Morphological and pseudomorphological effects in English visual word processing: How much can we attribute the statistical structure of the language?

Patience Stevens (pstevens@andrew.cmu.edu)

David Plaut (plaut@cmu.edu)
Carnegie Mellon University
Department of Pyschology, 4909 Frew St
Pittsburgh, PA 15213

Abstract

The statistical structure of a given language likely drives our sensitivity to words' morphological structure. The current work begins to investigate to what degree morphological processing effects observed in visual word recognition can be attributed to statistical regularities between orthography and semantics in English, without any prior knowledge or explicitly coded processes. We trained a simple feedforward neural network on form-to-meaning mappings for words from an English educational text corpus. Over the course of training, we originally examined the network's processing times for prime-target word pairs taken from two masked primed lexical decision studies (Rastle, Davis & New, 2004; Beyersmann, Castles, & Coltheart, 2012) to determine if the network was learning similar sensitivities to those seen in human participants. Results showed no morphological sensitivity to prime-target pairs with a transparent morphological relationship (e.g., teacher → TEACH) or an opaque morphological relationship (e.g., corner → CORN). To increase power, unique prime-target pairs from a larger set of studies (10 in total) were added to the testing set. With the larger testing set, strong transparent morphological priming effects were observed, while opaque morphological priming was nonexistent. This work shows that morphological sensitivity can emerge without any explicit knowledge of morphemes or word structure, and that opaque morphological priming cannot be explained solely by feedfoward mapping of existing orthographic-semantic regularities. Preliminary work on a more dynamic and neurally-plausible model meant to better capture emerging morphological processing effects is described.

Keywords: morphological processing; visual word recognition; neural network modeling

Introduction

Morphologically complex words contain multiple morphemes, or sequences of letters that convey meaning (e.g., unkindness = un + kind + ness). More than half of written words that English-speaking students encounter while reading between grades 3 and 9 are morphologically complex (Nagy & Anderson, 1984), indicating there is non-negligible regularity in the mappings that young readers learn from visual representations of words to their meanings. Decades of studies show that skilled English-speaking readers are sensitive to this morphological structure during visual word processing (e.g., Taft & Forster, 1975; Marslen-Wilson et al., 1994; Beyersmann et al., 2016; see review by Amenta & Crepaldi, 2012). Priming studies, particularly masked primed lexical decision studies, have found that morphologically-related primes facilitate target decisions (e.g., "teacher → TEACH") beyond what can be explained by summative effects of orthographic and semantic similarity (Rastle et al., 2000). Furthermore, even words which only have the *appearance* of a morphological relation (referred to as "pseudomorphological" or "morphologically opaque" primes, such as brother → BROTH or belly → BELL) facilitate target decisions to a greater extent than orthographic controls (such as "brothel → BROTH"; Rastle, Davis & New, 2004; see also Rastle & Davis, 2008; McCormick, Rastle & Davis, 2008; Beyersmann et al., 2016). Such findings suggest that morphemes take on a special mechanistic role in the early visual processing of words in a manner beyond what can be attributed to the semantic and orthographic similarities between morphologically related words, but the exact nature of that role has been subject to much debate. The primary goal of this paper is to determine whether the emergence of such morphological effects, both transparent and opaque, can be demonstrated as a natural outcome of learning word-to-meaning mappings for a typical English vocabulary, without the need for additional information or processing features.

