faster response times to transparent than to opaque targets when preceded by unrelated primes. In other words, lower-OSC words like WHISK, often used in opaque conditions following WHISKER, are responded to more slowly than higher-OSC transparent targets like CHEER (primed by CHEERFUL), regardless of what prime precedes them. It is possible that, because the targets in transparent conditions are already being recognized quickly, the degree to which their recognition could be further facilitated is reduced, constraining transparent priming to a level more comparable with opaque priming.

Jared et al. (2017) provided perhaps the most informative recent empirical contribution to the debate over early transparency effects in English. Using the same masked primed lexical decision paradigm implemented in many prior studies, they consistently found graded effects of semantic transparency across six experiments, using both behavioral and ERP measures. In their design, each condition consisted of substantially more prime-target pairs than were used in prior studies investigating semantic transparency (104, compared to the typical 40 to 50), and a quasi-transparent condition (e.g., DRESSER-DRESS, BOOKISH-BOOK) was included to examine intermediate transparency effects. Two notable discrepancies across previous studies' methods—relatedness of primes for nonword distractors and masked prime duration—were addressed and tested. Neither of these factors changed the pattern of results: for both ERP and behavioral results, transparent masked primes showed significant facilitation across all experiments, and opaque effects were weak and never significantly stronger than the orthographic control. Quasi-transparent effects fell between those of transparent and opaque conditions in magnitude. Jared et al. (2017) also replicated ERP evidence of graded transparency effects with a masked primed *semantic* decision task ("Is this a sea creature or not?"), and demonstrated that color boundaries between morphemes yielded a boost in ERP signals only for the transparent and quasi-transparent conditions.[6] Jared et al. (2017) not only repeatedly demonstrates the effect of semantic transparency on masked morphological priming magnitudes, but also suggests that opaque priming in English may be notably weaker than previously suspected, or possibly nonexistent. Similar results were found recently by Chee and Yap (2022); for both semantic categorization and lexical decision tasks, opaque masked priming was found to be no stronger than orthographic priming, and significantly weaker than transparent priming.

Despite some counterexamples and potential confounds, opaque priming has been found in a wide range of studies in English. Thus for the subsequent discussion of these effects with respect to decomposition and distributed theories, we will presume that opaque morphological facilitation in English masked priming studies is significantly greater than orthographic priming, though weaker than transparent morphological priming.

In decomposition theories of morphological processing, semantically opaque priming is interpreted as evidence that

[5]However, as pointed out by Davis and Rastle (2010), the results of Feldman et al. (2009) are also in question due to a higher incidence of non-morphological orthographic transformations from prime to target (e.g., BLISTERY-BLISS and COYNESS-COIN) in the opaque condition than in the transparent condition.

[6]In the *illusory conjunction* paradigm, a word is presented partly in one color and partly in another: the boundary between the colors either aligns with a boundary between components of proposed psychological import (e.g., morphemes) or not. If an advantage in recognition is observed when boundaries align, the components are thought to be cognitively represented (see Rapp, 1992, for another example of this paradigm being used to investigate complex word processing).

early in visual word recognition, morphological decomposition is applied to all words that have the appearance of being complex (Rastle & Davis, 2008). Furthermore, these effects have been cited as evidence against the distributed view, characterizing distributed models as only being able to explain morphological effects for words that are semantically transparent (e.g., Beyersmann, Coltheart, & Castles, 2012; Duñabeitia, Perea, & Carreiras, 2007; Rastle & Davis, 2008). However, simulations relying on distributed mechanisms do provide explanations of opaque effects. Using the Naive Discriminative Model, Baayen et al. (2011) simulated transparent and opaque priming effects that were not significantly different, using the same stimuli employed in Rastle et al. (2004). However, this result relied on how these items were represented when input to the model: Baayen et al. (2011) argued that for several of the opaque primes used in Rastle et al. (2004), the suffix plays a functional role in the words' meaning (e.g., -ER in ARCHER denotes "one who", with ARCH etymologically referring to a bow). The orthographic inputs of prime words for which Baayen et al. (2011) judged this to be a concern were thus associated with the meaning of the whole word (ARCHER) and that of the suffix (-ER), and those words drove the strong facilitation observed in priming simulations for the opaque condition. Although the transparency of certain prime affixes may strengthen the opaque priming magnitudes found in some studies, Beyersmann et al. (2016) showed that this argument was insufficient to entirely account for opaque priming. Using a much more tightly controlled set of opaque prime-target pairs for which no affix or stem could be argued to be contributing semantically to the surface meaning, they still found opaque priming to be significantly greater than orthographic priming, and not significantly different than transparent priming.

Plaut and Gonnerman (2000) provided an alternative distributed explanation of why opaque priming occurs. After a neural network was trained on a morphologically rich artificial language (likened to Hebrew), it reached stable semantic representations for target inputs faster when preceded by an opaque morphological prime than an orthographic prime—in other words, opaque priming was observed to be significantly stronger than orthographic priming. The network's learned weights that handle opaque words' inputs were more strongly influenced by the input's components due to the strong role of morphological structure present in the rest of the language. Sensitivity was not just acquired for particular stems in this case, but for morphological structure in word forms more generally. For the morphologically impoverished lanugage, in which no morphological structure existed save in the testing words shared with the rich language, Plaut and Gonnerman (2000) did not find significant opaque priming. In this case, learned processing of opaque words was less affected by their constituent morphemes, as morphology had less impact on the overall learning of the language and thus on the manner in which word forms, regardless of their transparency, were processed.

In retrospect, considering an artificial language almost entirely lacking in morphological structure to be analogous to English was an exaggeration, and thus the fact that the impoverished artificial language did not show opaque priming does not preclude an account of subsequent findings of opaque priming in English (Rastle et al., 2004). Presumably, opaque priming can be expected to occur with a degree of strength or robustness corresponding to the language's morphological richness, and English may be sufficiently morphologically rich to account for opaque priming that has been observed in masked priming contexts. If this is the case, Plaut and Gonnerman (2000) does provide an explanation of opaque priming, even in English, via distributed processing, and additionally explains the graded transparency effects found by Jared et al. (2017) and others.

The fact that opaque priming is often only observed in masked priming contexts, not when the prime is overt, is sometimes interpreted as evidence that the effect occurs during faster, earlier phases of word processing; however, this interpretation implies that processing is halted the moment the stimulus changes, which does not align with a general understanding of visual processing. As argued by Tzur and Frost (2007, p. 323), masked priming duration might be better interpreted as "the amount of energy that is provided to the cognitive system for the perception and identification of the stimulus". If word presentation initiates a wave of propagated activation, the duration of the prime could determine the strength of that activation. Consistent with this view, Tzur and Frost (2007) demonstrated that both the luminance of the prime and the amount of contrast between the prime and the background yield variability in effects comparable to that produced by manipulating prime duration. For example, a 20 ms prime with brighter luminance showed similar effects on lexical decision latencies as a 40 ms prime presented more dimly. From this perspective, graded semantic transparency effects as shown by Jared et al. (2017) and others could be interpreted, not as semantic, word-level processing happening "early", but as weak lexical input being sufficient to activate semantic information to some extent. Likewise, opaque effects during masked but not opaque priming in English can be explained as the weak influence of the lanuage's overall morphological structure on how the word form is processed, an influence which is overwhelmed when the signal is strong enough to more thoroughly activate semantic and surface-form level information.

To be clear, under this interpretation of masked priming effects, opaque and graded transparency effects in masked morphological priming are not inconsistent with morpho-orthographic decomposition. It could be that words with apparent morphological structure (pseudo or actual) are in fact split along morphemic boundaries regardless of those boundaries' semantic significance, but that this purely morpho-orthographic phase is too brief to be detectable in the absence of semantic feedback effects. This claim would be difficult to falsify. The distributed account, however, provides an alternative explanation: that simultaneous influence of both word-level and morpheme-level characteristics occurs even during early morphological processing, with the relative weighting of each dependent on the word, morpheme, and language in question. Thus, instead of providing evidence against the distributed view, opaque morphological priming can be explained by either perspective, while evidence of graded transparency effects in even masked priming contexts is more simply accounted for by the distributed view.

As a final note regarding semantic transparency, in favoring a view of morphological processing that does not regard morphemes as functional units, the distributed view provides an alternative means of quantifying and conceptualizing semantic transparency. Instead of viewing the contributions of constituent morphemes, particularly affixes, in terms of a linear addition (e.g., the meaning of TEACHER = the meaning of TEACH + the meaning of -ER), affixes' semantic role can be conceptualized as functions which transform the meaning of the stem in typically predictable ways (Marelli & Baroni, 2015). Once these functions are learned for each affix, compositional meanings for derived words and novel words can be constructed to depict what the word's morphology conveys about its meaning (e.g., the compositional meaning of SUMMER is "one who sums"). Furthermore, the similarity between the word's actual meaning, represented via traditional distributional semantics techniques (e.g., Landauer & Dumais, 1997), and its compositional meaning can be construed as a new measure of semantic transparency (Marelli & Baroni, 2015). Using compositional semantics techniques on compound words, Günther and Marelli (2018) demonstrated that a higher-dimensional treatment of semantic transparency yields better predictors of lexical decision latencies (also see Günther et al., 2020). The construction and comparison of compositional meanings has promise as an improved approach to understanding the role of semantic processing in learning and recognizing complex words (Amenta et al., 2020; Baayen, Chuang, Shafaei-Bajestan, & Blevins, 2019). Such tools can also be applied to quantifying the overall transparency of an entire language's morphology (Günther, Smolka, & Marelli, 2019).