Although morphemes have traditionally been thought of as discrete units, recent work suggests that a more probabilistic view of morphology might be the most fruitful path forward for researchers in this area (as noted by Seidenberg & Gonnerman, 2000, and reiterated by Hay & Baayen, 2005, and Crepaldi, Marelli, and Amenta, 2019). Blurring morphological distinctions (e.g. morpheme vs. not, opaque vs. transparent, bound vs. free) helps to highlight the impact of a morpheme's frequency and usage in the language on how that morpheme is perceived. For example, Xu and Taft (2015) found that complex words with a higher base frequency, how often its base appears alone or in other words, are recognized more quickly for high-transparency words (EATER is recognized more quickly than FEVERISH) but not for low-transparency words (FAIRY and BADGER are recognized equally quickly; for frequency effects also see Taft & Forster, 1975; Giraudo & Grainger, 2000; Davis, van Casterend & Marslen-Wilson, 2003; Beyersmann & Grainger, 2018). The "diagnosticity" of derivational suffixes (the number of words having a certain suffix that fall in a given grammatical category, divided by the total number of words with that suffix) impacts how skilled readers make decisions about novel words with those suffixes (Ulicheva et al., 2018). For example,-ICAL is a highly diagnostic suffix for adjectives (nearly all words ending in -ICAL are adjectives), while -Y has low diagnosticity for adjectives, and this characteristic impacts how humans react to novel words with those suffixes across word reading, sentence reading, and spelling tasks. Plaut & Gonnerman (2000) demonstrated via neural network simulations that manipulating the degree of systematicity in form-to-meaning mappings for an artificial language impacts priming magnitudes between morphologically but not semantically related (i.e., morphologically opaque) words. Their demonstration provided an explanation for why opaque morphological effects are found weakly in English but more strongly in languages with greater semantic transparency of morphologically complex words, such as Hebrew. Along similar lines, Günther, Smolka, & Marelli (2019) used compositional distributional models of semantics to confirm that German overall exhibits greater semantic transparency for morphologically complex words than English, and argued that this characteristic of German explains its stronger opaque effects relative to those found in English. All of this work illustrates that the best explanation of behavioral effects in morphological processing, and specifically opaque effects, likely lies with the structure and usage of the language in which the study is run.

A central question for the view of morphological sensitivity as a result of a language's statistical structure is whether statistics are *sufficient* (Seidenberg & Gonnerman, 2000, p. 359). In other words, how much of human behavioral effects related to morphological processing can be explained by the statistical structure of the language, without needing to introduce additional mechanisms to the process? Note that "additional mechanisms" could refer to the explicit morpheme identification and word-splitting processes that are often proposed (e.g., Taft & Forster, 1975; Rastle, Davis & New, 2004; Rastle & Davis, 2008; Beyersmann & Grainger, 2017), or to gradated and distributed but dynamic and nonlinear processes as of yet relatively unexplored. The work presented here starts to test the sufficiency of language statistics by determining whether the emergence of transparent and opaque morphological priming effects in English are a direct result of learning the form-to-meaning mappings of a typical developing reader's vocabulary.

Many masked priming studies have been conducted to test various aspects of transparent and opaque morphological effects. The stimuli from two such studies were used to test the networks' morphological sensitivity: Rastle, Davis & New (2004), the first study clearly demonstrating opaque morphological priming in English-speaking adults, and Beyersmann et al. (2012), a cross-sectional study which aimed to localize the developmental emergence of this effect. Rastle et al. (2004) reported a masked primed lexical decision study: adult participants were presented with words and nonwords and asked to determine if each was word. Prior to the presentation of the "target" word on which the decision was to be made, a forward mask (######) was shown for 500 milliseconds (ms), followed very briefly (42 ms) by a priming word so as to not be consciously perceived. Rastle et al. compared the degree to which lexical decisions were sped up by the masked prime across three conditions: morphologically transparent, morphologically opaque, and form-related (i.e., orthographically related). They found that at this brief prime duration, opaque and transparent priming effects were equivalent and greater than orthographic effects, and interpreted this as evidence for rapid visual decomposition of complex words. In other words, when the word BROTHER is presented, they argue it is rapidly and automatically decomposed in BROTH and ER. That segmented representation remains to facilitate the recognition of BROTH when the prime presentation is too short for slower-retrieved information to arrive and override it. Prior studies have shown that such opaque effects don't occur at longer prime durations in English (e.g., Rastle et al., 2000). Beyersmann et al. (2012) conducted a cross-sectional masked priming lexical decision study with 7- to 9-year-olds, 9- to 11-year-olds and adults, also comparing the reaction times for transparent, opaque, and form-related pairs. Stimuli were based on those used in Rastle et al. (2004) but included more high-frequency words to maximize the number of words younger participants recognized. They found transparent but not opaque morphological priming in the two younger age groups, while opaque priming was present but weaker than transparent priming in adults. A more recent version of this study using identical methodology and stimuli but recruiting participants from a much broader age range (Dawson, Rastle & Ricketts, 2019) confirmed that pseudomorphological priming emerges later and more weakly than morphological priming. Dawson et al. additionally observed that age and word reading ability measures are both strong predictors of opaque priming effect size.