Task effects

The predominant task used across morphological processing studies is the lexical decision task. The use of a standard task allows for ease of comparison across many studies' results, an advantage exploited in this paper and other reviews. However, variation in task is essential for determining what aspects of morphological processing are unique to, or at least more common for, a specific paradigm. Marelli, Amenta, Morone, and Crepaldi (2013) and Duñabeitia, Kinoshita, Carreiras, and Norris, (2011) both demonstrated, with tasks emphasizing more semantic and orthographic aspects of word processing, respectively, that morphological effects are quite sensitive to the goal of the reader.

Marelli et al. (2013) used a variation on the masked priming paradigm in which, following a morphologically transparent, morphologically opaque, or orthographic prime, a word is presented where the reader is fixating. Simultaneously, a number is shown elsewhere on the screen. The participant is motivated to recognize the presented word as quickly as possible so as to have time to look at the number. After some trials, a prompt appears asking about the most recent word or number (e.g., "Is it a tool?" or "Is it even?"), requiring that the participant access and retain semantic information about the stimuli. In this context, transparent priming significantly shortens word fixation duration, but opaque and orthographic priming do not. However, opaque priming is detected in a lexical decision paradigm with the same stimuli (Marelli et al., 2013).

Duñabeitia et al. (2011) investigated both transposed-letter and semantic transparency effects in the context of "same-different" judgments: participants are presented with two words (a reference and a target) and are asked to determine if they are the same or different. In the context of a lexical decision task, Duñabeitia et al. (2007) had found that masked primes with two transposed letters (e.g., AMNOG from AMONG) prime the original word more strongly than do masked primes with two replaced letters (AMELG from AMONG), but that this effect goes away when the transposed letters cross a morphological boundary (WALEKR). In the context of a same-different task, however, the transposed-letter effect does not weaken when crossing a morphological boundary (Duñabeitia et al., 2011). In a second study, same-different judgments were paired with transparent, opaque and orthographic masked primes, comparable to Rastle et al. (2004) and Jared et al. (2017). Same-different judgments yielded similarly strong priming across all three conditions, in contrast to the pattern of decreasing magnitudes found using lexical decision.

Marelli et al. (2013) and Duñabeitia et al. (2011) both demonstrate that sensitivity to morphological structure during word processing varies with the nature of the task. Critically, opaque and orthographic effects are undetectable when necessary information is semantic in nature ("Is it a tool?") whereas they are close in magnitude to transparent effects when orthographic information is needed ("Are these words the same?"). This suggests that the processes giving rise to decomposition phenomena are not applied indiscriminately, as stated in Rastle and Davis (2008), but are sensitive to the needs of the context (although some contexts can yield similar results; see Chee & Yap 2022). In a distributed account, these influences would arise via top-down feedback from higher-level task representations (e.g., Cohen, Dunbar, & McClelland, 1990; Gilbert & Shallice, 2002), although such effects have not been simulated in a morphological processing context. Task-varying effects could also be accommodated in a decomposition model—although perhaps somewhat less naturally—by allowing for alternative processing routes whose use varies with context. Kuperman et al. (2008) made a proposal along these lines to account for eye tracking data recorded while reading compounds (also see Kuperman, Schreuder, Bertram, & Baayen, 2009): in their model, multiple routes simultaneously processing different types of information (compound frequency, right constituent family size, etc.) interact with each other and impact behavior to the degree necessary for efficient recognition in the given context.

Orthographic flexibility and specificity

The orthographic constructions of complex words vary widely, both within and across languages. In some cases, the manner in which an orthographic form might be decomposed into its constituents is easy to imagine: In English, SINGER can be split into SING and ER. In Chinese, 公众 can be split into 公 and 众. Even in languages with nonconcatenative morphologies, such as Arabic and Hebrew, word forms can typically be separated into interleaving letter subsets corresponding to each morphological component (e.g., TIZMORET into root ZMR and pattern TI_O_ET; Frost et al., 1997). However, words for which morphemes are not easily split apart illuminate the role of orthographic form in morphological processing. Complex words that cannot be perfectly split into their morphological constituents are not merely exceptional cases, to be viewed in contrast to an easily-split norm. In the English language, for example, around 39% of complex words fall into this category (Baayen, Piepenbrock, & Gulikers, 1995). Studying such cases is essential for getting a complete picture of morphological processing.

It has been demonstrated in several experiments that morphological sensitivity is robust to lost, added, and replaced letters. For example, some derivations in Setswana are formed using letter replacement as well as concatenation (Kgolo & Eisenbeiss, 2015). MOREKI, "buyer", and THEKO, "a manner of buying" are derived from REKA, "to buy", and MORERI, "preacher", and THERO, "sermon", are derived from RERA, "to preach". Although it's not clear how these derivations would be divided into stem and affix if decomposed, these words still show significant morphological masked priming in lexical decision experiments (Kgolo & Eisenbeiss, 2015). In English, ADORABLE primes ADORE despite a missing -E-, LOVER primes LOVE despite a shared -E-, and DROPPER primes DROP despite the added -P- (McCormick, Rastle, & Davis, 2008). Importantly, semantically opaque words with similar transformations such as BADGER-BADGE and FETISH-FETE also show morphological priming despite not being perfectly segmentable into stems and affixes.

To accommodate orthographic flexibility in a decomposition view of morphological processing, McCormick et al. (2008) proposed that stem representations activated during the morpho-orthographic decomposition phase are underspecified (also see Taft, 1979). In other words, mutable letters like E at the end of LOVE are missing or optional in stem representations so as to be compatible with the decomposition of LOVING, whereas the M at the end of DRUM can be optionally duplicated to map onto DRUMMER and DRUMMING. Although this theory fits some English results nicely, it is challenged by more complicated morphological transformations such as the Setswana examples above. The orthographic transformation needed to generate THEKO from REKA and THERO from RERA would result in REKA and RERA's underspecified representations consisting of only -EK- and -ER-, respectively, which would likely risk confusing these stems with other linguistic entities and eliminate orthographic information (R-) that is useful for other derivations (e.g., MOREKI and MORERI). A theory reliant on underspecified stem representations would benefit from context-sensitivity when applied to Setswana, such that the flexible interchanging of a beginning R- with TH- is contingent upon the interchanging of an ending -A with -O.

The instances in which morphological operations (e.g., dropping -E and adding -ABLE) do *not* appear to generalize can also inform our understanding of morphological processing mechanisms. For example, idiosyncratic inflections, such as FELL from FALL, show masked priming effects, but word pairs which follow that orthographic transformation but are not morphologically related do not (e.g., BELL-BALL; Crepaldi et al., 2010). This finding has two implications: first, morphological masked priming for FELL and FALL can't be purely driven by the appearance of morphological relatedness (or the same effect would occur for BELL and BALL). Second, sensitivities to morphological transformations that occur only in one or very few contexts aren't learned to the extent that they impact opaque word processing. Perhaps if the past tenses of ENTHRALL, STALL, and APALL were ENTHRELL, STELL and APELL, a priming effect would be found between BELL and BALL.

Weak and defective roots in Hebrew provide another informative instance where morphological priming does not occur (Velan, Frost, Deutsch, & Plaut, 2005). Most derived forms of weak roots in Hebrew contain only two of the three formal root consonants, whereas derived forms of defective roots contain the third consonant sometimes, but not always. In a masked primed lexical decision paradigm, derivations of weak roots are facilitated by the two letters that appear across all forms, but not by the three letters of the formal root. Conversely, recognition of defective-root derivations are facilitated by both two- and three-letter primes, even if the target contains only two of the three root letters. These findings imply that morphological processing is sensitive to the variability of morphological family members' orthographic forms.

Customized orthographic flexibility is also seen in the context of position-specificity. Although morphologically structured nonwords have been found to be rejected more slowly than other nonwords, inverted suffixed words like NESSKIND are rejected as quickly as suffix-less forms like NELSKIND (Crepaldi et al., 2010), suggesting that suffixes are processed as morphologically significant only if they appear at the end of a word. Similarly, although BEGNESS is rejected more slowly than non-suffixed BEGNUSS, the inverted NESSBEG is not rejected more slowly than NUSSBEG (Crepaldi et al., 2010). Position-specific effects are also found for both prefixes and suffixes in Spanish nonwords (Carden, Barreyro, Segui, & Jaichenco, 2019). On the other hand, Taft (1985) showed that transposed English compound words (e.g., DROPRAIN) were rejected more slowly than other pseudo-compound nonwords in a lexical decision task, suggesting that morpheme representations may be less position-sensitive when the morphemes can occur in varying positions or as standalone words (e.g., DROP in RAINDROP, GUMDROP, DROPLET, DROPPER). Similarly, Duñabeitia, Laka, Perea, and Carreiras, (2009) found masked priming between Basque compound words whether the shared constituent was in the same position (LANORDU and LANPOSTU) or in a different position (SUMENDI and MENDIKATE). Thus, morphological processes are position-specific where that is sufficient (e.g., suffixation and prefixation) but can be more flexible for morphemes that appear in a variety of positions.

Overall, morphological sensitivity persists despite changes in a morpheme's appearance. Even opaque priming is robust to orthographic alterations in some cases, but only if such changes are common across derivations or for that particular morpheme. To account for orthographic flexibility of morphological processing effects, Crepaldi et al. (2010) proposed a variation of the decomposition view in which lemma representations are activated following morpho-orthographic representations. Inflected variations activate the same lemma, whereas derived forms have distinct lemmas due to their more varied semantic and syntactic roles. Within their model, irregular inflected priming, such as that seen for FELL and FALL but not BELL and BALL, can be attributed to FELL and FALL activating of the same lemma despite having different morpho-orthographic representations. The lemma modification explains the described results in English, but does not explain the uniquely flexible priming of defective roots in Hebrew or differences in position-specificity across morphemes. Carden et al. (2019) suggested that differences in position-specificity are due to inherent differences in affix and stem representations (see also Grainger and Beyersmann, 2017). However, the nature of those differences are underspecified. Additionally, it is unclear where prefixes that are more stem-like in nature (e.g., PSYCHO- and ANTHRO-) would fall within this framework.