In the simulation work described below, we take an initial step towards explaining the developmental emergence of transparent and morphological effects found in primed lexical decision studies by investigating to what degree a neural network trained on a developmentally appropriate English vocabulary demonstrates similar patterns of performance to those observed in developing readers. Specifically, we trained a network to map from visual representations to semantic representations of words for a vocabulary based on the Touchstone Applied Sciences Association (TASA) corpus of educational texts (Zeno et al., 1995). We then tested it for morphological sensitivity and pseudomorphological sensitivity over the course of training. Neither the visual nor the semantic representations contained any explicit morphological information, and were derived directly from the words' spelling and usage. The network was tested on a procedure comparable to priming, using the same prime-target word pairs used in the Rastle et al. (2004) and Beyersmann et al. (2012) studies.

Simulation

A network was trained to map from orthographic input to a semantic representation for each word in a vocabulary typical of a developing English-speaking reader. Following training the network was tested on stimuli used in previously reported priming studies (Rastle et al., 2004; Beyersmann et al., 2012), to determine how well such results can be explained simply by exposure and sensitivity to the letter-to-meaning regularities present in a developmentally realistic collection of written words.

Stimuli

The vocabulary presented to the network for training was drawn from the Touchstone Applied Sciences Association (TASA) corpus of school texts used from first grade through the end of high school (Zeno et al., 1995), so as to avoid over-representing adult-oriented texts which do not accurately reflect word frequency distributions during reading acquisition. The TASA corpus was compiled for the purpose of approximating the frequency with which school-age students encounter particular words (e.g., Landauer & Dumais, 1997; Bhide et al., 2014). Words with more than 8 letters were removed to account for the fact that longer words may require multiple fixations to recognize, potentially impacting the visual processing involved. Any words not also contained within the American Heritage Dictionary (Watkins, 2000) were also excluded to account for aberrant words (such as character names in literature textbooks) and typos. Finally, the vocabulary was limited to types appearing 30 or more times within the TASA corpus, resulting in training set of 9,970 words.

To approximate the visual input received when looking at a word, a variation of the Overlap Model, proposed by Gomez, Ratcliff & Perea (2008), was used. Letters within a word were presented as overlapping normal distributions of activation, as opposed to more common slot-filling approaches to orthographic representations of words in connectionist models of reading (e.g. Plaut & McClelland, 1993; Plaut, McClelland, Seidenberg & Patterson, 1996; Zorzi, Houghton, & Butterworth,1998). This more continuous approach to orthographic representations prevents morphologically related words from being recognized by identical beginnings. For example, the first two columns of Figure 1 show that although ”kind” and ”kindly” have the same first 4 letters, there is no simple operation by which to identify morphological relatedness from their orthographic representations. Additionally, the probabilistic presentation of letters accounts for lower-level word reading phenomena such as the confusability of adjacent letters (Gomez, Ratfliff & Perea, 2008) which might impact the ease with which reoccurring subsets of letters are detected.

Unlike in the original overlap model, in which letters closer to the start of a word had sharper distributions, letters' distributions in our representations had the same standard deviation of 2 regardless of their position in the word. This was meant to reflect the visual experience of a novice reader, as fixating closer to word onset (facilitating greater acuity for earlier letters) is presumably learned through large amounts of visual word recognition experience (as suggested by differing fixation patterns across languages, e.g. Alhama et al., 2019). The center of each letter distribution was determined by word length and letter position: spacing between letters was calculated such that shorter words had slightly

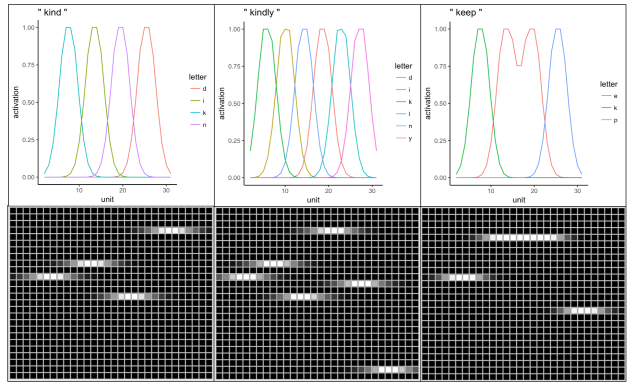

Figure 1: Orthographic representations visualized graphically (above) and unit-by-unit (below) for ”kind”, ”kindly” and ”keep”.

more spread between their letters than longer words[1], and all words were centered within the 30-unit space. If the same letter appeared multiple times in a word, the activations from normal distributions with different centers were summed to yield the total activation across the units for that letter, with activation capped at 1.0 (the rightmost column of Figure 1 illustrates this summation for the repeated *e* in *keep*).

Meaning representations for words in the training vocabulary were based on GloVe, an algorithm for learning vector space representations of words using co-occurrence information from text corpora that has been shown to capture human performance on semantic judgments well relative to similar methods (Pennington et al., 2014). Real-valued 300-dimensional semantic vectors generated from the Common Crawl internet text corpus were converted to 200-dimensional binary vectors using a binary multidimensional scaling algorithm (Rohde, 2002). Binary rather than real-valued vectors were used because it is easier for a network to drive sigmoid unit activations to extreme values as compared to intermediate values. The binarization process greatly reduced the dimensionality of the semantic representations while preserving the similarity structure (the pairwise distance matrices for the real-valued GloVe and binarized vectors were well correlated: r = 0.68, $p < .0001$). On average 33.61% of units were active for any word's semantic representation (SD = 7.27%).

Network Architecture

The LENS neural network simulator (Rohde, 2003) was used to build, train and test the network, which consisted of 780 input units, two 2000-unit hidden layers, and a 200-unit output layer. Hidden units used a rectified linear unit function

[1] Letter spacing was calculated as $\frac{N}{l+1} - exp(\frac{aN}{(l+1)-b})$ where N is the number of units over which the word is represented (30 for this simulation), l is the length of the word, and the parameters a and b were set to 1 and 9.1 so as to prevent the longest words from spilling beyond the available units and to maintain some letter overlap for short words.

(in an effort to speed up learning, given the multiple hidden layers and large example set), whereas the output units used a sigmoid unit function to avoid over-activation. At the start of training, all units from one layer were connected unidirectionally to the units of the downstream layer. Additionally, all units in the output and hidden layers received input from a bias unit.

Training Procedures

The initial weight values prior to training were somewhat constrained: weights connecting the bias unit to the hidden layers were randomly initialized (mean weight value 0.25, range of 0.1), while weights to the output layer were all set to -1.0 to aid the network's suppression of output activation during initial training epochs. All other weight values were randomized with a mean of 0 and a range of 0.03. The network was trained for 16,000 epochs using the delta-bar-delta learning algorithm, with a learning rate of $5\text{x}10^{-8}$ and momentum of 0.8. During training, the error and output unit error derivatives for each example was scaled by that word's frequency in the TASA corpus. By the end of training, 99.99% of output activations were within 0.5 of their targets, and 84.93% were within 0.3 of their targets. Weight values were saved every 100 epochs up to 1000 epochs, and every 1000 epochs subsequently, to allow analysis during multiple phases of training.

Testing Stimuli

The stimuli used for initial network testing were prime-target pairs used in the masked priming studies reported by Rastle et al. (2004) and Beyersmann et al. (2012). The Rastle set consisted of 50 morphologically transparent pairs (e.g., teacher → teach), 50 morphologically opaque pairs (e.g., corner → CORN), and 47 form-related pairs, as well as a control pair with the same target but an unrelated prime for each experimental pair. The Beyersmann set consisted of 34 prime-target word pairs in each of the same three conditions (although they referred to these conditions as "true-suffixed", "pseudosuffixed" and "nonsuffixed"), as well as matched control pairs. Of these 498 original prime-target pairs, 9 pairs and their controls were dropped because they occurred in both studies, 101 pairs and their controls were excluded because either a related prime or a target word was not present in the training vocabulary, and 14 unrelated pairs were altered to use an unrelated word present in the training vocabulary (selected randomly from the other primes used within that study). This left 278 prime-target pairs for testing procedures: 47 transparent pairs, 45 opaque pairs, 47 form pairs and 139 matched control pairs. The correlation between word representations for each testing pair was calculated to ensure that intended relations between primes and targets held true for our representations (see Table 1). A two-way ANOVA with an interaction term confirmed that orthographic correlations were significantly higher for related pairs than for controls ($t(2) = 33.04, p < 0.0001$) and this difference did not vary across conditions (ps > 0.25 for both interaction terms). A similar model with semantic correlations as the dependent variable and planned contrasts confirmed that semantic correlations were only higher for related relative to control pairs in the transparent condition ($t[2,2] = 17.60, p < 0.0001$).