In a distributed account, additional features or phases are not needed to account for the orthographic flexibility of morphological processing effects. More complex and abstract operations are learned by a neural network in contexts in which they are frequent and useful, such as dropping Es and doubling consonants in English. For simpler instances in which straightforward orthographic-semantic regularity is present (e.g., the single case of FELL and FALL, or among Hebrew words with a root that always contains all three consonants), this is all that is learned. The complex-when-necessary learning style of neural networks aligns well with the empirical results surrounding orthographic flexibility of morphological processing, while also providing a mechanistic account of how these effects might emerge with reading experience. Both Rumelhart and McClelland (1986) and Baayen et al. (2011) used neural network simulations to demonstrate irregularity effects in morphological processing, showing that these models can account for such phenomena without the need for separate routes. However, neither paper captures the mapping of orthographic representations to distributed meaning representations: Rumelhart and McClelland's 1986 network trained on production of the past-tense instead of semantic access, and Baayen et al.'s 2011 network used localized lemma representations as output. Additionally, both simulations only examined inflected words in isolation. More simulations using neural networks are necessary to complete the distributed account of orthographic flexibility in morphological processing. Simulating masked priming effects such as those presented by Crepaldi et al. (2010), Velan et al. (2005) and McCormick et al. (2008) would be particularly enlightening.

Difficulties with dividing complex words into morphological constituents have prompted several prominent linguists to reject the morpheme as a useful unit of meaning. Anderson (1992) emphasized the lack of isolable morphological constituents in contexts such as reduplication in Tagalog (e.g., MAGLALAKBAY from MAGLAKBAY and PAGBUBUKSAN from BUKSAN) and apophonic relations in English (SANG from SING and DOVE from DIVE). He argued for an approach to morphological structure focused on relational processes rather than components. For example, instead of viewing the A in SANG as an indicator of past, past is associated with the *process* of changing the I in SING to an A. Similarly, Blevins (2016) advocated for the "Word and Paradigm" view, stating that morphological structure provides information via patterns of change across words, not via separable contributing constituents. The difficulty of the morpheme unit with respect to explaining the nature of complex word forms has driven linguists as well as psychologists to pursue more flexible accounts.

Emergence and divergence of morphological processing mechanisms

As demonstrated in the sections above, morphological processing effects are found to be very sensitive to the morpheme's context and its general usage in the language in question. The pervasive importance of morpheme usage characteristics, from orthographic-semantic consistency to inflectional entropy, highlights the need for a better understanding of how the statistics of language drive the emerge of these mechanisms through experience. Broadening our scope to consider how morphological processing effects differ across languages further underscores this need: How does the visual system of a reader adapt to the distinct morphological systems of a certain language and develop appropriate processing mechanisms? Below, we review salient differences in morphological processing known to exist across languages, as well as existing evidence concerning morphological processing development in both young readers and second-language learners.

Differences across languages

Although cross-linguistic in its scope, to this point our review has focused on characteristics of morphological processing that are similar across languages; however, noting where processes differ can provide essential insights into how language characteristics shape these mechanisms.

As mentioned previously, in some languages semantically opaque morphological priming persists even when the prime is consciously perceived. Bentin and Feldman (1990) and Frost et al. (2000) showed in Hebrew that when morphologically related primes were presented overtly, lexical decisions for visual targets were significantly facilitated even if the prime wasn't semantically related to the target. In fact, opaque priming wasn't even significantly weaker than transparent priming in this context. Similar results have been found in Arabic (Boudelaa & Marslen-Wilson, 2001). If the effect were found only in these two languages, one might attribute its emergence to a unique characteristic of the Semitic language family, such as embedded morphology; however, German participants also show strong opaque priming by an overt visual prime (Smolka, Gondan, & Rösler, 2015).

The language characteristics driving such differences in semantic transparency effects have been discussed and assessed in some depth. An idea previously put forward by Plaut and Gonnerman (2000) attributed overt opaque priming in certain languages to the greater morphological systematicity of those languages. In other words, languages in which a word's meaning can usually be inferred directly from its morphological structure might show greater opaque morphological priming, because broadly applied morpho-orthographic sensitivity would, in this context, have more advantages than drawbacks. Günther et al. (2019) set out to quantify morphological systematicity in both English and German, to determine whether it differed sufficiently to account for the contrasting transparency effects. Using a distributional semantics approach (Landauer & Dumais, 1997), they calculated "observed" meanings for complex words based on what other words they co-occurred with in a large text corpus for each language. Linear functions to capture how word meanings are transformed when a derivational affix (e.g., -NESS or -ITY) was added to a free-standing stem (e.g., DARK or PROSPER) were also approximated, and used to calculate what the "compositional" meaning of each derived word would be (i.e., the compositional meaning of DRESSER might be roughly "one who dresses"). Comparisons of compositional and observed semantic vectors showed evidence of a more systematic morphology in German relative to English, supporting the hypothesis of Plaut and Gonnerman (2000). This is only one example of how variations in morphological systems across languages can shape how readers process complex words.

Differences in the productivity of a language's morphology also appear to impact how complex words are processed. A cross-linguistic study comparing lexical decision for derived words found greater sensitivity to stem frequencies in English than in Finnish (Vannest, Bertram, Järvikivi, & Niemi, 2002). Given the rich and productive morphological system in Finnish, this pattern of stem frequency effects, often interpreted as a measure of morphological sensitivity, is initially surprising. However, the large number of words containing more than one affix in Finnish, resulting in many multimorphemic words and large morphological family sizes, may explain this result. For example, the suffix -TÖN in Finnish is followed by an additional suffix in 87.5% of its occurrences, while the suffix -ABLE in English is followed by an additional suffix in only 11.8% of its occurrences (Vannest et al., 2002). Moscoso Del Prado Martín et al. (2004) demonstrated that for languages in which morphological family sizes are large, statistics for subsets of the family that are more semantically similar predict processing effects better than those of the whole family. For example, the word TYÖTÖNMYYS ("unemployment") might be more informatively considered relative to other words beginning with TYÖTÖN ("workless" or "unemployed"), not all words beginning with TYÖ ("work"). Thus, TYÖTÖN is functionally a more relevant "stem" than TYÖ in this context. Finnish words with a single derivational suffix, such as those used for lexical decision by Vannest et al. (2002), may be perceived in a manner more similar to monomorphemic stems in English, explaining their weak stem frequency effects.

One final cross-linguistic difference that is not yet well understood can be found in the morphological processing of dyslexic readers. Dyslexic readers in orthographically transparent languages (in which letter-to-sound correspondences are consistent), such as Italian and Spanish, appear to take advantage of morphological structure to support their word recognition. Conversely, in more orthographically opaque languages, such as English and Chinese, dyslexic readers show a deficit in morphological processing (see review by Deacon, Tong, & Mimeau, 2019). This finding is somewhat surprising, as cases where spelling-to-sound mappings are less clear in written English words often correspond to a boost in spelling-to-meaning regularity (Rastle, 2019a). For example, JEALOUS, NERVOUS, BONUS, and NECKLACE all end in the same sound, but the -OUS spelling uniquely conveys adjective status (Berg & Aronoff, 2017). As argued by Rastle (2019b), sensitivity to these meaningful variations in spelling may be particularly essential for reading acquisition in English, yet dyslexics do not generally appear to benefit from it. One explanation for this relies on grain-size theory (Ziegler and Goswami, 2005; also see Law & Ghesquière, 2021; Marelli et al., 2020). For languages in which small visual units (one or two letters) reliably map to speech sounds, developing readers' progress towards direct mapping of form to meaning requires a large jump in grain-size (letter to word). Morphemes may provide an efficient intermediate stepping stone for readers in these languages, but less so in orthographically opaque languages in which the grain-size for phonological decoding is already large. This wouldn't imply that morphological information does not play an important role in English reading acquisition, but rather that it is not salient to dyslexic readers in the same way that phonological information is not. Understanding why the role of morphology varies for dyslexic readers across languages would enhance our understanding of the emergence of morphological processing and how it interacts with other well-studied aspects of word processing.

Across all three of these examples, it is clear that the manner in which complex words are processed is customized to the language in question. As a given individual might have been born into any linguistic community, a complete theory of complex word recognition processes must account for these cross-linguistic differences in terms of how language characteristics shape morphological representations and processes during language and reading acquisition. As illustrated by Plaut and Gonnerman (2000), neural network simulations are uniquely well-suited to addressing such questions.

Morphological processing during reading acquisition

As children learn to read, they come to know an increasingly large set of words by sight (Ehri, 2005). With reading experience, even their earliest learned orthographic representations continue to evolve and become more specific. For example, Castles, Davis, Cavalot, and Forster, (2007) found in a longitudinal study that although 3rd graders showed orthographic priming of real words (e.g., PLAY) from both replaced-letter (RLAY) and transposed-letter masked primes (LPAY), they showed only transposed-letter priming two years later. This suggests increasing specificity of orthographic representations, moving towards no orthographic priming in either condition, as is seen in adults (Castles et al., 2007; also see Castles, 1999).

Early readers appear to be sensitive to the presence and meaning of embedded words. Seven-year-olds reading in English are slower to reject words as belonging to a certain category (e.g., "body parts") if the words have embedded strings that pertain to that category, such as SHIP (containing HIP) and CLIP (containing LIP; Nation & Cocksey, 2009). Sensitivity to embedded strings, potentially a byproduct of less precisely specified orthographic representations, could play an important role in the detection and exploitation of orthographic-semantic regularities as the system is refined. Hasenäcker, Solaja, and Crepaldi, (2021) extended this finding to Italian children (ages 8 to 11 years) and adults, and additionally investigated the effect of embedded strings on category rejection for both words with suffix and non-suffix endings (BURRONE versus RAPACE, comparable to CORNER versus PEACE in English). They found that participants at all ages show similar effects of the embedded string in slowing rejection latencies, regardless of the word's ending, but interestingly younger children are more likely to make errors for words with affix endings.

Indeed, children develop sensitivity to the morphological structure of words quite early. In lexical decision tasks, early readers (ages 7 to 10) are less accurate in rejecting morphologically structured nonwords relative to non-suffixed or non-morphological nonwords in Italian (Burani, Marcolini, & Stella, 2002), English (Dawson, Rastle, & Ricketts, 2018), and French (Casalis, Quémart, & Duncan, 2015). McCutchen, Logan, and Biangardi-Orpe (2009) found in an overt priming task that 5th and 8th graders recognized English words (e.g., PLAN) more quickly following a morphologically related word (PLANNER) than following one that was only orthographically (PLANET) or semantically (STRATEGY) related. French 4th graders, similarly, showed stronger morphological priming relative to orthographic priming in an overtly primed lexical decision task (Casalis, Dusautoir, Colé & Ducrot, 2009). When primes were presented more briefly (but still visible; 75 ms), morphologically and orthographically related primes facilitated children's lexical decision latencies equally, suggesting that recognition of morphological relatedness at this phase of reading acquisition takes longer and is perhaps more reliant on semantic similarity (Casalis et al., 2009).

Developmental masked priming studies that manipulate semantic transparency (i.e., including opaque primes such as CORNER) reinforce the importance of semantic factors in the development of morphological priming. In English, Beyersmann, Castles, and Coltheart (2012) found that 3rd and 5th grade readers demonstrated transparent but not opaque morphological priming in a masked primed lexical decision task, whereas adult participants showed significant priming in both conditions. Dawson et al. (2021b) replicated this finding with participants across a broader range of ages to show the gradual emergence of opaque effects with increased word reading proficiency. Likewise, in Hebrew, transparent morphological primes facilitate lexical decision for both 4th and 7th graders, whereas only 7th graders showed a marginal opaque morphological priming effect (Schiff, Raveh, & Fighel, 2012). In German, Fleischhauer et al. (2021) found no masked morphological priming in 1st and 2nd graders, only transparent morphological priming in German 3rd graders, and both transparent and opaque priming in 4th graders and adults. Thus, in English, Hebrew, and German, opaque morphological priming appears to emerge later than transparent priming, and only after a great deal of reading experience.

In contrast, both Quémart and Casalis (2015) and Quémart, Casalis, and Colé (2011) found evidence for opaque morphological priming effects in relatively young French children. Quémart and Casalis (2015) found equally strong transparent and opaque masked priming effects for typically developing French readers in 4th and 8th grade, although dyslexic 8th graders showed only transparent priming effects. In Quémart et al. (2011), French children in 3rd, 5th, and 7th grade showed opaque priming (e.g., BAGUETTE, "breadstick", priming BAGUE, "ring") for both short (60 ms) and long (250 ms) prime durations, whereas adults showed such priming only for the short duration. Their results suggest that French children may acquire sensitivity to morphological form separate from meaning earlier than children reading in English and Hebrew. The persistence of opaque priming for long prime durations in developing French readers could potentially be due to less-developed semantic processing or slower orthographic inhibition in children. All of the developmental morphological priming studies run in languages other than French (described above) only included one prime duration (between 50 and 65 ms). It may be that overt prime durations would also reveal longer-lasting opaque priming for children in these languages, a possibility also supported by the recent results of Hasenäcker et al. (2021) in Italian. Additional developmental studies comparing morphological processing effects for different prime durations would clarify these findings' significance.

Direct cross-linguistic comparisons of morphological processing development suggest that children reading in French show earlier sensitivity to morphological structure than children reading in German (Beyersmann et al., 2021) or English (Casalis et al., 2015). The described difference across languages is somewhat puzzling: French has a richer morphological system than English but a less-productive one than German; additionally, its orthography is opaque, different from German but similar to English. The early onset of morphological processing effects in French children could be due to the joint pressures of a morphologically rich and orthographically opaque language, but further investigation and simulation work is merited to fully understand this difference.

The emergence of semantic transparency effects has also been investigated among adult second language learners. Silva and Clahsen (2008) found that morphological masked priming effects in non-fluent English language learners (whose native languages were German, Chinese, or Japanese) are much weaker or nonexistent compared with those of native English speakers, reinforcing that these effects rely on extensive written word experience. Low-proficiency Chinese-English bilingual participants show transparent, but not opaque, morphological masked priming (Li, Taft, & Xu, 2017). However, proficient Spanish-English, Dutch-English, and Italian-English bilinguals show both transparent and opaque morphological masked priming, although opaque effects are weaker than transparent effects (Diependaele et al., 2011; Viviani & Crepaldi, 2019). Interestingly, both of these studies also found that opaque and transparent effects were of more similar magnitudes in mostly proficient bilinguals than in highly proficient bilinguals, suggesting that opaque priming does not necessarily become steadily stronger with experience, but instead may peak and then lessen relative to transparent priming. This could be due to the differing impacts of masked presentations on readers, or additive effects of morphological and orthographic priming (in that non-native native readers showed orthographic priming whereas native readers did not).

Studies with adults have found morphological effects to be surprisingly robust to orthographic variability (as discussed previously; e.g., McCormick et al., 2008). Thus, another question of interest in the developmental literature is how morphological effects for irregular and less orthographically consistent paradigms emerge with development. It appears that the automatic recognition of morphological relatedness for irregular forms comes later: Quémart and Casalis (2015) showed that neither 4th nor 8th grade French children's lexical decisions were facilitated by masked primes with an orthographic shift (e.g., SOIGNEUX, "careful", and SOIN, "care"). Interestingly, those same children *did* show significant opaque morphological priming effects in the same paradigm (Quémart & Casalis, 2015, as discussed above), so morphological processing effects not supported by typical form-base similarity may emerge more slowly than those not supported by semantic transparency. In Hebrew, masked priming experiments with 3rd and 7th grade children using stimuli derived from defective roots showed similar sensitivity to orthographic variation (Schiff, Raveh, & Kahta, 2008). Significant facilitation was found when both prime and target contained all three letters of the root (e.g., [NiGuV] and [leNaGeV]), and when both prime and target contained only two letters of the root ([maGeVet] and [maGaV]), but not when one contained two letters of the root and the other contained three. Overall, abstraction away from precise orthographic representations of morphological paradigms during visual word processing appears to emerge relatively late in reading acquisition.

In summary, among early readers of a given language, masked priming effects for semantically opaque and orthographically inconsistent primes are generally weak or nonexistent. Thus, these effects might be best thought of as late properties of a well-trained system, not as foundational features of morphological processing across all readers. Within a distributed view, the emergence of morphological processing could be explained in the following manner: regularity between words' forms and their meanings provides initial informative associations that are relatively simple. With more practice and exposure to a wider variety of words, orthographic processing becomes increasingly sensitive to frequent and informative higher-order structures, to the point that even simple words containing those structures (e.g., CORNER) are affected. The described evidence of changes over acquisition mostly aligns with this sequence. However, relatively few distributed simulations of morphological processing have focused on developmental phenomena, and so emergence has not been adequately demonstrated. Doing so could illuminate the need for more authentic approaches to model learning, such as increasing exposure to complex words or allowing orthographic representations to change over the course of training.

As mentioned previously, accounts of the emergence of morphological sensitivity within decomposition frameworks also look to the statistics of the language. Rastle and Davis (2008) discussed three methods by which a developing reader might learn which letter sequences to represent on morphemes, derived from strategies discussed in the speech segmentation literature: 1) identifying boundaries from low probability sequences, 2) noting letter sequences that occur with high probability, and 3) detecting more meaningful letter groupings via form-meaning regularity. A separate proposal put forward by Grainger and Beyersmann (2017) focuses on the representation of affixes, reasoning that as embedded words are already salient due to their appearance as standalone words (flanked by spaces), the differentiation of BROTHER and BROTHEL relies on affix recognition. By activating embedded words alongside full words, complex words receive a semantic boost if the embedded word is related, and the accompanying letter string's semantic and orthographic representations are strengthened. In some ways, these developmental accounts are not so different from that of distributed accounts: regularities across contexts lead to learning of higher-order structures which impact processing. However, the proposal by Grainger and Beyersmann (2017) is limited in that it applies only to languages for which the vast majority of complex words contain only two morphemes (e.g., English and French) and for which words are delineated by spaces (which is not the case for Chinese). Additionally, the mechanisms described by both Rastle and Davis (2008) and Grainger and Beyersmann (2017) rely on a concatenative morphology (not present in Arabic and Hebrew) and isolable morphological constituents (violated in English by apophonic relations such as SANG-SING and DOVE-DIVE; Anderson, 1992). Although both accounts describe statistical features that likely contribute to the emergence of morphological sensitivity in certain languages, the distributed view provides a more fully specified and general explanation for how sensitivity to morphological structure is learned.