	Orthographic		Semantic	
Condition	Related	Control	Related	Control
Form	0.43	0.02	0.04	0.03
Opaque	0.45	0.03	0.04	0.01
Transp.	0.40	0.05	0.40	0.02

Table 1: Mean correlations of prime and target representations used for network testing. Calculated for both orthographic and semantic representations.

Testing Procedure

In order to gather settling time data to compare with reaction times, during testing the network was run over multiple time steps, with all units in hidden and output layers integrating their inputs incrementally (with a time constant of 0.01). During each test trial, a particular prime-target pair was presented to the network. The orthographic representation of the prime was presented for 100 time steps (or "ticks"), and activations initiated by this input propagated through the network. Then the orthographic representation of the prime was replaced with that of the target. The network continued to run until the average amount of change in output units' activations between two ticks was less than 0.0001 (i.e., until the network settled to a semantic representation of the target word). The number of ticks needed for the output units to settle was recorded and used as the dependent variable in all analyses. This measure of processing time was used as a proxy for lexical decision reaction times measured in the modelled studies. On a single trial, the network could run for a maximum of 1500 ticks, but this maximum was never reached. Experimental pairs' settling times (e.g., teacher → TEACH) were subtracted from those of their corresponding control pairs (robbery → TEACH) to calculate the magnitude of a priming effect for each item in each condition.

Results

Simulated reaction times (number of ticks to settle to a semantic representation) for the testing stimuli taken from Rastle et al. (2004) and Beyersmann et al. (2012) are shown in the left panel of Figure 2. Priming magnitudes (prime-target pair reaction time subtracted from that of the matched control pair) are shown in the left panel of Figure 3.

A one-way ANOVA was run for testing performance after 1,000, 6,000, 11,000 and 16,000 epochs with priming magnitude as the dependent variable and condition as the independent variable. Despite showing a numerical trend consistent with predicted and experimental findings at every training phase, none of the tests showed a significant effect of condition (for 1,000 epochs: $F[2] = 2.03, p = 0.135$; for 6,000 epochs: $F[2] = 2.03, p = 0.135$; for 11,000 epochs: $F[2] = 1.45, p = 0.239$; for 16,000 epochs: $F[2] = 1.28, p = 0.281$).

Network reaction times increased over the course of training ($F[3] = 517.2, p < 0.0001$). This of course does not reflect the developmental trajectory of lexical decision reaction times (children tend to respond more slowly than adults) but is rather an artifact of the testing procedure: as the network achieves better performance, the semantic representations the output units settle to are more differentiated (closer to 0 and 1 and farther from 0.5) and thus take longer to reach.

The most likely explanation for the lack of a significant effect of condition on priming magnitude is low power: there is only one network performance being simulated with these testing pairs, compared with dozens of participants in each experiment. Increasing the power by running multiple, slightly varied simulations would allow our results to be more comparable to these studies, but that would require introducing our own theory of what causes individual differences in word recognition and lexical decision reaction times. Instead, power was increased via post-hoc analyses in which prime-target pairs from 8 other studies were added to the testing stimuli.

First Author	Year	Pairs Used	Unrelated Listed
Marslen-Wilson	2008	22	No
Jared	2017	114	No
Diependaele	2013	47	No
Feldman	2009	20	Yes
Beyersmann	2016	20	Yes
Li	2017	29	Yes
Rueckl	2008	13	No
Morris	2007	18	No

Table 2: Papers from which prime-target pairs were used for post-hoc analyses.