Artificial language learning studies with adults provide additional insights into how morphological sensitivity is acquired. Tamminen et al. (2015) investigated affix acquisition by teaching participants artificial words containing novel affixes (e.g., SLEEPNULE, SAILAFE). The impact of changes in learning context and materials on generalization of those affixes to new words was measured during sentence reading, using both explicit (judgments of semantic congruency) and implicit (semantic congruency effects in speeded reading aloud) tasks. They found that allowing time for consolidation—by conducting the testing session a week later rather than immediately after training—was necessary for semantic congruency effects to be detected during sentence reading, although performance with explicit judgments was unaffected (see also Merkx, Rastle, & Davis, 2011). The essential role of consolidation has been found in several other artificial language learning studies (e.g., Merkx et al., 2011; Tamminen, Davis, Merkx, & Rastle, 2012), supporting a complementary learning systems framework of episodic and semantic memory. The importance of consolidation is not sufficiently discussed in either decomposition-based or distributed accounts of morphological acquisition (although, see Kumaran, Hassabis, & McClelland, 2016 for a discussion of a neural network implementation of complementary learning systems in a more general context).

Tamminen et al. (2015) also found that generalization of affix learning was stronger for affixes that appeared in a greater variety of trained words, and for affixes that altered stem meaning in a semantically consistent manner. By training adults to recognize "words" constructed from morpheme-like chunks in an artificial script, Lelonkiewicz, Ktori, and Crepaldi, (2020) demonstrated that sensitivity to letter-like chunks can be acquired even without mapping to meaning, supporting the role of mechanisms such as boundary detection and identifying frequent letter chunks (Rastle & Davis, 2008) in bolstering morphological sensitivity. Together, these papers emphasize that both the semantic utility and statistical regularities of an affix determine how well it is learned. Although initial acquisition of morphological sensitivity in developing readers may differ from how skilled readers learn a few new affixes, artificial language learning studies provide notable insights regarding favorable conditions for morpheme learning. Implementing such paradigms with children (as done in Dawson, Rastle, & Ricketts, 2021a) could illuminate how learners who are refining relevant representations along many dimensions simultaneously may differ in their learning behavior.

More research is needed in order to fully understand how morphological processing emerges over development. As noted by Dawson et al. (2021b), as readers become more proficient they also recognize words more quickly, which means that a 50 ms prime presented to less experienced readers may yield a weaker or less informative signal than it would if presented to more experienced readers. Additionally, orthographic representations are developing and differentiating at the same time as morphological processing, and the timing with which these emerge relative to each other, and how they interact at different phases, is not well-specified. These challenges could be addressed by running masked priming studies that use multiple prime durations for several age groups (as done in Quémart et al., 2011), and always including an orthographic priming condition. Additionally, measurements of participants' word recognition proficiency may be a more precise metric of their phase of reading development than their age or grade level, particularly for the purposes of visual word processing research (Andrews & Lo, 2013; Dawson et al., 2021b; Kahraman & Kırkıcı, 2021). Taking such measurements as part of the experiment can help explain variability among participants of purportedly the same level, clarifying when certain phenomena occur. More generally, detailed studies of morphological effects in reading acquisition may provide key insights into how experience can shape mature, language-specific representations and processes.

General discussion

The picture that emerges from our review of studies of morphological processing in visual word recognition is of a flexible processing system that is very fine-tuned to the statistics of the language (e.g., relative entropy effects in Serbian), demonstrating graded degrees of morphological sensitivity depending on context (e.g., word frequency, word transparency, task goal) and morpheme characteristics (e.g., stem frequency, position-specificity). Morphological priming when the prime is not consciously perceived demonstrates greater sensitivity to the appearance of morphological structure, while also being robust to common orthographic variations. Morphological sensitivity in semantically opaque and orthographically inconsistent contexts emerges only after a great deal of reading experience.

In our view, the distributed account is better equipped to explain the full range of morphological processing effects. Distributed processing via propagated activation through optimally-adjusted weights can accommodate the graded, language- and context-specific nature of these results. Additionally, an explanation of how morphological sensitivity that is customized to statistics of the language emerges with experience is inherent in this framework, providing a fruitful lens for understanding both acquisitional and cross-linguistic comparisons.

Decomposition-based accounts, on the other hand, often need to be extended with added paths or features to address new challenges (e.g., Baayen et al., 1997; Crepaldi et al., 2010; Dawson et al., 2021b; McCormick et al., 2008; Taft, 1979). Because these accounts for the most part have not been implemented computationally, they are in many respects also underspecified (Rueckl, 2010). For example, at intermediate points in the acquisition of morpheme representations described by Grainger and Beyersmann (2017), what constitutes a partially-formed affix representation, and how does that lead to weaker or stronger decomposition? An answer to this question would be necessary to understand the gradual increase in opaque priming magnitudes over acquisition seen in Dawson et al. (2021b).

The primary shortcoming of the distributed account is that it is not yet complete. Few neural-network simulations have been published that directly explore known effects in morphological processing in visual word recognition, and thus many questions regarding whether certain empirical phenomena can be demonstrated within this framework have not been given a satisfactory response. For example, there have been no simulations of effects surrounding nonword priming, orthographic-semantic consistency, or orthographic variability in the context of morphological processing. Neural network simulations that account for word processing dynamics, capturing the distinctions between effects in masked versus overt priming contexts, are essential for determining whether the phenomena that most strongly suggest a role for decomposition can also be generated via distributed processes. Recurrent neural networks have been successful in capturing word processing dynamics in other contexts (e.g., Armstrong and Plaut, 2016; Cheyette & Plaut, 2017), so applying such methods to morphological questions is feasible.

More adequate treatment of semantic processing and semantic representations in distributed models would also advance our understanding of morphological processing. In particular, authentic depictions of the sparse and dynamic nature of lexical semantics and the nature of morphological contributions to word meaning could greatly enhance our understanding of morphological family size effects, task- and context-varying phenomena, and pseudo-complex nonword reading. Such improvements would also enhance the relevance of morphological processing models to higher-level processes (e.g., sentence comprehension).

In addition to its suitability for explaining the existing range of empirical morphological processing effects, some more general considerations favor transitioning to a distributed view. These are discussed briefly below.

Aligning with a graded view of morphology

The distinction between morphemes and non-morphemes is less discrete than it may appear at first glance. For example, the bound stem -MIT-, though etymologically considered a Latinate morpheme meaning "send", makes a very unclear semantic contribution across its appearances in ADMIT, COMMIT, PERMIT, and other words (Aronoff, 1976). Meanwhile, phonaesthemes such as GL- (GLIMMER, GLEAM, GLINT) and SN- (SNORT, SNIFF, SNEEZE) are not counted as morphemes, despite their prevalence in words with certain meanings and their morpheme-like priming effects (Bergen, 2004; see Baayen et al., 2011 for simulations of these effects using the Naive Discriminative Reader). There are a variety of ways in which spellings convey meaning, and the categorization schemes proposed by structuralist linguistic theories do not determine how morphological processing emerges; instead, these mechanisms are shaped by regularities learned via reading exposure.

Marelli et al. (2020) directly challenges the "morpheme-as-unit" assumption that is prevalent in morphological processing studies, arguing that "psycholinguistic studies have found evidence for morphological units because they were designed to look for that evidence" (p. 566; also see Anderson, 1992; Blevins, 2016; Gonnerman et al., 2007; Hay and Baayen, 2005). However, some experimentalists in morphological processing have already begun to shift towards a more graded approach to morphological information in their study designs. Ulicheva, Harvey, Aronoff, and Rastle (2020) investigated regularities between spelling and lexical category (e.g., -OON occurs predominantly in nouns such as SPOON, NOON, and MOON). They quantified two such regularities—diagnosticity and specificity—and demonstrated that skilled readers are sensitive to the magnitudes of these graded metrics. Amenta et al. (2020) also used a graded metric to demonstrate the spectrum between meaningful and meaningless letter strings: a high-frequency masked prime facilitates recognition of its stem (characteristic of morphological priming) if the stem has high orthographic-semantic consistency, but inhibits recognition of its stem (characteristic of orthographic priming) if it has low orthographic-semantic consistency. Such graded characterizations of regularity in the form-to-meaning mapping have become more possible in recent years thanks to advances in computational approaches such as distributional semantics (see Amenta et al., 2020), and they are essential for more informative characterization of the impact of morphological information on word processing. However, it is worth noting that both of these efforts still quantify degree of morphological information with respect to units of contiguous letters, an approach that cannot be generalized to all cases (e.g., Hebrew or Setswana, as discussed above). Overcoming this barrier will likely require more involved metrics of shared orthographic structure.

Decomposition accounts of morphological processing are less naturally reconciled with a graded view of morphological information. Splitting a word in some cases and not in others would require a threshold judgment (Rastle & Davis, 2008) leading to unnecessarily lost information from less informative letter strings such as -OON and

GL-. For an interpretation of decomposition in which the word is not necessarily split but constituent morphemes are represented, determining which letter strings demonstrate sufficient regularity to be represented, and how to represent potentially innumerable letter strings, remains a challenge. In a distributed account, all letters within a word can contribute to its processing via stronger and weaker weights, customized to useful regularities previously detected in the language. Separate representations of all slightly meaningful letter combinations are not necessary.

Integrating into broader theories of cognition

Written word recognition is primarily a task accomplished by the visual system, and several established orthographic effects, such as bigram frequency effects, have been shown to occur for other types of complex visual stimuli as well (Vidal, Viviani, Zoccolan, & Crepaldi, 2021). This suggests that many mechanisms involved in word recognition are also applicable to vision more generally. Neural network simulations have proved an essential tool for understanding the visual system (Kriegeskorte, 2015; Yamins & DiCarlo, 2016), and for appreciating complex word recognition within this broader context (Hannagan et al., 2021). Shared use of a distributed approach may allow insights from morphological processing research to inform our understanding of other visual phenomena, and vice versa.

Many researchers studying morphological processing are not doing so out of an interest in componential processing or mid-level visual processing per se, but because of its relevance to reading and language more generally. Integrating morphological processing into broader theories of reading and language requires a theoretical approach that can be easily combined with our understanding of orthographic, phonological, semantic, sentence and discourse processing. Neural network models rely on general learning mechanisms that can be applied to any domain or modality, and have been shown to be applicable in many of these contexts (e.g., Cheyette & Plaut, 2017; Plaut, McClelland, Seidenberg, & Patterson, 1996; Rabovsky & McClelland, 2020).

Distributed approaches represent a general theory of cognition, approximating how principles of neural computation and learning can give rise to complex behavior. Although this may not be the most effective way of capturing every higher-level cognitive phenomenon at present (e.g., conscious problem solving, or social interactions), the evidence we have reviewed suggests that it is the most productive approach to understanding morphological processing. Contextualizing stimuli- and task-specific phenomena within a broader common theory of learning and cognition promotes integration across separately studied phenomena. For example, the relationship between morphological processing in visual and auditory modalities might be better understood with a distributed approach.

Compatibility with natural language processing research

One final notable advantage of embracing a distributed account of language processing is that current state-of-the-art systems for natural language processing (NLP) rely on similar mechanisms and assumptions. For example, GPT-3 is a language model developed to produce high-quality human-like text (Brown et al., 2020, also see Radford et al., 2019). The model was generated by training a larger Transformer architecture (similar to recurrent neural networks, but parallelized for faster training; see Vaswani et al., 2017) on a massive text corpus, to predict natural language sequences. Without being trained on any of the tasks explicitly, GPT-3 can translate, answer questions, and perform simple arithmetic. It can even write entire articles (GPT-3, 2020). Although substantial gaps remain between the performance of such models and that of humans (Floridi & Chiriatti, 2020), it is worth appreciating that the most successful methods to date for processing and generating language in a human-like manner rely on graded and distributed learning mechanisms. The insights we can gain from NLP relevant to human word recognition are currently limited, as most efforts in this area take tokenized words as input (for some interesting exceptions, see Cao & Rei, 2016; Liu, Lu, Lo, & Neubig, 2017; Zhang, Zhao, & LeCun, 2015). Additionally, NLP research is generally not focused on understanding the dynamics of language processing, so data comparable to reaction times and priming magnitudes are not examined. However, greater alignment in scientists' and engineers' understanding of core mechanisms of language cognition could lead to a more productive exchange of information between these perspectives.

Conclusion

Our review of the morphological processing literature suggests that, although both decomposition and distributed models may be consistent with known empirical effects, the distributed approach provides a more straightforward account, while also allowing for contexts in which decomposition-like phenomena occur. A distributed approach also better explains cross-linguistic similarities and differences in morphological processing, and how the underlying mechanisms can be acquired via language experience. Contextualizing decomposition effects within a graded and distributed view of morphological processing, not only in our approach to experimental design but also in our theoretical models of underlying mechanisms, has benefited and will continue to benefit this subfield. Additional simulation work with neural network models, particularly with dynamic, recurrent models, will be essential to reaching a deeper understanding of how readers make use of morphological structure in written language.

Acknowledgment The authors wish to thank Marco Marelli, Kathleen Rastle, Marlene Behrmann, Anna Fisher, Nazbanou Nozari, and Charles Perfetti for helpful feedback and comments on previous versions of this manuscript. The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305B150008 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Practices Statement There are no data or experiments associated with the current work. Thus, no data has been made publicly available and no experiments were preregistered.

References

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, *40*(1), 41–61.

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Psychology*, *3*, 1–12.

Amenta, S., Crepaldi, D., & Marelli, M. (2020). Consistency measures individuate dissociating semantic modulations in priming paradigms: A new look on semantics in the processing of (complex) words. *Quarterly Journal of Experimental Psychology*, *73*(10), 1546–1563.

Amenta, S., Günther, F., & Marelli, M. (2020). A (distributional) semantic perspective on the processing of morphologically complex words. *The Mental Lexicon*, *15*(1), 62–78.

Amenta, S., Marelli, M., & Crepaldi, D. (2015). The fruitless effort of growing a fruitless tree: Early morpho-orthographic and morpho-semantic effects in sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1587.

Anderson, S. R. (1992). *A-morphous morphology (No. 62)*. Cambridge: Cambridge University Press.

Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, *68*(3), 279–296.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, *31*(7), 940–966.

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

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge: MIT Press.

Baayen, R. H. (2010). The directed compound graph of English: An exploration of lexical connectivity and its processing consequences. *New Impulses in Word-Formation*, *17*, 383–402.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. *Complexity*.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, *37*(1), 94–117.

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

Baayen, R. H., Milin, P., Revic, 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.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The celex lexical database (release 2). *Distributed by the linguistic data consortium*, University of Pennsylvania.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *The celex lexical database (cd-rom)*. Philadelphia: University of Pennsylvania.

Baayen, R. H., & Smolka, E. (2020). Modeling morphological priming in german with naive discriminative learning. *Frontiers in Communication*, *5*, 17.

Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon*, *2*(3), 419–463.

Badecker, W., & Allen, M. (2002). Morphological parsing and the perception of lexical identity: A masked priming study of stem homographs. *Journal of Memory and Language*, *47*(1), 125–144.

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.

Berg, K., & Aronoff, M. (2017). Self-organization in the spelling of english suffixes: The emergence of culture out of anarchy. *Language*, *93*(1), 37–64.

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, *80*(2), 290–311.

Bertram, R., Baayen, R. H., & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language*, *42*(3), 390–405.

Beyersmann, E., Castles, A., & Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *The Quarterly Journal of Experimental Psychology*, *65*(7), 1306–1326.

Beyersmann, E., Cavalli, E., Casalis, S., & Colé, P. (2016). Embedded stem priming effects in prefixed and suffixed pseudowords. *Scientific Studies of Reading*, *20*(3), 220–230.

Beyersmann, E., Coltheart, M., & Castles, A. (2012). Parallel processing of whole words and morphemes in visual word recognition. *Quarterly Journal of Experimental Psychology*, *65*(9), 1798–1819.

Beyersmann, E., & Grainger, J. (2018). Support from the morphological family when unembedding the stem. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *44*(1), 135–142.

Beyersmann, E., Mousikou, P., Schroeder, S., Javourey-Drevet, L., Ziegler, J. C., & Grainger, J. (2021). The dynamics of morphological processing in developing readers: A cross-linguistic masked priming study. *Journal of Experimental Child Psychology*.

Beyersmann, E., Ziegler, J. C., Castles, A., Coltheart, M., Kezilas, Y., & Grainger, J. (2016). Morpho-orthographic segmentation without semantics. *Psychonomic Bulletin & Review*, *23*(2), 533–539.

Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford: Oxford University Press.

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

Bradley, D. (1979). Lexical representation of derivational relation. In M. Kean, & M. L. Aronoff (Eds.) *Juncture*. Cambridge: Academic Press.

Broadbent, D., & Gregory, M. (1968). Visual perception of words differing in letter digram frequency. *Journal of Verbal Learning & Verbal Behavior*.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Burani, C., Dovetto, F. M., Spuntarelli, A., & Thornton, A. M. (1999). Morpholexical access and naming: The semantic interpretability of new root–suffix combinations. *Brain and Language*, *68*(1–2), 333–339.

Burani, C., Marcolini, S., & Stella, G. (2002). How early does morpholexical reading develop in readers of a shallow orthography? *Brain and Language*, *81*(1–3), 568–586.

Burani, C., Salmaso, D., & Caramazza, A. (1984). Morphological structure and lexical access (Vol. 18, No. 4).

Burani, C., & Thornton, A. M. (2011). The interplay of root, suffix and whole-word frequency in processing derived words. In *Morphological structure in language processing* (pp. 157–208). De Gruyter Mouton.

Butterworth, B. (1983). Lexical representation. In Butterworth, B. (Ed.) *Language production* (Vol. 2, pp. 257–294). London: Academic Press.

Cao, K., & Rei, M. (2016). A joint model for word embedding and word morphology. In *Proceedings of the 1st workshop on representation learning for nlp* (pp. 18–26).

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, *28*(3), 297–332.

Carden, J. R., Barreyro, J. P., Segui, J., & Jaichenco, V. (2019). The fundamental role of position in affix identity. *The Mental Lexicon*, *14*(3), 357–380.

Casalis, S., Dusautoir, M., Colé, P., & Ducrot, S. (2009). Morphological effects in children word reading: A priming study in fourth graders. *British Journal of Developmental Psychology*, *27*(3), 761–766.

Casalis, S., Quémart, P., & Duncan, L. G. (2015). How language affects children's use of derivational morphology in visual word and pseudoword processing: Evidence from a cross-language study. *Frontiers in Psychology*, *6*, 452.

Castles, A. (1999). Neighbourhood effects on masked form priming in developing readers. *Language and Cognitive Processes*, *14*(2), 201–224.

Castles, A., Davis, C., Cavalot, P., & Forster, K. I. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology*, *97*(3), 165–182.