Post-hoc analysis: adding more testing pairs

To increase the number of prime-target pairs being compared during testing, 8 peer-reviewed papers were identified that (1) included a morphologically transparent, morphologically opaque, and form/orthographic condition, and (2) listed all related pairs in the paper itself or supplementary materials. No nonword stimuli were included. See Table 2 for a summary of the papers from which prime-target pairs were taken.

All prime-target pairs from these papers were considered for inclusion in post-hoc analyses. Pairs were not included if the related prime or the target word were not in the network's training vocabulary. If an unrelated prime wasn't present in the training vocabulary, it was replaced by a different randomly selected prime from the same paper that was in the training vocabulary. For papers that did not list their paired control primes, they were generated by permuting the related primes and re-pairing them to the targets. Doing this increased the total number of related test pairs from 139 to 422 (170 transparent, 120 opaque, 132 form). All representation correlation effects described for the previous stimulus set also held true for the extended set.

	Orthographic		Semantic	
Condition	Related	Control	Related	Control
Form	0.34	0.04	0.05	0.01
Opaque	0.38	0.05	0.04	-0.01
Transp.	0.37	0.05	0.37	0.01

Table 3: Mean correlations of prime and target representations in extended testing set.

Figure 2: Mean number of ticks for experimentally related prime-target pairs and their matched controls. Error bars denote standard error.

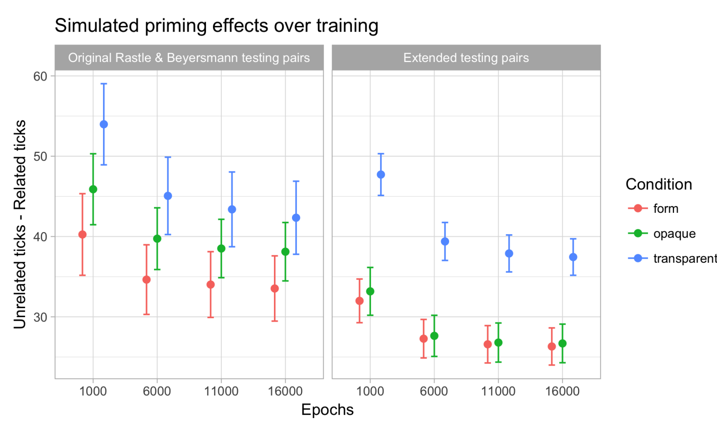

Figure 3: Simulated priming effects over training for stimuli from masked-priming lexical decision studies.

The extended testing pairs were then analyzed using the same one-way ANOVA analyses, and a significant effect of condition was found at all four phases of training (all Fs

> 7.5, all ps < 0.001). Planned contrasts showed a significant difference between the transparent and orthographic conditions at all phases of training (1,000 epochs: $t(419) = 4.66, p < 0.0001$; 6,000: $t(419) = 4.14, p < 0.0001$; 11,000: $t(419) = 4.02, p < 0.0001$; 16,000: $t(419) = 4.66, p < 0.0001$), as well as a marginal negative difference between the opaque and orthographic conditions at all phases of training (1,000 epochs: $t(419) = -1.89, p = 0.060$; 6,000: $t(419) = -1.82, p < 0.070$; 11,000: $t(419) = -1.79, p = 0.074$; 16,000: $t(419) = -1.73, p < 0.085$). In short, after increasing the number of testing pairs the network showed transparent morphological effects early and robustly, while opaque morphological effects do not appear.

Discussion

In this simulation work, we explored whether a simple neural network trained on a developmental English vocabulary could capture the emergence of morphological processing effects. The trained network showed strong transparent morphological priming effects when tested on a large stimulus set. This is the first computational demonstration to our knowledge that morphological sensitivity can emerge from authentic linguistic stimuli without any explicit knowledge of word structure being provided. All word representations were generated from letter strings and co-occurrence-based semantic vectors. However, the network did not capture the later, weaker increase in opaque morphological priming observed in experiments (e.g., Beyersmann et al., 2012). This suggests that while the regularities in form-to-meaning mappings can account some aspects of morphological processing in English, a more realistic and dynamic model is called for.

It is worth noting explicitly that this simulation is not meant to be a model of word processing. This is evident from aspects of the simulation that depart notably from human performance: response latencies increase instead of decreasing as training progresses and temporal dynamics such as the reversing impacts of orthographic and semantic priming with increased prime duration are not captured. These differences are unsurprising given the simple and entirely feed-forward training imposed on the network, and do not detract from the central aim of the simulation.