Chee, Q. W., & Yap, M. (2022). Are there task-specific effects in morphological processing? Examining semantic transparency effects in semantic categorization and lexical decision. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/17470218221079269

Cheyette, S. J., & Plaut, D. C. (2017). Modeling the n400 erp component as transient semantic over-activation within a neural network model of word comprehension. *Cognition*, *162*, 153–166.

Ciaccio, L. A., Kgolo, N., & Clahsen, H. (2020). Morphological decomposition in Bantu: A masked priming study on Setswana prefixation. *Language, Cognition and Neuroscience*, *35*(10), 1257–1271.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of german inflection. *Behavioral and Brain Sciences*, *22*(6), 991–1013.

Clahsen, H., & Ikemoto, Y. (2012). The mental representation of derived words: An experimental study of–sa and–mi nominals in Japanese. *The Mental Lexicon*, *7*(2), 147–182.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the stroop effect. *Psychological Review*, *97*, 332–361.

Colé, P., Beauvillain, C., & Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: A differential frequency effect. *Journal of Memory and Language*, *28*(1), 1–13.

Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). 'Fell' primes 'fall', but does 'bell' prime 'ball'? Masked priming with irregularly-inflected primes. *Journal of Memory and Language*, *63*(1), 83–99.

Crepaldi, D., Rastle, K., & Davis, C. J. (2010). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & Cognition*, *38*(3), 312–321.

Dasgupta, T., Sinha, M., & Basu, A. (2015). Computational modeling of morphological effects in bangla visual word recognition. *Journal of Psycholinguistic Research*, *44*(5), 587–610.

Davis, M. H., & Rastle, K. (2010). Form and meaning in early morphological processing: Comment on Feldman, O'Connor, and Moscoso del Prado Martín (2009). *Psychonomic Bulletin and Review*, *17*(5), 749–755.

Dawson, N., Rastle, K., & Ricketts, J. (2018). Morphological effects in visual word recognition: Children, adolescents, and adults. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(4), 645–654.

Dawson, N., Rastle, K., & Ricketts, J. (2021a). Bridging form and meaning: Support from derivational suffixes in word learning. *Journal of Research in Reading*, *44*(1), 27–50.

Dawson, N., Rastle, K., & Ricketts, J. (2021b). Finding the man amongst many: A developmental perspective on mechanisms of morphological decomposition. *Cognition*, *211*, 104605.

Deacon, H., Tong, X., & Mimeau, C. (2019). Semantic processing in developmental dyslexia across languages Morphological A theoretical and empirical review. In L. Verhoeven, C. A. Perfetti, & K. R. Pugh (Eds.) *Developmental dyslexia across languages and writing systems: The big picture* (pp. 327–349). Cambridge: Cambridge University Press.

De Grauwe, S., Lemhöfer, K., & Schriefers, H. (2019). Processing derived verbs: The role of motor-relatedness and type of morphological priming. *Language, Cognition and Neuroscience*, *34*(8), 973–990.

De Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., & Baayen, R. H. (2002). The processing and representation of dutch and english compounds: Peripheral morphological and central orthographic effects. *Brain and Language*, *81*(1–3), 555–567.

De Jong, N. H., Schreuder, R., & Baayen, H. R. (2000). The morphological family size effect and morphology. *Language and cognitive processes*, *15*(4–5), 329–365.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

De Rosa, M., & Crepaldi, D. (2021). Letter chunk frequency does not explain morphological masked priming. *Psychonomic Bulletin & Review*, 1–11.

Diependaele, K., Duñabeitia, J.. A.., Morris, J., & Keuleers, E.. (2011). Fast morphological effects in first and second language word recognition. *Journal of Memory and Language*, *64*(4), 344–358.

Diependaele, K., Grainger, J., & Sandra, D. (2012). Morphological processing. In *The Cambridge handbook of psycholinguistics* (Vol. 211, pp. 311–350).

Diependaele, K., Sandra, D., & Grainger, J. (2009). Semantic transparency and masked morphological priming: The case of prefixed words. *Memory and Cognition*, *37*(6), 895–908.

Drews, E., & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 1098.

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.

Duñabeitia, J. A., Laka, I., Perea, M., & Carreiras, M. (2009). Is milkman a superhero like batman? Constituent morphological priming in compound words. *European Journal of Cognitive Psychology*, *21*(4), 615–640.

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

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2008). Does darkness lead to happiness? Masked suffix priming effects. *Language and Cognitive Processes*, *23*(7–8), 1002–1020.

Ehri, L. C. (2005). Learning to read words: Theory, Findings, and Issues. *Scientific Studies of Reading*, *9*(2), 167–188.

Elman, J. L., Bates, E. A., & Johnson, M. H. (1996). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). Cambridge: MIT Press.

Feldman, L. B. (1994). Beyond orthography and phonology: Differences between inflections and derivations. *Journal of Memory and Language*, *33*, 442–470.

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

Feldman, L. B., Milin, P., Cho, K. W., Moscoso del Prado Martín, F., & O'Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Frontiers in Human Neuroscience*, *9*, 1–19.

Feldman, L. B., O'Connor, P. A., & 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 and Review*, *16*(4), 684–691.

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

Fiorentino, R., & Fund-Reznicek, E. (2009). Masked morphological priming of compound constituents. *The Mental Lexicon*, *4*(2), 159–193.

Fiorentino, R., Politzer-Ahles, S., Pak, N. S., Martínez-García, M.. T.., & Coughlin, C. (2015). Dissociating morphological and form priming with novel complex word primes: Evidence from masked priming, overt priming, and event-related potentials. *The Mental Lexicon*, *10*(3), 413–434.

Fleischhauer, E., Bruns, G., & Grosche, M. (2021). Morphological decomosition supports word recognition in primary school children learning to read: Evidence from masked priming of German derived words. *Journal of Research in Reading*, *44*(1), 90–109.

Floridi, L., & Chiriatti, M. (2020). Gpt-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), 681–694.

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

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635.

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680–698.

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

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

Gilbert, S. J., & Shallice, T. (2002). Task switching: A pdp model. *Cognitive Psychology*, *44*, 297–337.

Giraudo, H., Dal Maso, S., & Piccinin, S. (2016). The role of stem frequency in morphological processing. In *Mediterranean morphology meetings* (Vol. 10, pp. 64–72).

Giraudo, H., & Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Language and Cognitive Processes*, *15*(4–5), 421–444.

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.

GPT-3 (2020). A robot wrote this entire article. are you scared yet, human? The Guardian. Retrieved from https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

Grainger, J., & Beyersmann, E. (2017). Edge-aligned embedded word activation initiates morpho-orthographic segmentation. In *Psychology of learning and motivation* (Vol. 67, pp. 285–317). Elsevier.

Grainger, J., & Beyersmann, E. (2020). Effects of lexicality and pseudo-morphological complexity on embedded word priming. *Journal of Experimental Psychology, Learning, Memory, and Cognition*.

Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language*, *30*(1), 370–384.

Grainger, J., & Ziegler, J. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, *2*, 1–13.

Günther, F., & Marelli, M. (2018). Enter sandman: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental psychology. Learning, Memory, and Cognition*, *45*(10), 1872–1882.

Günther, F., Marelli, M., & Bölte, J. (2020). Semantic transparency effects in german compounds: A large dataset and multiple-task investigation. *Behavior Research Methods*, *52*(3), 1208–1224.

Günther, F., Smolka, E., & Marelli, M. (2019). 'Understanding' differs between English and German: Capturing systematic language differences of complex words. *Cortex*, *116*, 168–175.

Hannagan, T., Agrawal, A., Cohen, L., & Dehaene, S. (2021). Emergence of a compositional neural code for written words: Recycling of a convolutional neural network for reading. *Proceedings of the National Academy of Science USA*, *118*, e2104779118.

Hasenäcker, J., Solaja, O., & Crepaldi, D. (2021). Does morphological structure modulate access to embedded word meaning in child readers? *Memory & Cognition*, *49*(7), 1344–1347.

Hay, J. B., & Baayen, R. H. (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences*, *9*(7), 342–348.

Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(1), 157.

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

Jared, D., Jouravlev, O., & Joanisse, M. F. (2017). The effect of semantic transparency on the processing of morphologically derived words: Evidence from decision latencies and event-related potentials. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(3), 422–450.

Kahraman, H., & Kırkıcı, B.. (2021). Letter transpositions and morphemic boundaries in the second language processing of derived words: An exploratory study of individual differences. *Applied Psycholinguistics*, *42*(2), 417–446.

Kgolo, N., & Eisenbeiss, S. (2015). The role of morphological structure in the processing of complex forms: Evidence from Setswana deverbative nouns. *Language, Cognition and Neuroscience*, *30*(9), 1116–1133.

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, *1*, 417–446.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences*, *20*(7), 512–534.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*(7–8), 1089–1132.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 876.

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), 211–240.

Laudanna, A., Badecker, W., & Caramazza, A. (1989). Priming homographic stems. *Journal of Memory and Language*, *28*(5), 531–546.

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

Lavric, A., Elchlepp, H., & Rastle, K. (2012). Tracking hierarchical processing in morphological decomposition with brain potentials. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 811–816.

Law, J. M., & Ghesquière, P. (2021). Morphological processing in children with developmental dyslexia: A visual masked priming study. *Reading Research Quarterly*.

Leinonen, A., Grønholm-Nyman, P., Järvenpää, M., Søderholm, C., Lappia, O., Laine, M., & et al. (2009). Neurocognitive processing of auditorily and visually presented inflected words and pseudowords: Evidence from a morphologically rich language. *Brain Research*, *1275*, 54–66.

Lelonkiewicz, J. R., Ktori, M., & Crepaldi, D. (2020). Morphemes as letter chunks: Discovering affixes through visual regularities. *Journal of Memory and Language*, *115*, 104152.