Nevertheless, a more realistic model is likely the best next step towards understanding these processes. Though still relying on neural networks as the vehicle of learning and processing, the authors' more recent work exploring complex word reading incorporates neurally plausible recurrent connections (as in Laszlo & Plaut, 2012) and a more naturalistic and dynamic training environment. With the right combination of network structure and language environment, such a neural network model may be "sufficient" to explain opaque morphological effects in a manner more easily integrated into other theories of cognition. Such a model would simultaneously provide an account of these processes' developmental emergence, which is lacking in the current literature.

The work presented here is unique in training a neural-network model on an authentic English vocabulary to better understand the emergence of morphological sensitivity.

References

Alhama, R. G., Siegelman, N., Frost, R., & Armstrong, B. C. (2019). The role of information in visual word recognition: A perceptually-constrained connectionist account. In *Proceedings of the 41st annual conference of the cognitive science society.*

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*, 232.

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.

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

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

Bhide, A., Schlaggar, B. L., & Barnes, K. A. (2014). Developmental differences in masked form priming are not driven by vocabulary growth. *Frontiers in psychology*, *5*, 667.

Crepaldi, D., Marelli, M., & Amenta, S. (2019). For a probabilistic and multidisciplinary approach to the investigation of morphological processing. *Cortex*, *116*, 1-3.

Davis, M. H., Casteren, M. van, & Marslen-Wilson, W. (2003). Frequency effects in the processing of inflectional morphology: A distributed connectionist account. In R. H. Baayen & R. Shreuder (Eds.), *Morphological structure in language processing.* Berlin: Mouton de Gruyter.

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. (2019). Individual differences in morphological processing in developing and skilled readers. *Society for the Scientific Study of Reading*, *116*, 1-3.

Diependaele, K., Morris, J., Serota, R., Bertrand, D., & Grainger, J. (2013). Breaking boundaries: Letter transpositions and morphological processing. *Language and Cognitive Processes*, *28*(7), 988-1003.

Feldman, L., O'Connor, P., & Prado Martin, F. del. (2009).

Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin Review*, *16*(4), 684-691.

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.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological review*, *115*(3), 577-600.

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

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

Jared, D., Jouravlev, O., & Joanisse, M. (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.

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.

Laszlo, S., & Plaut, D. C. (2012). A neurally plausible parallel distributed processing model of event-related potential word reading data. *Brain and Language*, *120*(3), 271-281.

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

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

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

McCormich, S. F., Rastle, K., & Davis, M. H. (2008). Is there a 'fete' in 'fetish'? effects in 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.

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

Nagy, W. E., & Anderson, R. C. (1997). How many words are there in printed school english? *Reading research quarterly*, 304-330.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing* (p. 1532-1543).

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. (1993). Generalization with componential attractors: Word and nonword reading in an attractor network. In *Proceedings of the 15th conference of the cognitive science society* (p. 824-829).

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (2000). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological review*, *103*(1), 56-115.

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 & review*, *11*(6), 1090-1098.

Rohde, D. L. (2002). Methods for binary multidimensional scaling. *Neural Computation*, *14*(5), 1195-1232.

Rohde, D. L. (2003). Lens: The light, efficient network simulator (version 2.63).

Rueckl, J., & Aicher, K. (2008). Are corner and brother morphologically complex? not in the long term. *Language and Cognitive Processes*, *23*(7).

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in cognitive science*, *4*(9), 353-361.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of verbal learning and verbal behavior*, *14*(6), 638-647.

Ulicheva, A., Harvey, H., Aronoff, M., & Rastle, K. (2018). Skilled readers' sensitivity to meaningful regularities in english writing. *Cognition*.

Watkins, C. (Ed.). (2000). *The american heritage dictionary of indo-european roots*. Houghton Mifflin Harcourt.

Xu, J., & Taft, M. (2015). The effects of semantic transparency and base frequency on the recognition of english complex words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *41*, 904-910.

Zeno, S. M., Ivens, S. H., Millart, R. T., & Duwuri, R. (Eds.). (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associate, Inc.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? a connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(4), 1131-1161.