Li, J., Taft, M., & Xu, J. (2017). The processing of English derived words by Chinese-English bilinguals. *Language Learning*, *67*(4), 858–884.

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, *84*(1), 50–64.

Liu, F., Lu, H., Lo, C., & Neubig, G. (2017). Learning character-level compositionality with visual features. In *Proceedings of the 55th annual meeting of the association for computational linguistics* (Vol. 1: Long papers, pp. 2059–2068).

Longtin, C. M., & Meunier, F. (2005). Morphological decomposition in early visual word processing. *Journal of Memory and Language*, *53*(1), 26–41.

Longtin, C. M., Segui, J., & Hallé, P. A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes*, *18*(3), 313–334.

Lüdeling, A., & De Jong, N. (2002). German particle verbs and word-formation. *Verb-Particle Explorations*, 315–333.

Manelis, L., & Tharp, D. A. (1977). The processing of affixed words. *Memory & Cognition*, *5*(6), 690–695.

Marelli, M., & Amenta, S. (2018). A database of orthography-semantics consistency (osc) estimates for 15,017 English words. *Behavior Research Methods*, *50*(4), 1482–1495.

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

Marelli, M., Amenta, S., Morone, E. A., & Crepaldi, D. (2013). Meaning is in the beholder's eye: Morpho-semantic effects in masked priming. *Psychonomic Bulletin & Review*, *20*(3), 534–541.

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review*, *122*(3), 485–515.

Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as abstract operation in semantic space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition*, *166*, 207–224.

Marelli, M., Traficante, D., & Burani, C. (2020). Reading morphologically complex words: Experimental evidence and learning models. In V. Pirrelli, I. Plag, & W. U. Dressler (Eds.) *Word knowledge and word usage: A cross-disciplinary guide to the mental lexicon*, (pp. 553–592). de Gruyter Mouton.

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

Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*(1), 3–33.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*(5), 375–407.

McCormick, S. F., Brysbaert, M., & Rastle, K. (2009). Short article: Is morphological decomposition limited to low-frequency words? *Quarterly Journal of Experimental Psychology*, *62*(9), 1706–1715.

McCormick, S. F., Rastle, K., & Davis, M. H. (2008). Is there a 'fete' in 'fetish'? Effects of orthographic opacity on morpho-orthographic segmentation in visual word recognition. *Journal of Memory and Language*, *58*(2), 307–326.

McCutchen, D., Logan, B., & Biangardi-Orpe, U. (2009). Making meaning: Children's sensitivity to morphological information during word reading. *Reading Research Quarterly*, *44*(4), 360–376.

Merkx, M., Rastle, K., & Davis, M. H. (2011). The acquisition of morphological knowledge investigated through artificial language learning. *Quarterly Journal of Experimental Psychology*, *64*(6), 1200–1220.

Meunier, F., & Longtin, C. M. (2007). Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, *56*(4), 457–471.

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

Milin, P., Filipovic Durdevic, 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.

Milin, P., Smolka, E., & Feldman, L. B. (2018). Models of lexical access and morphological processing. In E. M. Fernández, & H.

S. Cairns (Eds.) *The handbook of psycholinguistics* (pp. 240–268). Wiley.

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

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.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In *Processing of visible language* (pp. 259–268). Boston: Springer.

Moscoso Del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R. H. (2004). Morphological family size in a morphologically rich language: The case of finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(6), 1271–1278.

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.

Mulder, K., Dijkstra, T., Schreuder, R., & Baayen, H. R. (2014). Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language*, *72*, 59–84.

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, *102*(6), 963.

Myers, J., Huang, Y. C., & Wang, W. (2006). Frequency effects in the processing of Chinese inflection. *Journal of Memory and Language*, *54*(3), 300–323.

Nation, K., & Cocksey, J. (2009). Beginning readers activate semantics from sub-word orthography. *Cognition*, *110*(2), 273–278.

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.

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.

Moscoso del Prado Martın, F. (2003). Paradigmatic effects in morphological processing: Computational and cross-linguistic experimental studies. MPI Series in Psycholinguistics Max Planck Institute for Psycholinguistics, Nijmegen.

Moscoso del Prado Martín, F., Deutsch, A., Frost, R., Schreuder, R., De Jong, N. H., & Baayen, R. H. (2005). Changing places: A cross-language perspective on frequency and family size in Dutch and Hebrew. *Journal of Memory and Language*, *53*(4), 496–512.

Moscoso del Prado Martín, F., Ernestus, M., & Baayen, R. H. (2004). Do type and token effects reflect different mechanisms? Connectionist modeling of Dutch past-tense formation and final devoicing. *Brain and Language*, *90*(1–3), 287–298.

Quémart, P., & Casalis, S. (2015). Visual processing of derivational morphology in children with developmental dyslexia: Insights from masked priming. *Applied Psycholinguistics*, *36*(2), 345–376.

Quémart, P., Casalis, S., & Colé, P. (2011). The role of form and meaning in the processing of written morphology: A priming study in French developing readers. *Journal of Experimental Child Psychology*, *109*(4), 478–496.

Rabovsky, M., & McClelland, J. L. (2020). Quasi-compositional mapping from form to meaning: A neural network-based approach to capturing neural responses during human language comprehension. *Philosophical Transactions of the Royal Society B*, *375*(1791).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, *1*(8), 9.

Rapp, B. C. (1992). The nature of sublexical orthographic organization: The bigram trough hypothesis examined. *Journal of Memory and Language*, *31*, 33–53.

Rastle, K. (2019a). Eps mid-career prize lecture 2017: Writing systems, reading, and language. *Quarterly Journal of Experimental Psychology*, *72*(4), 677–692.

Rastle, K. (2019b). The place of morphology in learning to read in English. *Cortex*, *116*, 45–54.

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

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

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

Rice, G. A., & Robinson, D. O. (1975). The role of bigram frequency in the perception of words and nonwords. *Memory & Cognition*, *3*(5), 513–518.

Rueckl, J. G. (2010). Connectionism and the role of morphology in visual word recognition. *The Mental Lexicon*, *5*(3), 371–400.

Rueckl, J. G., & Aicher, K. A. (2008). Are CORNER and BROTHER morphologically complex? Not in the long term. *Language and Cognitive Processes*, *23*(7–8), 972–1001.

Rueckl, J. G., & Raveh, M. (1999). The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language*, *68*, 110–117.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1: Foundation, pp. 318–362). Cambridge: MIT Press.

Rumelhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2: psychological and biological models, pp. 216–271). Cambridge: MIT Press.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1.

Schiff, R., Raveh, M., & Fighel, A. (2012). The development of the hebrew mental lexicon: When morphological representations become devoid of their meaning. *Scientific Studies of Reading*, *16*(5), 383–403.

Schiff, R., Raveh, M., & Kahta, S. (2008). The developing mental lexicon: Evidence from morphological priming of irregular Hebrew forms. *Reading and Writing*, *21*(7), 719–743.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. *Morphological Aspects of Language Processing*, *2*, 257–294.

Schreuder, R., & Baayen, R. H. (1997). How complex simplex words can be. *Journal of Memory and Language*, *37*(1), 118–139.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Seidenberg, M. S., & Plaut, D. C. (2014). Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive Science*, *38*(6), 1190–1228.

Shapiro, B. J. (1969). The subjective estimation of relative word frequency. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 248–251.

Siegelman, N., Rueckl, J. G., Lo, J. C. M., Kearns, D. M., Morris, R. D., & Compton, D. L. (2022). Quantifying the regularities between orthography and semantics and their impact on group-and individuallevel behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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

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).

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

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 399–412.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory & Cognition*, *7*(4), 263–272.

Taft, M. (1985). The decoding of words in lexical access: A review of the morphographic approach. In D. Besner, T. Waller, & G. MacKinnon (Eds.) *Reading research: Advances in theory and practice*, (pp. 83–126). New York: Academic Press.

Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Language and Cognitive Processes*, *9*(3), 271–294.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *57*(4), 745–765.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words (Vol. 14(No. 6)).

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, *15*(6), 607–620.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition*, *125*(1), 107–112.

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology*, *79*, 1–39.

Tsang, Y. K., & Chen, H. C. (2013). Early morphological processing is sensitive to morphemic meanings: Evidence from processing ambiguous morphemes. *Journal of Memory and Language*, *68*(3), 223–239.

Tzur, B., & Frost, R. (2007). Soa does not reveal the absolute time course of cognitive processing in fast priming experiments. *Journal of Memory and Language*, *56*(3), 321–335.

Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2020). Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition*, *195*, 1–21.

Vannest, J., Bertram, R., Järvikivi, J., & Niemi, J. (2002). Counterintuitive cross-linguistic differences: More morphological computation in English than in Finnish. *Journal of Psycholinguistic Research*, *31*(2), 83–106.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Velan, H., Frost, R., Deutsch, A., & Plaut, D. C. (2005). The processing of root morphemes in Hebrew: Contrasting localist and distributed accounts. *Language and Cognitive Processes*, *20*(1–2), 169–206.

Vidal, Y., Viviani, E., Zoccolan, D., & Crepaldi, D. (2021). A general-purpose mechanism of visual feature association in visual word identification and beyond. *Current Biology*, *31*(6), 1261–1267.

Viviani, E., & Crepaldi, D. (2019). Masked morphological priming tracks the development of a fully mature lexical system in L2. *Journal of Memory and Language*.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365.

Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. *Advances in Neural Information Processing Systems*, *28*, 649–657.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3.

Zwitserlood, P. (2022). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and Cognitive Processes*, *9*(3), 341–368.